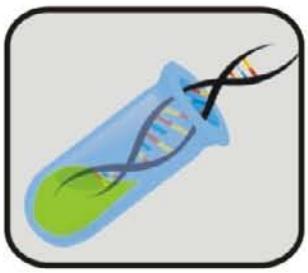
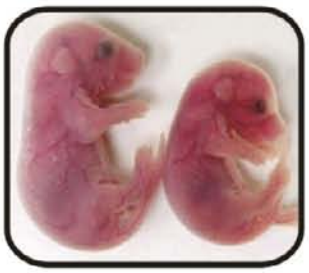
NEW AGE

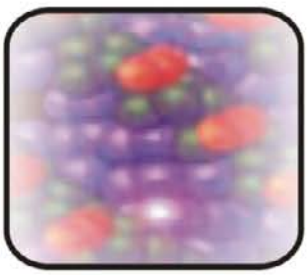
SECOND EDITION

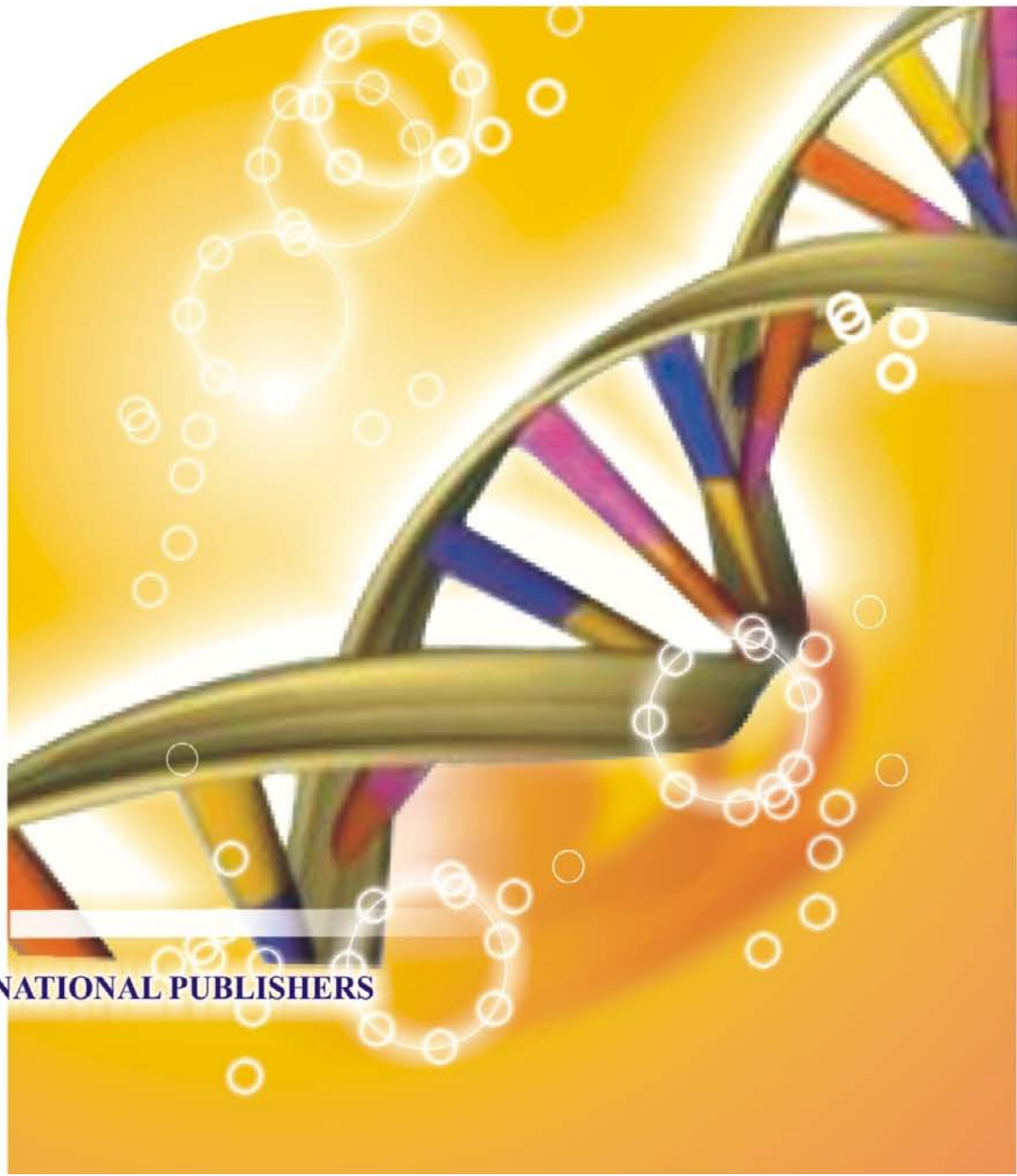
GENETICS

Karvita B. Ahluwalia



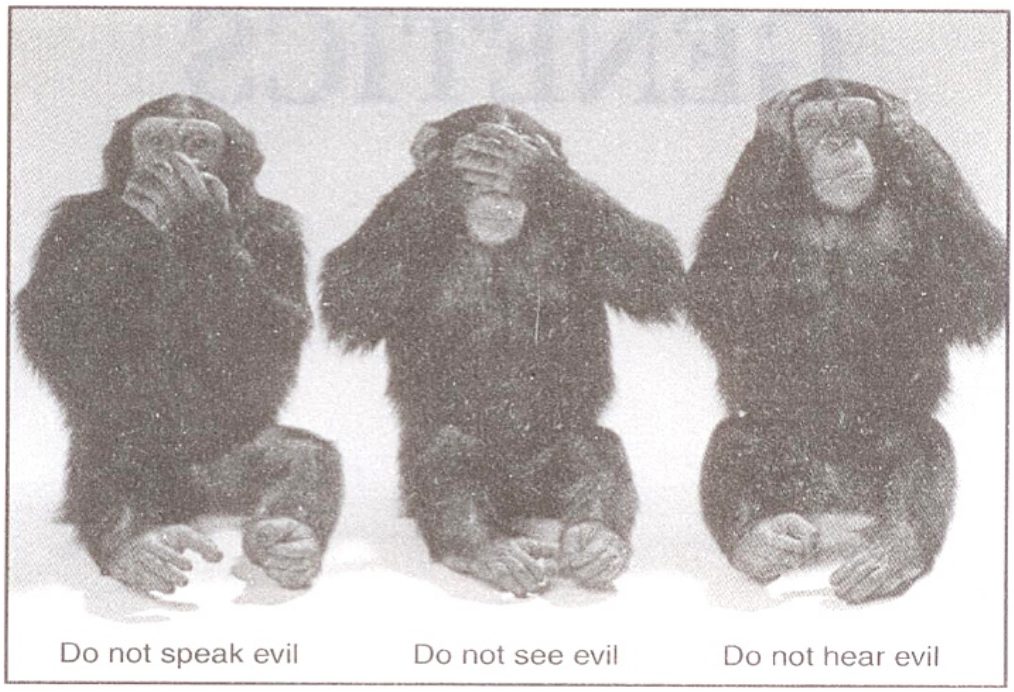






NEW AGE INTERNATIONAL PUBLISHERS

GENETICS



Turn Off Control Elements of Evil Genes

GENETICS

Second Edition

Karvita B. Ahluwalia
Formerly Professor
Cell Biology and Electron Microscope Section
Department of Biophysics
All India Institute of Medical Sciences
New Delhi, India

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS

New Delhi • Bangalore • Chennai • Cochin • Guwahati • Hyderabad
Jalandhar • Kolkata • Lucknow • Mumbai • Ranchi

Visit us at **www.newagepublishers.com**

Copyright © 2009, 1985, New Age International (P) Ltd., Publishers
Published by New Age International (P) Ltd., Publishers

All rights reserved.

No part of this ebook may be reproduced in any form, by photostat, microfilm, xerography, or any other means, or incorporated into any information retrieval system, electronic or mechanical, without the written permission of the publisher.
All inquiries should be emailed to ***rights@newagepublishers.com***

ISBN (13) : 978-81-224-2880-3

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
4835/24, Ansari Road, Daryaganj, New Delhi - 110002
Visit us at **www.newagepublishers.com**

Dedicated

to

My Parents

This page intentionally left blank

Preface to the Second Edition

In view of the dramatic advances in this rapidly progressing field, *Genetics* has been updated to present the most current information available. About 22 years ago when the first edition was printed, recombinant DNA technology was an emerging field, beginning to find its way into molecular biology laboratories. Over the years, the invention of new technologies have expanded recombinant DNA technology significantly. The sequencing of whole genomes, not only of prokaryotes and eukaryotes, even humans has been achieved successfully. In the scenario that now exists, sophisticated techniques have become more and more routine and available in many research laboratories. Bioinformatics has made inroads into, and become an integral component of molecular biology research. It is now time to develop new approaches for elucidating gene function and processes involved in running the metabolic machinery of the whole cell.

While modifying text for the second edition of *Genetics*, the sequence of chapters has not been changed so that the chronological order of developments in the field is kept in place. I consider topics in classical genetics not as mere historical landmarks. We cannot underestimate the work of Gregor Mendel who discovered heterozygosity, and dominant *versus* recessive traits; cytologists who explained "sticky ends" in chromosomes; and geneticists who made linkage and recombination maps accurately. Because today these topics are key elements of research in recombinant DNA technology and sequencing of genomes. Results of Mendel's experiments are still the backbone of pedigree analysis of human genetic disease. All in all, classical genetics provided the necessary foundations on which the molecular framework could be laid. Therefore, the student of genetics will do well in first comprehending the basics of genetics, that are nonetheless exciting, before proceeding to the nucleotide sequence of DNA. The text of *Genetics* therefore, gives equal importance to classical and modern genetics, and is intended to be useful not only for students of genetics, but also for nonspecialists in various disciplines of life sciences and medicine.

Karvita B. Ahluwalia

This page intentionally left blank

Preface to the First Edition

I have no special gift, I am only passionately curious.

Albert Einstein

Genetics has been written for university students who have had prior exposure to the subjects of biology and biochemistry. To maintain the balance of the text, equal importance has been given as far as possible, to discoveries of the Mendelian era, genetic mechanisms in microorganisms and recent advances in gene structure and function. The material of the book has not been divided into classical, chromosomal and molecular genetics. On the contrary attempts were made wherever possible, to illustrate the essential unity of the various topics. Due to research material outpouring in scientific journals, the frontiers of genetics are growing at great speed. Each topic covered in a chapter consequently is, or is becoming a specialised branch of genetics. Some effort has been made to provide the earliest as well as the latest in the field. The text has been organised in view of these facts. Each chapter first introduces the reader to elementary principles. After the basic knowledge, advanced information is given in increasing depth with the hope that the reader should get an insight into the recent developments in the field. For an integrated approach, attempts were made to trace and interconnect the characteristics of inheritance through the various branches of genetics.

Some of the emerging fields have been covered in detail. These include the left-handed helix, ribosomes, gene transcription, chromatin structure, genetic regulation, as well as split, overlapping and movable genes. Special attention has also been paid to the chapters on human genetics and genetic engineering. While preparing material aimed to benefit the students, I have drawn upon 18 years of experience in different fields of genetics. If I have succeeded in these aims, then this book could meet the requirements of undergraduates, and also be of some use to postgraduate students.

Genetics requires observing traits in nature, solving problems and reading literature. A set of questions and references are provided at the end of each chapter. The references include those cited in the text, relevant review articles and papers from journals, providing the student a variety of ways of entering the literature. The original papers cited from journals like *Cell, Journal of Molecular Biology* and some others are for the few interested in pursuing genetics as a career.

While rapid advances are being made, and new information is flowing in, some challenges continue to face us. We are still groping in the dark for regulators of gene expression in eukaryotes. What are the controlling elements that turn genes on and off so that exact quantities of required proteins are synthesised at any moment is not known. The way the nucleosome gets folded up into the highly condensed state of a metaphase chromosome is an enigma. And even

though recombinant DNA technology is receiving tremendous support and attention, the expression of eukaryotic genes in prokaryotic cells is yet to be achieved. The same is true for *in vivo* expression of genes in the field of gene therapy. To add to these, geneticists are now wondering about cross talk between adjacent chromosomes, about members of gene families conversing with each other. Sure enough we are still ignorant about the eukaryotic genome.

A number of colleagues provided useful suggestions relevant to different portions of the book. In particular, I wish to express my gratefulness to Professor R.K. Mishra, Dr. Sushil Kumar, Dr. M.S. Grewal and Dr. (Mrs.) V. Kothekar for reading some chapters critically and helpful comments. My appreciation of students who became curious about the subject and shared its excitement with me. Some of them read beyond their syllabus and even pointed out a couple of errors. I record my thanks to Kulwant who suggested that I write the book and for constant encouragement, and to Gurvesh and Isha for being patient throughout.

Karvita B. Ahluwalia

Contents

1

The Pre-Mendelian Era and Mendelism

Man's curiosity to know about transmission of hereditary characters is as old as humanity itself. From the earliest times it had been noticed that the offspring may resemble their parents, grandparents, or other relations. Around 300 BC the great Aristotle had observed that peculiarities of hair, nails and even gait may reappear in offspring, that characters not visible in an individual may also be inherited. It appears that the existence of heredity was taken for granted. Yet, the rules which govern transmission of characters were not understood. Explanations were sought everywhere, even in the prevailing beliefs of the time such as magic, witchcraft, gods and folklore.

Many considered heredity to be some sort of a blending process, because of which the offspring showed different "dilutions" of the parental characteristics. The concept of blending inheritance fitted well with ancient thinking as it could explain why some children were more like their parents whereas others were less. As far back as 400 BC an ancient Greek writer suggested the role of environment in producing variations in inheritance. The idea appears similar to the "Theory of acquired characteristics" proposed by Lamarck twenty-two centuries later.

The earliest indications about the material basis of heredity came from plant breeding procedures practised by farmers who were aware of various techniques of hybridisation and selection for developing new varieties. In the middle of the eighteenth century Carolus Linnaeus (1707–1778) a Swedish taxonomist, and two German plant breeders Josef Gottlieb Kölreuter (1733–1806) and Karl Friedrich Von Gaertner (1722–1850) performed artificial cross pollinations in plants and obtained hybrid offspring. In 1760 Kölreuter published a book in which he described 500 plant hybridisation experiments. Although his observations were similar to those of Mendel, he was not able to interpret them correctly. For example, in the cross between *Nicotiana rustica* and *N. paniculata* he could not understand the vigorous and sterile nature of the hybrid offspring. Gaertner carried out similar experiments on garden peas, the very material which Mendel studied successfully a few decades later. In crosses of maize plants Gaertner found that if one parent has red kernels the other yellow, the hybrids were all of the yellow type. When these hybrids were self fertilised the second generation showed roughly three yellow to one red plant—identical to the 3 : 1 ratio found later by Mendel.

It was about the same time, in the first quarter of the nineteenth century that three British botanists, Knight, Seton and Goss were also experimenting with inheritance of seed color in garden peas. They observed that a cross between plants with yellow seeds and green seeds produced all yellow seeded plants in the first generation. On self fertilisation they obtained

second generation offspring with both yellow and green seeds. Plants with green seeds on self fertilisation produced only green seeded offspring showing that green seed colour always breeds true.

The works of all the plant hybridisers were studied and analysed in detail by Mendel. It is noteworthy that most of the important concepts of Genetics emerged on the basis of classical experiments done on plants. Among animals, efforts of scientists were focussed on the study of mode of sexual reproduction. Spallanzani in 1785 obtained offspring from artificial insemination of dogs. In 1875 Oscar Hertwig observed that in sea urchins, fertilisation involves the union of sperm nucleus with that of the egg. In human beings the inheritance of some diseases, so well understood today, was not known at all until the beginning of the twentieth century. It is however thought that if Mendel's laws had not been formulated through studies of garden peas, they would have been discovered in the early part of this century through studies of the inheritance of blood groups or of inborn errors of metabolism in man.

Although Mendel published his work in 1865, it was not until 35 years later that its significance was realised when three independent workers published similar results. Due to the vast impact which Mendel's work had on the discovery of later concepts, the year 1900 is considered to represent the birth year of modern Genetics.

MENDEL AND THE BIRTH OF MENDELISM

One of the peasant families settled in the Czech village of Heinzendorf, then part of Austria, maintained a farm where developing new varieties of apples was their main occupation. In this family was born Gregor Johann Mendel in 1822. Mendel grew up in this farm environment and finished high school at the age of eighteen. Thereafter he tried to become a tutor at the college in Ölmutz but did not succeed. Perhaps due to this failure, or due to lack of money, he returned to his family farm where he spent one year. After that he did find work as a tutor, but it seems he could not settle down comfortably due to combined pressures of studying, teaching and some financial problems. He therefore left Ölmutz (now Brno in Czechoslovakia). He started off as a substitute teacher and in 1848 was ordained as a parish priest. In 1851 Mendel entered the University of Vienna for training in physics, mathematics and natural sciences. It was at Vienna that Mendel was influenced by two scientists, Franz Unger a plant physiologist, and Christian Doppler, discoverer of the well-known Doppler effect in physics. Perhaps Mendel picked up knowledge here about Kölreuter's and Gaertner's hybridisation experiments which formed part of Unger's teaching courses. It also seems likely that Mendel sharpened his mathematical awareness through Doppler's influence on him. After completing his studies at Vienna he returned to Brunn in Moravia, Czechoslovakia in 1854 where he continued to work as priest and as a teacher in high school.

MENDEL'S EXPERIMENTS

In 1856 Mendel began his experiments on plant hybridisation with garden peas in the monastery garden. Although similar work had already been done by contemporary botanists, the significant features of all these experiments had been overlooked because the investigators made overall observations of all inherited characters instead of collecting and analysing data in a systematic, mathematical way. This is how Mendel achieved what his predecessors could not. First of all he concentrated his attention on a single character in his experiments on inheritance. Secondly, he kept accurate pedigree records for each plant. And third, he counted the different kinds of plants resulting from each cross. Fourthly, he analysed his data mathematically.

Mendel's success is in part also attributed to his choice of material. The garden pea (*Pisum sativum*) used in his experiments (Fig. 1.1) offers certain advantages: it is an easily growing, naturally self fertilising plant; it is well suited for artificial cross pollination therefore hybridisation (crossing of two different varieties) is easily accomplished; it shows pairs of contrasting characters which do not blend to produce intermediate types and can be traced through successive generations without confusion. For example tall and dwarf are a pair of contrasting conditions for the character height; similarly round and wrinkled seeds are contrasting forms for the character seed texture. On self pollination each character breeds true. Mendel worked with seven pairs of characters so that he had 14 pure breeding varieties.

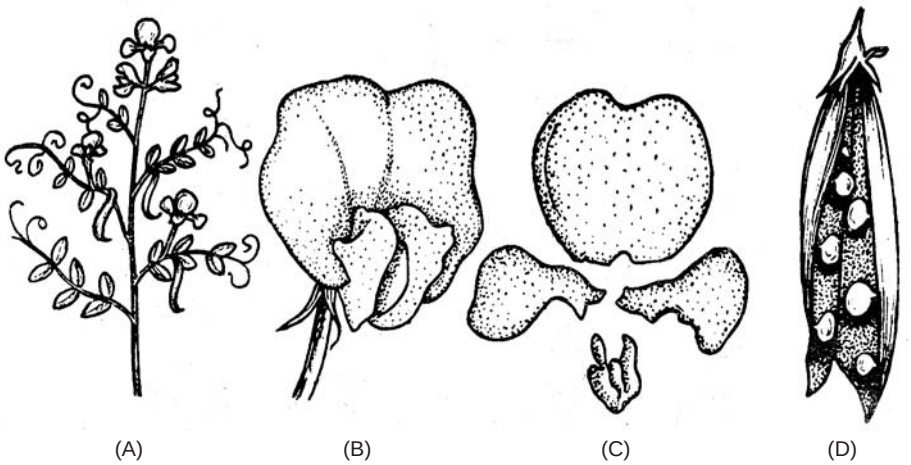


Fig. 1.1 The pea plant and parts of its flower and pod.

Monohybrid Cross

Mendel crossed varieties of edible peas which showed clearcut differences in morphological characters (Fig. 1.2) such as colour of flowers (red vs. white), shape of pod (inflated vs. constricted), colour of pod (green vs. yellow), texture of seed (round vs. wrinkled), colour of cotyledons (yellow vs. green), flower position (axial vs. terminal) and height of plant (tall vs. dwarf). He was dusting the pollen of one variety on the pistil of the other. To prevent self pollination of the female parent, he removed its stamens before the flowers had opened and shed the pollen. After making the cross he would enclose the flowers in bags to protect them from insects and foreign pollen.

Mendel's first experiments explain how a single gene segregates in inheritance. When Mendel crossed a true breeding tall plant (female parent) with a true breeding plant of the dwarf variety (male parent), he got tall plants like one parent in the first filial generation designated F_1. He used the term "dominant" for the *tall* character which dominated in the F_1 generation, and "recessive" for the character of *dwarfness* which remained hidden (latent) in the F_1 generation. Self fertilisation of the F_1 hybrids produced the second filial generation F_2 consisting of a total of 1064 plants of which 787 were tall and 277 were dwarf. That is tall and dwarf plants appeared in F_2 in the proportion of 2.84 : 1 which is roughly equal to 3 : 1. When he performed the reciprocal cross by reversing the sexes of the parents, the same results were obtained showing thereby that it did not matter which plant was used as male or as female parent. Similarly, Mendel crossed pea plants differing in other characters such as colour of flowers (red flowered versus white flowered), texture of seed (round versus wrinkled), colour of cotyledons (yellow versus green). Such a cross which involves only one character from each parent is called a *monohybrid cross*. In each case Mendel found one parental character

dominating in the F_1 hybrid, and after self fertilisation in F_2 generation both parental characters appeared in the proportion of three-fourths to one-fourth. He performed each experiment on several thousand plants and counted all the plants in F_2 progeny which gave an average ratio of 3 : 1.

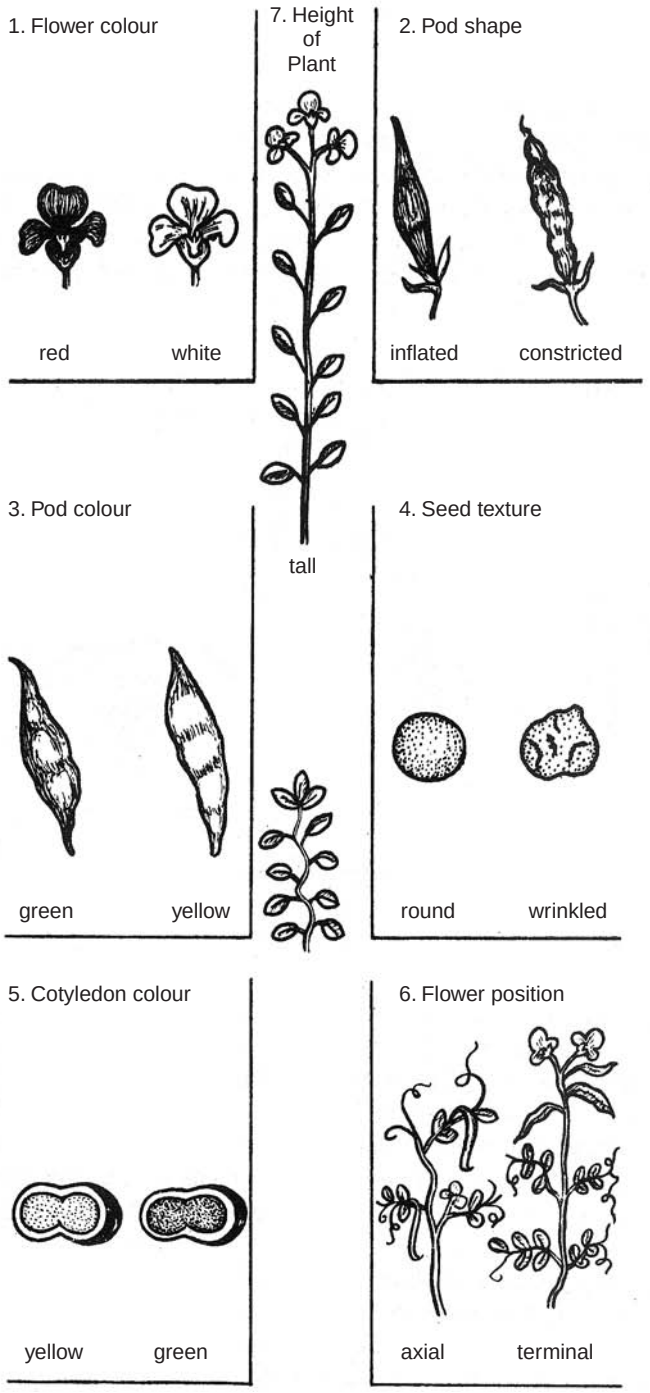


Fig. 1.2 Characters in pea plant used by Mendel in his crosses.

Segregation of Genes: From his experiments Mendel concluded that each parent contributes one *factor* for a character to the F_1 hybrid. In this way the F_1 hybrid has two factors for each character. When the F_1 hybrid forms gametes the two factors separate from each other. There is no mixing up of factors thus emphasising the purity of gametes. The phenomenon of separation became Mendel's First Principle and was later termed as the Law of Segregation. This is explained diagrammatically as follows:

P:	Tall (true breeding) × Dwarf (true beeding)		
Factors:	(*TT*)	(*tt*)	
Gametes:	*T*	*t*	
F_1 hybrid:		Tall × Self *i.e.* Tall (*Tt*)	
		(*Tt*)	
Gametes:		*T*, *t*	*T*, *t*
F_2:		3 Tall: 1 Dwarf	
		(*TT*, *Tt*, *tT*) (*tt*)	

Terms Used in Mendel's Crosses

Dominant versus *Recessive:* When two pure breeding varieties are crossed, the parental character that expresses itself unchanged in the F_1 generation hybrids is *dominant*: the one that does not appear in F_1 but appears in F_2 is called recessive. In the above cross three-fourths of the F_2 progeny show the dominant character and one-fourth the recessive character.

Alleles: Factors which control contrasting expressions of a character are said to be alleles or allelomorphs of each other. In the above cross the character in consideration is height, and factors *T* and *t* which control tallness and dwarfness are alleles of each other.

Homozygous and *Heterozygous:* These terms were coined by Bateson and Saunders in 1902. Mendel had concluded that each character is controlled by a pair of factors. When both factors are identical such as *TT* and *tt*, the individual is said to be *homozygous* for that character. When the factors are different (for example *Tt*), the term heterozygous is used. Mendel used capital letter of the alphabet to denote dominant factors, and small letters for recessive alleles.

Mendel's factors were later replaced by the term '*gene*' by a Danish botanist Johannsen in 1909. The word *genotype* refers to the genetic constitution of an individual, whereas *phenotype* refers to the external appearance or manifestation of a character.

Mendel selfed members of the F_2 progeny and found that out of the dominant types, one-third bred true for the dominant character, whereas two-thirds segregated into dominants and recessives in the ratio of 3 : 1. All the recessive plants of F_2 generation when selfed bred true for the recessive character. Mendel found similar results in monohybrid crosses with all the seven pairs of contrasting characters in *Pisum sativum.* After eight years of detailed investigations on thousands of pea plants, Mendel published his results in a paper entitled "Experiments in Plant Hybridisation" in the Proceedings of the Brunn Natural History Society in 1866. However, his work received no attention for 34 years until three scientists, DeVries in Holland, Correns in Germany and Tschermak in Austria working independently published their findings in 1900 and confirmed Mendel's results.

The Test Cross: Not satisfied with his work, Mendel himself subjected his results to a test. In the cross between tall and dwarf pea plants, the F_1 hybrids were all phenotypically tall but their genotypes were not only *TT* but also *Tt*. Consider a heterozygous hybrid plant *Tt*. When it forms gametes, the factors *T* and *t* segregate in the gametes in a 1 : 1 ratio. This means

that 50% of the gametes of an F_1 heterozygous hybrid carry the factor T and 50% the factor t. Mendel crossed such a hybrid plant (Tt) with a plant of the true breeding, dwarf variety (tt). All the gametes of the homozygous dwarf plant carried the recessive factor t. Every gamete of the recessive parent has 50% chance of combining with a gamete carrying T and 50% chance to combine with a t gamete from the heterozygous parent. This should result in 50% of progeny showing the tall phenotype and genetic constitution Tt, whereas 50% of the progeny should be phenotypically dwarf with genotype tt as explained diagrammatically below:

P:	Tall	(Tt)	×	Dwarf	(tt)
Gametes:	T(50%),	t(50%)		All	t
Test Cross	Tall	:	Dwarf	: :	1 : 1
Progeny:	(Tt)		(tt)		
	50%		50%		

Indeed Mendel's results of this cross agreed with the theoretical expectations thus providing additional experimental proof of the correctness of his interpretations. Such a cross where an individual is crossed to a double recessive parent to test and verify the individual's genotype is called a *testcross* or *backcross*.

In order to determine genotypes of the F_2 progeny, Mendel allowed the F_2 plants to self-fertilise and produce a third filial or F_3 generation. He found that the homozygous F_2 tall plants could produce only tall plants on self-fertilisation. This indicated their genotype to be TT. Similarly the F_2 dwarf homozygotes yielded only dwarf plants on selfing; their genotype was tt. The F_2 heterozygotes on self fertilising behaved identical to the F_1 hybrids and gave rise to tall and dwarf phenotypes in the ratio 3 : 1. This proved that their genotype was identical to that of F_1 hybrids *i.e.* Tt.

It is noteworthy that the genotypes of the parents are written as TT and tt instead of single T and t. This is in accordance with Mendel's hypothesis that each parent has two factors for a character. There is also a cytological explanation. The somatic chromosomes of all plants and animals exist in homologous pairs, one member of each pair coming from the paternal parent, other from maternal parent. A gene is a section of the chromosomal DNA which has information necessary for determination of a specific genetic trait. Suppose a hypothetical gene A occupies a particular site or locus on a given chromosome. The homologous chromosome contains at the identical locus an alternative gene a which controls the same trait as gene A, but in such a way as to produce a different phenotype for the same trait. The alternative genes at the same locus A and a are also called alleles. It is an astonishing fact that though Mendel knew nothing about genes, he could predict the existence of factors, which later turned out to be genes. During the reduction division of meiosis (Metaphase I), chromosomes of a pair separate and go to the opposite poles. Consequently genes or alleles segregate from each other and pass into different gametes.

Demonstration of Genetic Segregation

Mendel's F_1 hybrids (Tt) were all tall plants indistinguishable phenotypically. Sometimes homozygous and heterozygous plants show phenotypic differences. There is a seedling character for green pigment in soybeans. The homozygous (GG) soybean plant is dark green, the heterozygous (Gg) plant light green. The homozygous recessive (gg) produces a golden lethal seedling which dies in early stages due to lack of green pigment. If the heterozygous plants are grown to maturity and self pollinated, their progeny will again segregate as dark green, light green and lethal golden in the ratio of 1 : 2 : 1.

Differences between homozygous and heterozygous genotypes can sometimes be observed in the gametes. In rice, sorghum and maize, effect of the gene for waxy endosperm is visible in the pollen grains. Maize kernels which have waxy endosperm produce starch and stain blue with iodine; nonwaxy endosperm does not produce starch and stains red with iodine. In maize gene for waxy endosperm is located on chromosome 9. A homozygous plant with genetic constitution *WxWx* produces starch in endosperm and stains blue with iodine. In the heterozygous plant (*Wx wx*) the dominant gene causes starch production and the kernels stain blue with iodine. But kernels on homozygous recessive plants (*wx wx*) have no starch and stain red with iodine. If anthers of these plants are treated with iodine, the pollen grains stain in a similar way. In homozygous plants all the pollen grains stain blue. In heterozygous plants 50% of pollen grains stain blue (*i.e.* those containing *Wx*), whereas 50% stain red (*i.e.* pollen grains having *wx*). In the homozygous recessive plant, all the pollen grains stain red. If breeding tests are done by self pollinating the heterozygous F_1 plants, the F_2 progeny consists of blue staining kernels (*WxWx* and *Wxwx* plants) and red staining kernels (*wxwx* plants) in the ratio 3 : 1.

The Dihybrid Cross

Mendel made crosses between pea plants differing in two characters such as texture of seed and colour of cotyledons. Such a cross in which inheritance of *two* characters is considered is called a dihybrid cross.

First of all Mendel crossed a pea plant that was breeding true for round seeds with a plant that bred true for wrinkled seeds. The F_1 indicated that roundness was dominant over wrinkled texture of seedcoat. Similarly, by another cross he could determine that yellow colour of cotyledons was dominant over green. He now used as male parent a plant which bred true for both round and yellow characters and crossed it with a female parent that bred true for wrinkled green. As expected from the results of his single crosses, the F_1 was round yellow. When he selfed the F_1 hybrids, the F_2 progeny showed all the parental characters in different combinations with each other. Thus plants with round yellow seeds, round green seeds, wrinkled yellow seeds and wrinkled green seeds all appeared in the ratio 9 : 3 : 3 : 1. Reciprocal cross in which the female parent was round yellow and male parent wrinkled green gave the same results.

Mendel applied the principle of a monohybrid cross and argued that in the dihybrid cross the true breeding round yellow parent must be homozygous *RRYY*, and the wrinkled green parent *rryy*. Since each character is determined be two factors, in a dihybrid cross there must be four factors present in each parent. Likewise the F_1 hybrid must be *RrYy*. But the question remained as to how did the four different combinations of parental phenotypes appear in the progeny? Mendel argued that the pair of factors for roundness must be behaving independently of the pair of factors for yellow colour of seeds. In other words, one factor for a character must be passing independently of a factor for another character. Thus in the F_1 hybrids, *R* and *r* pass into different gametes. Now the probability of an *R* gamete formed is one-half, and of *r* gamete also one-half. Similar probabilities exist for *Y* and *y* gametes. It follows that the probability that *R* and *Y* should go to the same gamete is one-fourth, as also of *R* and *y*, *r* and *Y*, and *r* and *y*. Therefore, gametes containing factors *RY*, *Ry*, *rY* and *ry* should form in equal proportions.

The F_1 hybrid producing the four types of gametes mentioned above was selfed. The results expected in the F_2 progeny can be predicted by making a checkerboard or a Punnett Square. Gametes produced by one parent are plotted on top of the checkerboard, and gametes

of the other parent on the side. The sixteen squares of the checkerboard are filled up by making various possible combinations of male and female gametes during fertilisation. The phenotypes read out from the checkerboard indicate a 9 : 3 : 3 : 1 ratio exactly as observed by Mendel.

P:	Round Yellow	×		Wrinkled Green
	RRYY			*rryy*
Gametes:	*RY*			*ry*
F_1:		Round Yellow	×	Self (Round Yellow)
		RrYy		*RrYy*
Gametes:		*RY, Ry, rY, ry*		*RY, Ry, rY, ry*

	RY	*Ry*	*rY*	*ry*
RY	*RRYY*	*RRYy*	*RrYY*	*RrYy*
Ry	*RRYy*	*RRyy*	*RrYy*	*Rryy*
rY	*RrYY*	*RrYy*	*rrYY*	*rrYy*
ry	*RrYy*	*Rryy*	*rrYy*	*rryy*

As in the case of the monohybrid cross, Mendel verified his results by performing the test cross. He crossed the F_1 hybrid heterozygous for both characters with a double recessive parent (*rryy*) which should produce only one type of gamete *ry*. The uniformity in the gametes of the recessive parent determines the differences in the types of gametes produced by the heterozygous parent. Now the hybrid *RrYy* produces gametes carrying *RY*, *Ry*, *rY* and *ry* with equal frequency. It follows that during fertilisation if *all* these four types of gametes unite with *ry* gamete of the recessive parent, the resulting progeny should show all the four combinations of characters also in equal proportions. Indeed, Mendel observed the testcross progeny to consist of Round Yellow, Round Green, Wrinkled Yellow and Wrinkled Green plants in the ratio 1 : 1 : 1 : 1.

F_1:	Round Yellow	×	Wrinkled Green
	RrYy		*rryy*
Gametes:	*RY, Ry, rY, ry*		*ry*

F_2:		*RY*	*Ry*	*rY*	*ry*
	ry	*RrYy*	*Rryy*	*rrYy*	*rryy*

From the results of his dihybrid crosses, Mendel realised the following facts. At the time of gamete formation the segregation of alleles *R* and *r* into separate gametes occurs independently of the segregation of alleles *Y* and *y*. That is why the resulting gametes contain all possible combinations of these alleles, *i.e. RY*, *Ry*, *rY, ry*. In this way Mendel proved that when two characters are considered in a cross, there is independent assortment of genes for each character, and this became the *Law of Independent Assortment*.

Trihybrid and Multihybrid Crosses

Mendel extended his observations to trihybrid crosses involving three pairs of contrasting characters. The characters he considered were: seed shape—smooth (*S*) vs. wrinkled (*s*); colour of cotyledons—yellow (*Y*) vs. green (*y*); and flower colour—violet (*V*) vs. white (*v*).

P:	Smooth Yellow Violet	×	Wrinkled Green White
	SSYYVV		*ssyyvv*
Gametes:	*SYV*		*syv*
F_1:	Smooth Yellow Violet × Self		
	SsYyVv		
Gametes:	*SYV*, *SyV*, *SYv*, *Syv*, *sYV*, *syV*, *sYv*, *syv*		
F_2:	27 Smooth Yellow Violet: 9 Smooth Yellow White 9 Smooth Green Violet: 9 Wrinkled Yellow Violet 3 Smooth Green White: 3 Wrinkled Yellow White 3 Wrinkled Green Violet: 1 Wrinkled Green White.		

The F_1 hybrid produces 8 types of gametes. These on selfing have equal chances to combine with any of the 8 types of gametes produced by the other parent resulting in 64 different combinations. All the dominant phenotypes are expressed. In this way it is possible to predict genotypes and phenotypes in F_2 of crosses involving more than 3 genes (multihybrid crosses). In each case number of gametes formed by F_1 heterozygote is determined by the formula 2^n, where n represents the number of characters. Thus in a trihybrid cross $2^3 = 8$ gametes result. In a cross involving 4 characters, $2^4 = 2 \times 2 \times 2 \times 2 = 16$ gametes must result. The number of F_2 phenotypes resulting from selfing F_1 hybrid is a square of the number of gametes. Thus in a trihybrid cross there are $8 \times 8 = 64$ phenotypes, in a tetrahybrid cross $16 \times 16 = 144$ phenotypes, and so on.

The experiments of Mendel laid the foundations of our knowledge of heredity. After seven years of detailed investigations he formulated his Principles and presented them at two meetings of the Natural History Society of Brunn in 1865. His results were published in the Annual Proceedings of the Society in 1866, which was distributed in the libraries of Europe and America the same year. But, unfortunately, nobody paid attention to this work and it remained unrecognised for 34 years. It was only in 1900 that three scientists—Correns in Germany, DeVries in Holland and Tschermak in Austria independently rediscovered the same principles of heredity, and the significance of Mendel's work was realised.

QUESTIONS

1. Name two plant breeders of the 18th century who knew those hybridisation techniques which Mendel had used in his experiments. Why did they not succeed as Mendel did?
2. When a pure breeding brown cat is mated to a pure breeding white cat, all the kittens born are brown. What results would you expect when an F_1 brown kitten is mated with
   (*a*) a pure breeding white cat;
   (*b*) another F_1 brown kitten;
   (*c*) a pure breeding brown cat.
3. When the brown kittens of the above cross grow up, they can no longer be distinguished from the brown mother cat. Suggest a cross which could identify the parent brown cat from among its F_1 progeny.

4. An ambitious plant breeder wanted to raise true breeding maize plants with kernels rich in starch grains. He applied the iodine test to split maize kernels and found out which seeds could produce plants with waxy endosperm. By what method could he find out which seeds would produce true breeding plants for starch-filled kernels.
5. In man brown eyes (*B*) are dominant over blue (*b*). A brown eyed man married to a brown eyed woman has a blue eyed child. What would be the genotypes of the parents?
6. In sweet peas yellow seed coat (*Y*) is dominant over green (*y*), and round seed shape (*R*) is dominant over wrinkled (*r*). What phenotypic ratios would be obtained in F_1 progeny of the following crosses:

   (*a*) *YyRR* × *YYRr*; (*b*) *yyRr* × *YyRr*;

   (*c*) *Yyrr* × *yyrr*; (*d*) *YyRr* × *yyRR*.
7. A cross between yellow round and green wrinkled seeded pea plants gave the following four types of progeny in equal proportions: yellow round, yellow wrinkled, green round and green wrinkled. What are the genotypes of the parents?
8. In man brown eyes (*B*) are dominant over blue (*b*), dark hair (*D*) over red hair (*d*), and right handedness (*R*) over left handedness (*r*). What would be the outcome of marriages between persons of the following genotypes:

   (*a*) *BBDdRr* × *BbDdRR*

   (*b*) *bbDdRr* × *BBDDRR*

   (*c*) *BbDDRr* × *bbddrr*

SELECTED READINGS

Blixit, S. 1975. Why didn't Gregor Mendel Find Linkage? *Nature* **256**: 206.

Carlson, E.A. 1966. The Gene: A Critical History. Saunders, Philadelphia.

Dunn, L.C. 1965. A Short History of Genetics. McGraw-Hill, New York.

IItis, H. 1932. The Life of Mendel. Allen and Unwin, London.

McKusick, V. 1969. Human Genetics. 2nd ed. Prentice-Hall, N.J.

Mendel, G. 1866. Versuche über Pflanzen hybriden. Verh. naturf. Vereines in Brunn. IV.

Roberts, H.F. 1929. Plant Hybridisation before Mendel. Princeton Univ. Press, N.J.

Stern, C. 1950. The Birth of Genetics. *Genetics* **35**, Supplement.

Whitehouse, H.L.K. 1973. Towards an Understanding of the Mechanism of Heredity. 3rd ed. St. Martin's Press, New York.

2

Interaction of Genes

The phenotypic ratios obtained by Mendel in garden peas demonstrate that one gene controls one character; of the two alleles of a gene, one allele is *completely* dominant over the other. Due to this the heterozygote has a phenotype identical to the homozygous parent. Soon after Mendel's work was rediscovered, instances came to light where a gene was not producing an individual effect. On the contrary, genes were interacting with each other to produce novel phenotypes which did not exhibit dominance relationships observed in Mendel's experiments.

In one of the first cases reported by Kölreuter, the heterozygote showed a phenotype intermediate between the parental phenotypes. This was termed *incomplete dominance* or *intermediate inheritance.* In *codominance* the heterozygote expresses both the parental phenotypes equally. Sometimes a gene masks the expression of another gene at a different locus. This is known as *epistasis.* On still other occasions a gene does not completely mask another gene as in epistasis, but, in some way modifies the effect of the second gene. Known as *modifying* genes, such genes either enhance or suppress the expression of a different gene. Interaction between genes enables some genes to act together to produce an effect that neither gene can produce separately. Such genes are said to be *complementary.* There are genes that copy other genes, so to say, to produce a similar effect. Thus independent genes that produce the same effect are given the name *duplicate* genes. Lastly, but most serious are genes causing death. They are known as *lethal genes.*

INCOMPLETE DOMINANCE

A monohybrid cross between a red-flowered snapdragon (*Antirrhinum majus*) and a white flowered variety does not produce red or white flowered plants in F_1 as expected from mendelism. Instead the flowers are pink, *i.e.* intermediate between the two parents. This is because neither red flower colour nor white is dominant, but each allele has its influence in color development and the hybrid appears pink. If the F_1 pink flowers are self-pollinated, the F_2 progeny shows red, pink and white flowered plants in the proportion 1 : 2 : 1. It may be recalled that this is the same genotypic ratio that Mendel obtained in garden peas. The difference is that in the present case the heterozygous progeny is distinct in appearance from the homozygotes. The name *intermediate inheritance* is also given to crosses where F_1 hybrids show incomplete or partial

dominance with no phenotypic resemblance to either parent. This type of inheritance has been found in crosses involving many other plants.

CODOMINANCE

In 1900 Bateson and his colleagues studied inheritance of comb shape in fowls. There are four types of combs in fowls: rose, pea, walnut and single. Bateson first performed a cross between rose and single. The F_1 hens all had a rose comb, and on inbreeding gave rise to an F_2 progeny of rose and single in the ratio 3 : 1. The cross indicates that rose and single comb are controlled by a single gene and that rose is dominant over single. In the second cross when chickens with pea comb were mated with single comb, the F_1 progeny had pea comb, and F_2 had pea and single in the proportion 3 : 1. Obviously, the gene for pea comb is dominant over single. This raises an interesting question—are the genes for rose and pea comb same or different? Bateson then crossed rose and pea. Surprisingly, the F_1 birds had an altogether different comb of the walnut type! An F_2 progeny raised by inbreeding the walnut type consisted of four types of chickens—walnut, rose, pea and single (Fig. 2.1) in the ratio 9 : 3 : 3 : 1. As this ratio is typical for dihybrid inheritance it became clear that rose and pea combs were controlled by two pairs of genes.

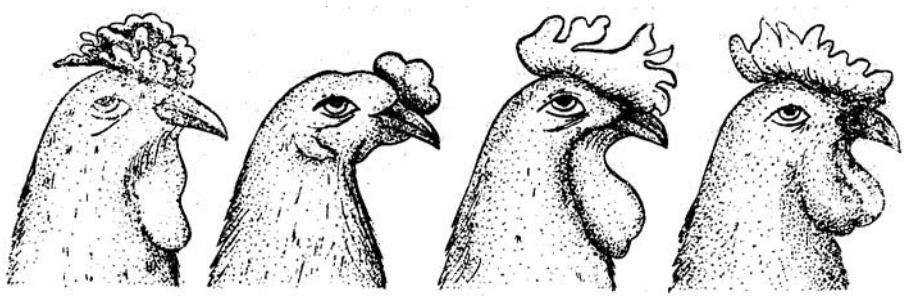


Fig. 2.1 Comb types in fowls.

P:	Rose comb			×		Pea comb
	RRpp					*rrPP*
Gametes:	*Rp*					*rP*
F_1:		Walnut		×		Walnut
		RrPp				*RrPp*
Gametes:		*RP, Rp, rP, rp*				
F_2:	9 Walnut	:	3 Rose	:	3 Pea	: 1 Single
	RRPP		*RRpp*		*rrPP*	*rrpp*
	RrPP		*Rrpp*		*rrPp*	
	RRPp					
	RrPp					

The appearance of walnut comb in F_1 of cross between rose and pea shows that both of the independent dominant genes *P* and *R* are jointly responsible for the walnut comb. When

present together in the zygote, *P* and *R* genes interact to produce the walnut comb. When present alone, they produce rose or pea comb. The recessive alleles of rose and pea combs produce the fourth type of chicken with the single comb.

Another example where the heterozygote can be recognised distinctly is offered by the "roan" (reddish grey) coat colour of short horn cattle. When homozygous red-haired cattle are crossed with homozygous white-haired type, the F_1 has reddish grey hair and is designated "roan". It must be noted that there is no mixture of red and grey pigments in a roan. But some hair are all red, others all white, so that the final effect is a reddish grey coat colour. The cross therefore demonstrates a difference from intermediate inheritance. In F_2, codominant genes segregate in the ratio 1 : 2 : 1.

P:	Red		×		White	
	r_1r_1				r_2r_2	
F_1:	"Roan" (reddish grey)			×	"Roan"	
		r_1r_2				r_1r_2
F_2:	1 Red	:	2 "Roan"	:	1 White	

The *ABO* blood group system in man is controlled by multiple alleles of a gene *I*, each allele producing a different antigen. Likewise there are four blood group phenotypes in man designated *A*, *B*, *AB* and *O*. Individuals with blood group *O* have no antigen, whereas those with *AB* represent heterozygotes in which both *A* and *B* antigens are present. The *AB* heterozygote has both dominant alleles I^A and I^B equally expressed to produce distinct *A* and *B* antigens. Similarly, the rare blood group *MN* also shows both *M* and *N* antigenic specificities expressed equally in the heterozygote.

EPISTASIS

Due to the phenomenon of dominance a recessive *allele* remains obscure in the hybrid. But when two different genes which are *not* alleles, both affect the same character in such a way that the expression of one masks, inhibits or suppresses the expression of the other gene, it is called *epistasis.* The gene that suppresses is said to be epistatic, and the gene which remains obscure is hypostatic.

Dominant Epistasis

In poultry white birds belong to two *different* varieties namely white leghorns or white wyandottes. Experiments reveal that the gene for white plumage of white leghorns is *dominant* over the gene for coloured plumage of coloured varieties. But the gene for white plumage of white wyandottes is recessive to the gene for coloured plumage of coloured varieties. Therefore the gene which produces white plumage in white leghorns is different from the gene for white plumage in white wyandottes.

A cross between a white leghorn and a white wyandotte gives an F_1 of white birds with small dark flecks. When such birds are inbred, the F_2 progeny segregates in the ratio of 13 white to 3 coloured birds. The experiment is explained below by postulating two genes *C* and *I* for the white leghorns:

P:	White Leghorn	×	White Wyandotte	
	IICC		*iicc*	
Gametes:	*IC*		*ic*	
	White with small			
F_1:	dark flecks	×	Inbred	
	IiCc			
F_2 :	White	:	Coloured	
	13		3	

A checkerboard for the 16 phenotypes and genotypes of the F_2 birds indicates that only three out of sixteen genotypes, that is *iiCC, iiCc, iiCc* produce coloured birds. The white leghorns obviously contain a gene *I,* which in the dominant state inhibits or suppresses the expression of the dominant colour gene *C,* resulting in white plumage. The recessive alleles of the inhibitor gene (*ii*) produce coloured birds due to expression of gene *C.* In other words gene *I* is epistatic to gene *C*. This is a case of *dominant epistasis* because even one dominant allele of gene *I* is able to express itself.

Recessive Epistasis

Epistasis due to recessive genes is called recessive epistasis. In mice albinism (white coat) is produced by a recessive gene *aa*. There is a different gene *B* which in the dominant state (*BB* and *Bb*) produces grey coat colour called agouti, and when recessive (*bb*) leads to black coat colour. The recessive gene for albinism (*aa*) is found to be epistatic to the gene for agouti (*BB* and *Bb*), and also to its recessive, homozygous allele (*bb*) for black. The presence of the dominant allele (*AA*) of the epistatic gene allows expression of gene *B* so that agouti (*BB* and *Bb*) and black (*bb*) coat colours can be produced (Fig. 2.2).

P:	Agouti	×	Albino	
	AABB		*aabb*	
Gametes:	*AB*		*ab*	
F_1:	Agouti	×	Agouti	
	AaBb		*AaBb*	
Gametes:	*AB*, *Ab*, *aB*, *ab*			

	AB	*Ab*	*aB*	*ab*
A B	*AABB* Agouti	*AABb* Agouti	*AaBB* Agouti	*AaBb* Agouti
Ab	*AABb* Agouti	*AAbb* Black	*AaBb* Agouti	*Aabb* Black
aB	*AaBB* Agouti	*AaBb* Agouti	*aaBB* Albino	*aaBb* Albino
ab	*AaBb* Agouti	*Aabb* Black	*aaBb* Albino	*aabb* Albino

F_2: 9 Agouti : 3 Black : 4 Albino

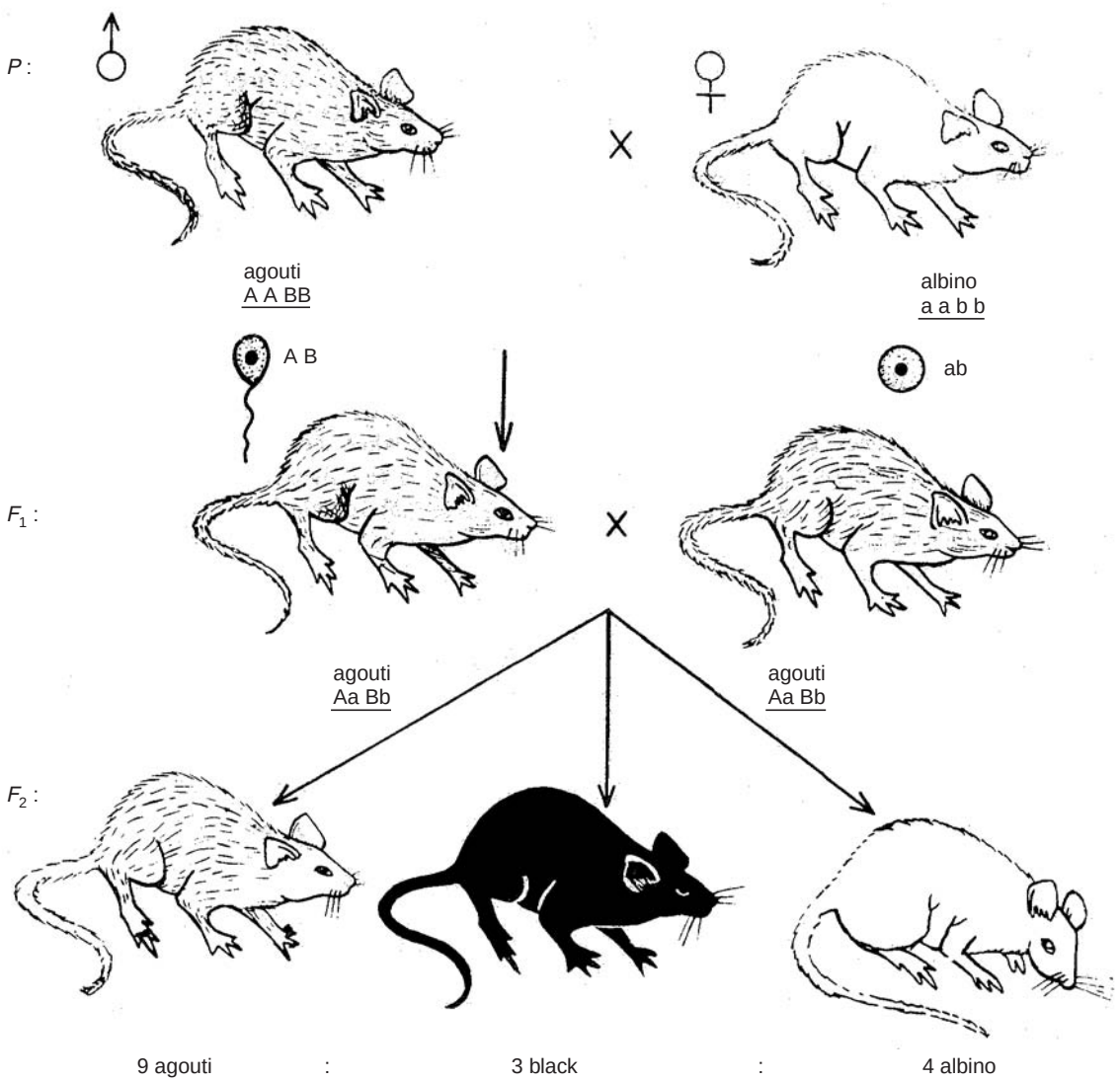
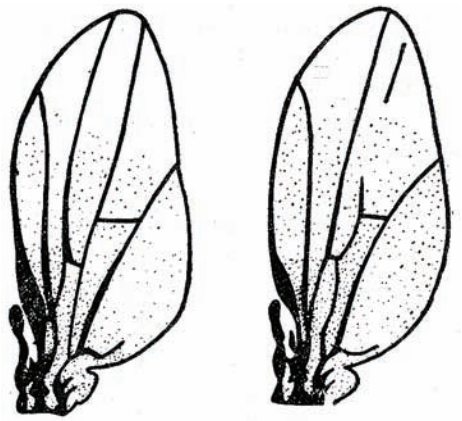


Fig. 2.2 Inheritance of the coat colour gene in mouse illustrating recessive epistasis.

The 9 : 3 : 4 ratio obtained is a modification of the classical 9 : 3 : 3 : 1 in which the last two classes (3 : 1) are phenotypically identical and are therefore added up together. In human beings also the recessive gene for albinism shows epistasis in a similar manner. Epistatic effect is usually only in one direction, from one particular gene pair to another.

Epistasis in Drosophila

There are two recessive wing mutants in *Drosophila: apterous* (*ap*) which produces small stubby wings instead of the normal transparent ones; the other called *cubitus interruptus* (*ci*) which causes a small interruption in the fourth longitudinal vein (Fig. 2.3). When the two mutants are used in a dihybrid cross, the F_2 progeny segregates in the ratio of 9 normal: 3 interrupted vein: 4 apterous wing.



normal cubitus interruptus

Fig. 2.3 Normal and interrupted veins in lateral wings of *Drosophila.*

P:	apterous	×		interrupted vein
	ap/*ap*+/+			+/+*Ci*/*Ci*
Gametes:	*ap*+			+*Ci*
F_1:		Wild type	×	Inbred
		ap/+*Ci*/+		
F_2:	9 Wild type	: 3 interrupted vein	:	4 apterous wing.

In the cross above, the presence of the homozygous recessive mutant gene for apterous wing (*ap*/*ap*) marks the expression of the gene for interrupted vein.

Epistasis and Blood Groups in Man

In the *ABO* blood group system, a person with blood group *A* has antigen *A* on the surface of red blood cells; a blood group *B* person has *B* antigen: an *AB* person has both *A* and *B* antigens, whereas type *O* has neither *A* nor *B* antigen. The antigens are controlled by an autosomal gene *I* (isohaemaglutinin) which has multiple alleles. Thus I^A controls antigen *A*, I^B controls antigen *B*. The recessive allele *i* produces no antigen and results in phenotype *O*.

Certain proteins present in the blood serum of a person show a precise relationship with the red cell antigen. These proteins are the *agglutinins* or *antibodies*. Thus the serum of an individual with blood group *A* has antibodies against antigen *B*, whereas the serum of a blood group *B* person is anti-*A*. In persons with blood group *AB* where both antigens *A* and *B* are present, the serum has neither antibody *A* nor *B*; individuals with blood group *O* have both types of antibodies.

It has been found that there is another gene *H* which controls production of a precursor substance *H*. In persons of blood type *A*, having gene I^A, precursor *H* gets converted into antigen *A* under control of gene I^A. Similarly when gene I^B is present, *H* gets converted to antigen *B*. In the presence of the recessive allele *i* in type *O* persons, precursor *H* is not converted and remains as such in the blood. It can be agglutinated by anti-*H* or by such substances as an extract of the seeds of a plant *Ulex europeus* or blood of eel. In persons of heterozygous blood groups such as I^A/i or I^B/i only part of the *H* substance which is unconverted is agglutinated. In

AB persons there is likewise no agglutination of *H* substance. It is found that most human beings are homozygous (*HH*) for gene *H*, few are heterozygous (*Hh*), and very rarely a homozygous recessive (*hh*) individual is met with that cannot produce any *H* substance at all. Such a person may have any blood group phenotype from the *ABO* system, but his red cells are not agglutinated by anti-*A*, anti-*B*, anti-*H*. First discovered in Bombay, India, a homozygous recessive (*hh*) person is said to show the *Bombay phenotype*. It demonstrates epistatic action of *hh* gene over *I* gene.

The secretor trait in humans also shows epistatic effect of genes. Just as red blood cells carry antigens on their surface, most persons have these antigens in a water soluble form in some body secretions such as saliva, gastric juice, fluids from nose, eyes and mammary glands of females. Such individuals are known as *secretors*. In a few persons regardless of blood group present, there are no antigens in body secretions. They are known as *nonsecretors*. The secretor trait is due to a dominant autosomal gene *Se*, whose recessive allele *sd* results in a nonsecretor. The dihybrid inheritance pattern of secretor trait and blood groups has been studied in a large number of families. Marriages between two double heterozygotes for both blood groups and secretor trait, *i.e.,* $I^AiSese \times I^AiSese$ or $I^BiSese \times I^BiSese$ show progeny phenotypes in the ratio 9:7 indicating epistatic interaction of genes.

$I^AiSese \times I^AiSese$

Gametes: I^ASe, I^Ase, iSe, ise

	I^ASe	I^Ase	iSe	ise
I^ASe	I^AI^ASeSe	I^AI^ASese	I^AiSeSe	I^AiSese
I^Ase	I^AI^ASese	I^AI^Asese	I^AiSese	I^Aisese
iSe	I^AiSeSe	I^AiSese	$iiSeSe$	$iiSese$
ise	I^AiSese	I^Aisese	$iiSese$	$iisese$

The checkerboard shows that there are 9 individuals showing both dominant phenotypes (blood group *A* secretors), 3 type 0 secretors and 4 nonsecretors. The data indicate epistatic interaction of genes.

COMPLEMENTARY GENES

Bateson and Punnett crossed two different white flowered varieties of sweet pea and obtained an F_1 progeny of red flowered plants. On self pollination the F_1 plants gave an F_2 progeny of 9 red and 7 white flowered plants. Single crosses between the red flowered variety and the two *different* white flowered varieties showed that the gene for red colour was dominant over the gene for each of the two white varieties.

The cross between the two white varieties can be explained by assuming two genes for red colour which must be present *together*, *i.e.,* must act in a complementary way to each other. Thus each gene independently contributes something different but essential for synthesis of red pigment. If one of the two genes for red colour is absent, the result is a white flower. This explanation can be verified by making a checkerboard.

The inheritance of the colour of aleurone layer in corn also demonstrates interaction of complementary genes. The outermost layers of endosperm in the maturing corn kernels become modified into a specialised *aleurone* tissue, so named because the cells have rich deposits of aleurone grains. In corn the aleurone layer is coloured due to anthocyanin pigments in the cells, and is controlled by complementary effect of two genes.

DUPLICATE GENES

When two or more genes have the same effect on a given trait, they are referred to as *duplicate* genes. In maize the gene for yellow endosperm is dominant over white endosperm. A pure breeding yellow endosperm plant when crossed to a white endosperm plant yields yellow endosperm in F_1. On self pollination of F_1 hybrids an F_2 generation of 15 yellow and 1 white endosperm is obtained. The yellow endosperm results from two independent dominant genes Y_1 and Y_2. When any one of these two dominant genes or both together are present, yellow endosperm is produced. When only recessive alleles are present in the homozygous condition ($y_1y_1y_2y_2$) it forms white endosperm. Thus the dominant genes Y_1 and Y_2 have an identical effect on endosperm colour and are consequently termed *duplicate* genes or *isogenes*.

In human beings 3 different genes can produce up to 12 similar lactic dehydrogenase enzymes called *isozymes* or *isoenzymes*. Lactic dehydrogenase consists of four polypeptide chains each of which is coded for by two different genes *A* and *B*. A third gene *C* codes for yet another polypeptide chain that is present in lactic dehydrogenase found in male germ cells.

POLYMERIC GENES

In *Cucurbita pepo* (summer squash) fruit shape could be spherical or cylindrical. The spherical fruit shape is dominant over cylindrical and is controlled by two independent genes. Thus there are two different varieties of spherical fruit plants. When two such genetically different spherical fruit plants are crossed, the F_1 progeny shows a novel fruit shape—discoid. Self pollination of F_1 discoid fruit plants produces an F_2 generation with all the three fruit shapes, *i.e.,* discoid, spherical and cylindrical in the proportion 9 : 6 : 1.

P:	Spherical	×		Spherical
	$S_1S_1s_2s_2$			$s_1s_1S_2S_2$
F_1:		Discoid	×	self
		$S_1s_1S_2s_2$		
F_2:	9 Discoid : 6 Spherical : 1 Cylindrical			

Out of the six spherical fruit plants of F_2 generation, three plants belong to one variety and have the dominant gene S_1. The other three spherical fruit plants belong to a genetically different variety with another dominant gene S_2. Due to an additive effect of genes S_1 and S_2 (also designated as *polymeric* effect), the discoid shape is produced.

MODIFYING GENES

As more and more instances of gene interaction were discovered, the earlier notion that one gene controls one phenotype independently of other genes had to be discarded. In reality an observed phenotype is the result of many complex processes within an organism. It is reasonable then that many genes should be involved in the final expression of a trait. There is a large group of genes that come under a general heading or *modifiers* which influence the activity of other genes and change their phenotypic effects. The modifying effect may be quantitative so that the expression of a phenotype is either enhanced or suppressed.

There is a recessive suppressor gene (*su*) in *Drosophila* which suppresses the effect of a mutant gene for hairy wing (*Hw*) so that even homozygous flies (*HwHw*) fail to develop hair on their wings. The same gene (it is also known as *su-Hw*) reduces the expression of a few other mutant phenotypes such as the ones caused by genes for *interrupted* wing veins and *forked* bristles. Another suppressor gene (*su-S*) in *Drosophila* is restricted in its action so that it reduces the expression of only one dominant gene which controls star eye shape.

In human beings the occurrence of minor brachydactyly (a form of brachydactyly in which only the index finger is of shorter length) in Norwegian families is due to a dominant gene *B*. There is a modifier gene *M* which modifies the effect of gene *B* to produce variable phenotypes. Thus in individuals with both genes *B* and *M*, the index finger is very much shortened. Persons having gene *B* and recessive alleles of the modifier gene (*mm*) show only a slight shortening of the same finger. Modifier genes by themselves do not seem to produce a visible phenotype.

LETHAL GENES

So far we have discussed modifications in Mendelian ratios caused by interaction of genes. Lethal genes can also alter the basic 9 : 3 : 3 : 1 ratio and lead to death of an organism.

Dominant Lethals

The yellow body colour in mice is dominant over brown, but the yellow mice are never true breeding. When yellow mice are inbred, the progeny consists of yellow and brown mice in the ratio 2 : 1 which does not fit any of the Mendelian expectations. Moreover, the litter size after inbreeding is smaller by one-fourth as compared to litter size resulting from a cross between yellow and brown. When yellow mice were backcrossed to true breeding brown mice, only heterozygous yellow mice were obtained.

Yellow	×	Yellow	Yellow	×	Brown
Yy	↓	*Yy*	*Yy*	↓	*yy*
2 Yellow	:	1 Brown	1 Yellow	:	1 Brown
Yy		*yy*	*Yy*		*yy*

Why were homozygous yellow mice never born? The answer came from a French geneticist L. Cuènot. He sacrificed *Yy* pregnant females after inbreeding and examined the embryos to determine if death occurred in embryonic stages or not. Indeed, one-fourth of the embryos were observed to die in late stages of development. Thus only heterozygous yellow and brown

mice in the ratio 2 : 1 were being born. The ratio 1 : 2 : 1 expected when a cross between two heterozygotes is made was never obtained proving the lethal expression of the homozygous yellow gene.

The brachyury gene (*I*) in mouse is lethal in the homozygous state, and when heterozygous the animal survives, but with a short tail. Embryos homozygous for brachyury show complete absence of notochord, a few abnormalities and die in *utero*. When two short tail mice heterozygous for brachyury are crossed, the viable offspring produced show a phenotypic ratio of 2 short tail : 1 normal tail (Fig. 2.4). In quite a few plants including maize, soybean and *Antirrhinum* (snapdragon) there is a dominant lethal gene which interferes with the process of photosynthesis and chlorophyll is not synthesised. Young seedlings which emerge from a seed carrying the homozygous dominant gene are yellow and die at a very young stage due to starvation. The heterozygous seedlings are light green in colour and able to survive.

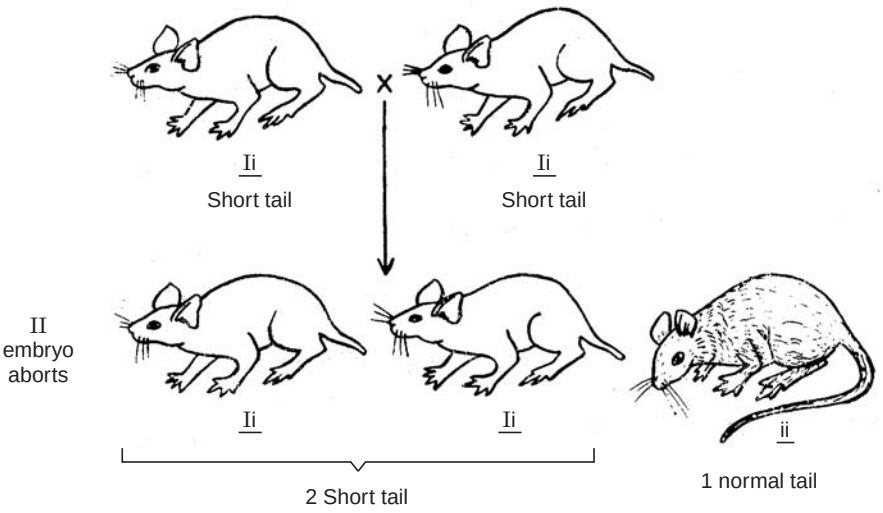


Fig. 2.4 Inheritance of the lethal gene for brachyury in mouse.

In all the cases of dominant lethal genes described above, the homozygous state of the gene leads to early death, whereas the heterozygote is viable. Perhaps the most serious effect a gene can have is to cause death even in the heterozygous state. The gene causing Huntington's chorea in man expresses itself when a single dominant allele is present. Whether homozygous or heterozygous, the phenotype of the disease becomes visible at middle age, usually after forty years. The individual suffers from muscular failure, mental retardation and finally death. Since the onset of Huntington's chorea is much after the start of the reproductive period, the gene can be transmitted to the next generation of offspring.

Another dominant gene which causes epiloia in human beings leads to death in early stages of life even in the heterozygous condition due to severe mental defects, tumours and abnormal skin growths. Dominant lethal genes which express lethality at an early stage in life are not detectable in the population.

Recessive Lethality

The recessive lethal gene remains unnoticed in the population because it does not produce a visible phenotype in the heterozygous state. In fact it may be transmitted through heterozygous

carriers for many generations without being detected. Consequently, a larger number of recessive lethal genes are known as compared to the dominant lethals.

There is a recessive lethal gene in man which causes death of newborn infants by producing internal adhesions of the lungs. A foetus homozygous for this gene completes its embryonic development with the help of oxygen supplied by the maternal blood. But death occurs soon after birth when the lungs fail to function normally. Such a recessive lethal gene is carried in heterozygous individuals without producing harmful effects. It is detected only when two heterozygous persons marry and about one-fourth of their children die after birth, as they receive both recessive alleles from their parents.

Another recessive lethal gene in humans known as Tay Sachs disease causes death of young children. Individuals homozygous for this gene lack one of the enzymes needed for the normal metabolism of fatty substances. The phenotype of the disease becomes visible after the first one year when fatty substances accumulate in the nerve sheaths. The transmission of nerve impulses becomes affected leading to loss of muscular control and mental deficiency. Within a few years the individual dies.

In humans there is a good chance of expression of recessive lethal genes in offspring of first cousin marriages. The single allele of the gene may have been present in the normal ancestors. It is only when the two alleles combine in the offspring of closely related persons that lethality is expressed.

In mouse hydrocephaly is due to a recessive lethal gene. During embryonal development there is abnormal growth of cartilage. This leads to irregularities in formation of skull and brain and excessive accumulation of cerebrospinal fluid. Embryos carrying the homozygous gene do not survive. The heterozygotes are phenotypically normal.

There is a recessive lethal gene in the beef producing cattle (dexter) in England. Dexter is the heterozygous breed which is highly prized for the larger quantity of beef meat it can produce. The common breed of cattle known as Kerry has two homozygous dominant genes, is normal like the dexter but produces less meat. When two dexters are crossed the progeny consists of 1 kerry : 2 dexters : 1 bulldog. The bulldog calf carries two recessive lethal genes, has very short legs, a few abnormalities and dies soon after birth.

Sex-Linked Lethals

This is a system in which the lethal gene is carried on the sex chromosome, usually *X*. In *Drosophila*, sex-linked recessive lethals are frequently employed to detect mutations. A recessive lethal gene carried on the *X* chromosome is especially important in the hemizygous male individuals because it can express lethality when only a single allele is present. The presence of a lethal *X*-linked gene can also alter the sex ratio so that more females are born instead of the expected ratio of 1 female : 1 male. Thus a female carrying a recessive lethal gene will produce a progeny in which one-half of the male offspring would not be viable. The disturbance in sex ratio is clearly visible in organisms like *Drosophila* which produce a large progeny. In human beings, existence of sex-linked lethal genes is suspected in those families where female births occur far more frequently than male births.

In human beings lethal effects among the progeny may be caused accidentally by radiation (*X*-ray) treatment of the reproductive organs of the parents. According to a study by R. Turpin in France, when women receive *X*-ray exposures in the pelvic region for abdominal ailments, recessive lethal mutations are induced in the *X* chromosome present in the ovum. Such a woman produces more females and very few males in the progeny. If the male parent is exposed to *X*-rays and dominant lethal mutations are induced on his *X* chromosome, there will be more

boys in the progeny and few females. This is because the single *X* chromosome is passed to the daughters resulting in their death.

Muscular dystrophy (duchenne type) is due to an *X*-linked recessive gene which shows a visible phenotype many years after birth. Boys having this gene are normal for about 10 years after which there is failure of muscular control and death results.

Conditional Lethals

Sometimes an organism lives normally under one set of conditions, but when certain changes are introduced in its environment, lethality results. One of the first conditional lethals known was recognised by Dobzhansky in *Drosophila pseudoobscura*. The flies live normally at a temperature of 16.5°C, but at 25.5°C the flies die. Similarly in the wasp *Bracon hebetor* the mutant gene which produces kindey-eyes at lower temperatures expresses lethality at a temperature of 30°C.

A number of conditional lethals have been described in *Drosophila melanogaster* by Suzuki in 1970. He has indicated that certain mutant strains became lethal when they were exposed to high temperatures *only* during the late larval stages. This is called the temperature-sensitive stage. If the larvae are kept at low temperatures during the specific temperature-sensitive stage, the flies born can live normally even at high temperatures throughout their life cycle. Perhaps a specific gene product—an enzyme or a protein becomes altered, causing death if the larvae are exposed to high temperature during the critical period. In fact with this perspective, conditional lethals are being studied extensively in micro-organisms for analysing genes, enzymes and proteins.

In poultry there is a recessive gene which causes feathers to break off. Chickens homozygous for this gene become devoid of feathers but are able to live normally if they are kept in a relatively warm environment. But if the temperature falls below the optimum, the chickens die due to lack of insulation provided by normal feathers.

Conditional lethals have been well studied in some haploid organisms such as yeasts, *Neurospora* and others. It is easy to study lethal genes in haploid organisms because the presence of even a single allele results in lethality. The wild type *Neurospora* is able to grow on a medium deficient in the amino acid arginine because it produces all the necessary enzymes required for synthesis of arginine from sugar and ammonia. But a mutant strain of *Neurospora* will not be able to grow on the same medium. A strain of yeast that grows normally on a glucose medium can show lethal effects if grown on a medium containing galactose. The mutant gene therefore, acts as a conditional lethal.

Early and Late Acting Lethals

The earliest stage at which lethal genes can act is evident from studies on mutations in gametes. Normal gametes are more viable and have better chances of effecting fertilisation and producing zygotes. The lethal genes are eventually lost with the death of unfertilised gametes. Such genes are referred to as gametic lethals. The phenomenon by which a certain class of gametes is specifically inhibited from taking part in fertilisation has been termed *meiotic drive* by Sandler and Novitski. There is a gene called segregation distorter (*SD*) present on the second chromosome of *Drosophila*. The dominant allele of this gene does not allow gametes to participate in fertilisation. Thus only gametes bearing the recessive allele (*sd*) are able to

fertilise eggs and produce viable zygotes. Since this also results in distortion of typical Mendelian ratios, the name segregation distorter has been given to this gene.

There are lethal genes that act after zygote formation resulting in embryonic death. In experimental animals such as the mouse, it is possible to determine this lethal effect by sacrificing impregnated females and analysing the dead embryos. If, however, death occurs very early in embryonal development, this technique does not succeed because actual observation of aborted embryos is not possible. There is a gene which exerts a killing effect by preventing normal cleavage of the zygote. Such a gene is called a zygotic lethal.

In human beings the gene described for causing adhesions in lungs expresses lethality soon after birth when the lungs of the newborn infant fail to function normally. In plants most lethal genes are known to act during or after seed germination.

Among late acting lethal genes, some clear cut examples can be cited from human diseases. Genes causing muscular dystrophy (duchenne type) and Tay Sachs disease cause death before or in the second decade of life, before the onset of reproduction. Huntington's chorea on the other hand is fatal when the person is middle-aged, and the gene may have already been passed on to future generations.

QUESTIONS

1. You have been asked to perform a cross which produces an F_2 progeny containing homozygotes and heterozygotes that can be distinguished phenotypically. What material and characters would you select to make such a cross ?
2. When a pure red-flowered snapdragon is crossed with a pure white-flowered one, the F_1 is intermediate—pink. When a pure red-haired short horn cow is mated to a pure white-haired type, the F_1 is reddish gray (roan). But the genetic basis for the F_1 phenotype is different in the two cases. Explain.
3. What conclusions could you draw about parents whose offspring had phenotypes in the following proportions:
   (*a*) 13 : 3 (*b*) 15 : 1 (*c*) 9 : 3 : 4
4. In maize plants alleles for dwarfism and striped leaves are recessive to alleles for normal height and green leaves. A plant heterozygous for both recessive alleles is self pollinated and 320 seeds are collected. How many of these seeds could produce plants—
   (*a*) of normal size
   (*b*) with striped leaves
   (*c*) with green leaves and dwarf size
   (*d*) with striped leaves and normal size
5. Four hens with walnut combs are each mated with a single combed hen. One mating produces only chickens with walnut comb; the second produces both walnuts and pea comb; the third only walnuts; the fourth produces rose, pea and single combs. What are the genotypes of the four walnut hens used as parents?
6. A rose combed hen is crossed with a walnut. The progeny consists of 3 walnut, 3 rose, one pea and one single. What are the genotypes of the parents?

7. In sweet peas the genes *C* and *P* when present together produce purple flowers. But when either *C* or *P* is present alone it produces white flowers. What flower colour would the progeny of the following crosses have:
(*a*) *Ccpp* × *ccPP* (*b*) *ccPp* × *CcPp* (*c*) *CCPp* × *ccPP*

SELECTED READINGS

Bateson, E.R. *et al.,* 1905. Reports to the Evolution Committee. Royal Society, Vol. II. Harrison and Sons, London.

Cuènot, L. 1903. L'heredite de la pigmentation cheq les souries. *Arch. Zool. Exper et Genet* **2**: 33.

Edwards, J.H. 1960. The Simulation of Mendelism. *Acta Genetica*, Basel **10**:63.

Hadorn, E. 1961. Developmental Genetics and Lethal Factors. Wiley, New York.

Hexter, W. and Yost, H.T. 1976. The Science of Genetics. Prentice-Hall, N.J.

Jenkins, J.B. 1979. Genetics. Houghton Mifflin Co. Boston.

Lawler, S.D. and Renwick, J.H. 1959. Blood Group and Genetic Linkage. Brit. Med. Bull., **15**:145.

McKusick, V.A. 1969. Human Genetics. Prentice-Hall, N.J.

Punnett, R.C. (ed.) 1928. The Scientific Papers of William Bateson. 2 Volumes. Cambridge University Press, London.

Swanson, C.P. 1957. Cytology and Cytogenetics. Prentice-Hall, New York.

3

Environment and Gene Expression

It is common observation that the phenotype shows modifications in response to the environment. The question arises as to whether the genotype also changes in response to an internal or external environment or not? Further, is the genotypic change, if any, inherited and transmitted to future generations? The answers to these questions have come mostly from experimental work on organisms that were made to live in different environments. It has been found that the environment plays a role in changing the form and degree of expression of some genes resulting in a corresponding change in the phenotype.

PENETRANCE AND EXPRESSIVITY

The presence of a gene does not always bear an absolute relationship with the appearance or absence of a trait. In the *ABO* blood group system, the genes are expressed in an absolute way. But in many other instances the gene is expressed in a variable manner, *i.e.* the visible phenotype shows varying intensities. The terms *penetrance* and *expressivity* are used to describe variable gene expression. *Penetrance* is the proportion of individuals that show an expected phenotype. When a gene is completely penetrant it is always expressed; when incompletely penetrant, the gene is expressed in some individuals, not in others, the proportions depending upon the degree of penetrance. For example in the recessive traits which Mendel studied, the phenotype was expressed fully when the gene was in homozygous condition; this is due to 100 per cent penetrance. Suppose instead that in a hypothetical cross, only 60 per cent of individuals show the expected trait when all 100 are carrying the gene; we say that in this case penetrance is 60 per cent. *Expressivity* is the degree to which a gene is expressed in the same or in different individuals. Thus the gene for lobe eye in *Drosophila* may show a complete range of phenotypic expression in different individuals. Some flies may have a normal sized eye, in others the eye is smaller, in still others the eye is absent.

TEMPERATURE

The earliest studies related to the effect of temperature on genetic constitution were done on the Himalayan breed of rabbits and Siamese cats. Coat colour in rabbits is controlled by multiple alleles of a gene. When one of the recessive alleles c^h is present in the homozygous condition ($c^h c^h$), the Himalayan coat colour results. Such a rabbit is a mosaic with white fur all over the body except the nose, paws, ears and tail which are black (Fig. 5.1). The black extremities are the portions which have lower temperature (less than 34°C) than the rest of the body. If the extremities are exposed to higher temperature artificially, the new hair which starts growing is white. Similarly, if some portion of the body bearing albino fur is artificially kept at a lower temperature, the new hair formed is black. These observations explain the temperature sensitive behaviour of the allele (c^h) which controls Himalayan trait in the homozygous state. The allele codes for an enzyme used in pigment formation which is temperature sensitive and is inactivated by temeratures above 34°C resulting in albino phenotype; if temperature is lower the same alleles promote synthesis of pigment and the phenotype is black. When rabbits of this genotype are grown at cold temperatures, they are completely black. The Siamese cat shows the same pigmentation pattern as the Himalayan rabbit due to the presence of similar type of temperature-sensitive allele.

In *Drosophila* temperature changes the penetrance of the gene known as *tetraptera* which controls wing development. At 25°C the gene has 35 per cent penetrance so that the corresponding number of flies develop wings whereas 65 per cent do not. At 17°C penetrance is much reduced so that only one per cent of flies show the winged phenotype. The recessive gene *vg*/*vg* which produces vestigeal wings in *Drosophila* is also influenced by temperature. At 32°F the wings are feebly developed and extend very little from the body (Fig. 3.1). At 40°F the wings are better developed and have some venation. At 88°F wings are well developed with conspicuous venation.

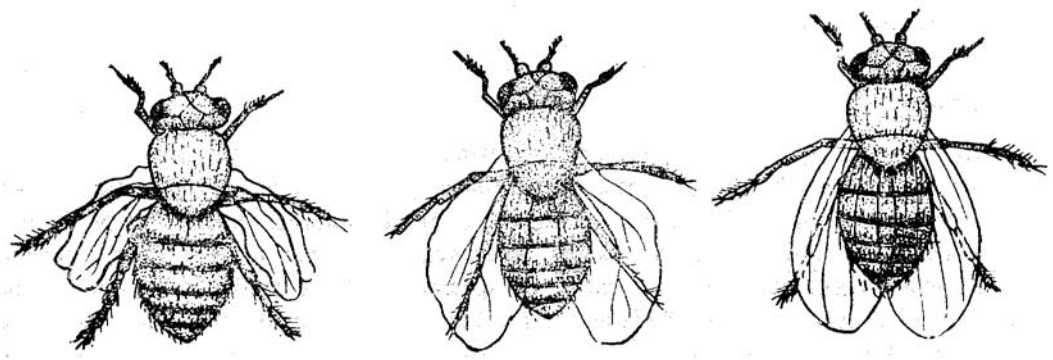


Fig. 3.1 Influence of temperature on the expression of the gene for vestigial wings (*vg*) in *Drosophila.*

Some temperature-sensitive mutations are exhibited in bacteriophages. In general the temperature at which normal phenotypes are produced is referred to as *permissive temperature:* that which produces mutant phenotypes is called *restrictive temperature.* Some lethal mutations in viruses and in *Drosophila* are *temperature sensitive.* Among plants, colour of flower in primrose changes from red to white when temperature is raised above 86°F.

LIGHT

There is a gene in maize plants which controls anthocyanin pigment formation. When ears of plants carrying the homozygous gene are exposed to sunlight by removing the green leafy coverings on the young cobs, the kernels become bright red in color ("sunred"). If however, the blue violet rays of the light spectrum are prevented from reaching ears of maize plants (by wrapping red cellophane paper around them so that only red rays penetrate the cells) the sunred phenotype is not visible. In this case sunlight interferes with one or more chemical reactions leading to pigment formation. The reddish freckles on the sensitive skin of white skinned human races are also caused by sunlight in a similar way.

In human beings a skin cancer known as *xeroderma pigmentosum* is caused by a homozygous recessive gene. The skin becomes extremely sensitive to sunlight so that even a minor exposure to faint light gives rise to pigmented spots on the facial skin. The spots can become cancerous and if they spread to other parts of the body, death results. If an individual homozygous for the recessive gene is not exposed to light, the gene is not able to express itself.

Environment and Sex Determination

The marine worm *Bonellia* demonstrates the effect of environment on sex. In this sexually dimorphic organism the female is very large, about 10 cm in length; the male is 3 mm long and lives inside the cloaca of the female. If the free swimming larvae that have arisen from fertilised eggs remain in the sea bed away from the females, they develop into female worms. But if females are available, the larva settles on the female proboscis, draws nourishment from it, and develops into a male. Of the many experiments performed with *Bonellia,* one is most interesting and relevant here. If *Bonellia* are raised in the laboratory in a tank containing artificial sea water, the free-swimming larvae settle down at the bottom of the tank and develop into females. But if the artificial sea water is agitated by some mechanical device, the larvae develop into males!

PHENOCOPIES

Depending upon the extent to which the environment influences the genotype, the changes in the phenotype may be subtle or dramatic. Sometimes the phenotype becomes altered by the environment in such a way that the new phenotype resembles another phenotype produced by known genes. The induced phenotype is not inherited and is called a *phenocopy*. In many instances phenocopies result from application of specific treatments like radiation, chemicals poisons, temperature shocks etc. The Himalayan rabbit described earlier in this chapter develops a coat that is all black if the rabbit is made to live in a cold environment. The Himalayan rabbit is a *phenocopy* of the genetically black rabbit. If both rabbits live together at moderately high temperature, the Himalayan rabbit has a phenotype very different from the genetically black rabbit.

One of the most striking examples of phenocopies could be observed in what were known as *thalidomide* babies in the early 1960's. A number of deformed children were born in West Germany and Great Britain to mothers who had taken the tranquilising drug *thalidomide* in their sixth week of pregnancy. The abnormal children showed deformities in limbs; some had one, two or three limbs, others had no limbs at all. The abnormalities showed a great resemblance to another phenotype known as *phocomelia* caused by a recessive gene.

Diabetes mellitus is a heritable human trait associated with reduced amounts of the hormone insulin that is secreted by the pancreas. In the presence of insulin glucose is absorbed by the cell membranes. When the hormone is not produced in sufficient quantity, the unabsorbed glucose passes into the blood and urine. The exact mode of inheritance of diabetes is not properly understood. There are different types of diabetes arising from different causes; it therefore seems likely that there are several gene pairs controlling the trait. On the other hand the study of a pair of genetically identical twins, one of whom had diabetes the other not, indicates that the condition is due to a recessive gene with low penetrance. If proper doses of insulin are administered to a diabetic person, he reverts to the normal phenotype. In other words, control of diabetes produces a phenocopy of the normal individual.

There are many other examples in human beings where, by giving drugs, the mutant genotype produces a phenocopy of the normal phenotype. In haemophiliac patients, a specific protein required for blood clotting is either defective or deficient. If however, an antihaemophiliac factor isolated from humans is injected into a patient, a phenocopy of the normal individual results. Similarly, if thyroxine is administered to a child whose thyroid gland does not secrete this substance in adequate quantities, the normal phenotype is produced.

The creeper trait in chickens is observed sometimes in domestic fowl when the newly hatched chickens have the legs drawn up under the body. The affected chicken is not able to walk but creeps along the ground. The creeper trait (Fig. 3.2) is expressed by the heterozygous condition of a dominant gene which is lethal when homozygous. Creeper chickens can also be produced if incubating eggs of normal fowls are treated with boric acid or insulin. Such induced creepers are phenocopies of the genetically controlled heterozygous creeper chickens.

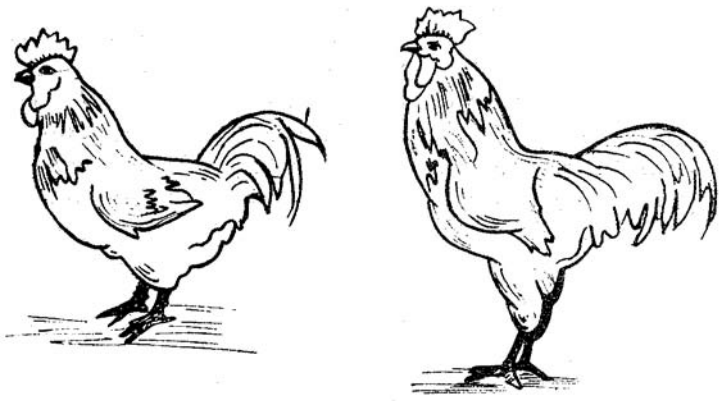


Fig. 3.2 Chicken expressing creeper trait (left) compared with a normal chicken (right).

Due to a recessive gene, maize plants become dwarfed, because they are deficient in the plant growth hormone known as gibberellic acid. But if the hormone is supplied to the dwarf plants they grow to normal height producing phenocopies of normal plants.

ENVIRONMENTAL EFFECTS AND TWIN STUDIES

In human beings it is not possible to perform controlled breeding experiments. Twin studies are perhaps the best way of determining as to whether the observed differences between individuals are due to heredity.

Twins are of two types—*monozygotic* or identical twins that arise from a single fertilised egg and have idential genotypes; *dizygotic* or fraternal twins which arise from two fertilised eggs and are therefore no more genetically alike than siblings (brothers and sisters). The correct identification of twin types is difficult and unreliable unless done by a physician. For assessing the role of environment in heredity, the percentage of *concordance* (both twins showing identical phenotype) and *discordance* (different phenotypes) for a given trait must be determined for twins of both types. In general if concordance percentage for a trait is high in the case of monozygotic twins, and much less in dizygotic twins, one can conclude that heredity has played a role. If the concordance rate is similar in monozygotic and dizygotic twins, it suggests that the environment is determining the phenotype. From studies of a large number of twins it has been found that measles (caused by infection with *Rubelia* virus in early pregnancy) is largely controlled by the environment. On the other hand conditions like diabetes, schizophrenia, Rickets and tuberculosis appear to be controlled by the genotype. Another useful aspect of twin studies is to determine the effects of different environments on identical genotypes by analysing those rare cases of monozygotic twins that have been separated from birth and reared apart. However in absence of adequate data it is not possible to conclude much on this aspect as yet.

HUMAN INTELLIGENCE

A number of studies have been done to determine how much of human intelligence and I.Q. are controlled by the genotype and how much by the environment. Both clarifications and complications have been revealed. The differences in intelligence among different racial groups have been extensively studied by Arthur Jensen in 1969. This work is highly controversial and has been much debated. Nevertheless, it is generally agreed that intelligence is under the control of several gene pairs interacting with the environment. From twin studies it has been further estimated that about one-half to three-fourths of human intelligence is determined genetically; the remainder is controlled by the environment.

DRUG RESISTANCE

It is fairly well established that mosquitoes develop resistance to *DDT* and other insecticides used for eradicating malaria. The resistance develops due to change in the genotype in response to the environment, and is inherited. Similar resistance is reported also in insects which carry the causal agent for some other diseases like dengue fever, yellow fever, filariasis and river blindness. A number of pests which are harmful to major crops such as rice, maize, cotton, wheat and potato are also known to have become resistant to a wide range of insecticides.

QUESTIONS

1. Describe two examples where temperature could alter the expression of a genotype.
2. Explain the following terms:
   (*a*) permissive temperature; (*b*) phenocopy;
   (*c*) penetrance; (*d*) concordance and discordance.

3. Name two human conditions in which administration of drugs can change the phenotype.
4. Twin studies are extremely useful for study of genes in relation to environment. Explain.

SELECTED READINGS

Farnsworth, M.W. 1978. Genetics. Harper & Row, New York.

Herskowitz, I.H. 1977. Principles of Genetics. Collier Macmillan, New York.

Jenkins, J.B. 1979. Genetics. Houghton Mifflin Co. Boston.

Osborn, R.H. and Degeorge, F.V. 1959. Genetic Basis of Morphological Variation. Harvard Univ. Press, Mass.

Sang, J.H. 1963. Penetrance, Expressivity and Thresholds. J. Heredity **54**: 143.

Searle, A.G. 1968. Comparative Genetics of Coat Colour in Mammals. Academic Press, New York.

4

Quantitative or Polygenic Inheritance

The classical experiments in Mendelian inheritance demonstrated transmission of obviously well-marked, distinct pairs of contrasting characters (tall vs. dwarf; yellow vs. green cotyledons, etc.). Moreover, the two alternative expressions of a phenotype were controlled by a single gene which strictly obeyed the rules of dominance to produce visible qualitative differences between phenotypes. Such traits are said to be *qualitative* and constitute what are known as *discontinuous variations* in a population. However, even Mendel's experiments with pea plants showed small, indistinct, quantitative differences so that all tall plants were not equally tall, nor were all dwarf plants equally short. Instead some degrees of tallness and shortness were noticeable. Quantitative characters like height, weight and intelligence in man, meat in animals, milk in cattle, all show these variations. These small *quantitative variations* could not be accounted for by distinct genes because they produced gradations in phenotypes that appeared to be continuous. These small differences were said to result from fluctuations in the environment.

It was at this time that Darwin's Theory of Evolution which emphasised inheritance of continuous variations was gradually being overthrown by ideas of geneticists like Bateson and DeVries. The latter proposed that continuous characters were not produced genetically and were not heritable, but only large discontinuous variations were inherited.

At the same time Francis Galton, a cousin of Charles Darwin, and his followers consisting of geneticists and biometricians put forward strong views about the heritable nature of continuous quantitative traits in humans such as height and intelligence. Human beings show all degrees of tallness and shortness between the two extremes. The biometricians used the terms "metrical" characteristics for such traits which showed small, measurable differences. By analysing the data statistically they could demonstrate the inheritance of gradations in height and intelligence between relatives in human families. But they could not account for the role of genes as separate units.

At the beginning of the 20th Century therefore existed two groups with opposing views. One led by Bateson supporting Mendelian views and emphasising the heritable nature of qualitative discontinuous traits; the second group led by followers of Galton who contended that heritable variation is basically quantitative and continuous.

The conflict was partially resolved in 1903 when a Danish geneticist W.L. Johannsen analysed a quantitative seed character in *Phaseolus vulgaris* (bean) and showed that continuous variation was due to both heredity and environment. He established 19 pure lines on the basis of seed weight that ranged from 15 centigrams in the light varieties to 90 centigrams in the heavy varieties. Johannsen found that the mean seed weight did not differ significantly in the succeeding generations. Although his pure lines were homozygous and true breeding, there were some variations in seed weight within a pure line which he attributed to the environment. It was Johannsen who coined the terms genotype, phenotype and gene in 1903, on the basis of his work on pure lines. In 1906 Yule suggested that continuous quantitative variation may be due to several genes each with a small effect. However, the controversy was finally resolved by the experiments of the Swedish geneticist Nilsson-Ehle in 1909 who demonstrated that in wheat the quantitative effect was in fact due to many genes. The hypothesis was further confirmed by East in 1910 from his studies on inheritance of corolla length in tobacco.

INHERITANCE OF KERNEL COLOUR IN WHEAT

Nilsson-Ehle performed many crosses between varieties of wheat having red seeds and those having white seeds. In all crosses except one the F_2 generation gave a ratio of 3 red to 1 white. In the one exceptional cross he used a very old red variety from the north of Sweden. The F_1 plants produced seeds intermediate in colour between the two parents, and the F_2 generation segregated in the ratio of 63 red : 1 white. The noteworthy feature of this experiment was the variation in the intensity of the red pigment in wheat grains produced by F_3 plants.

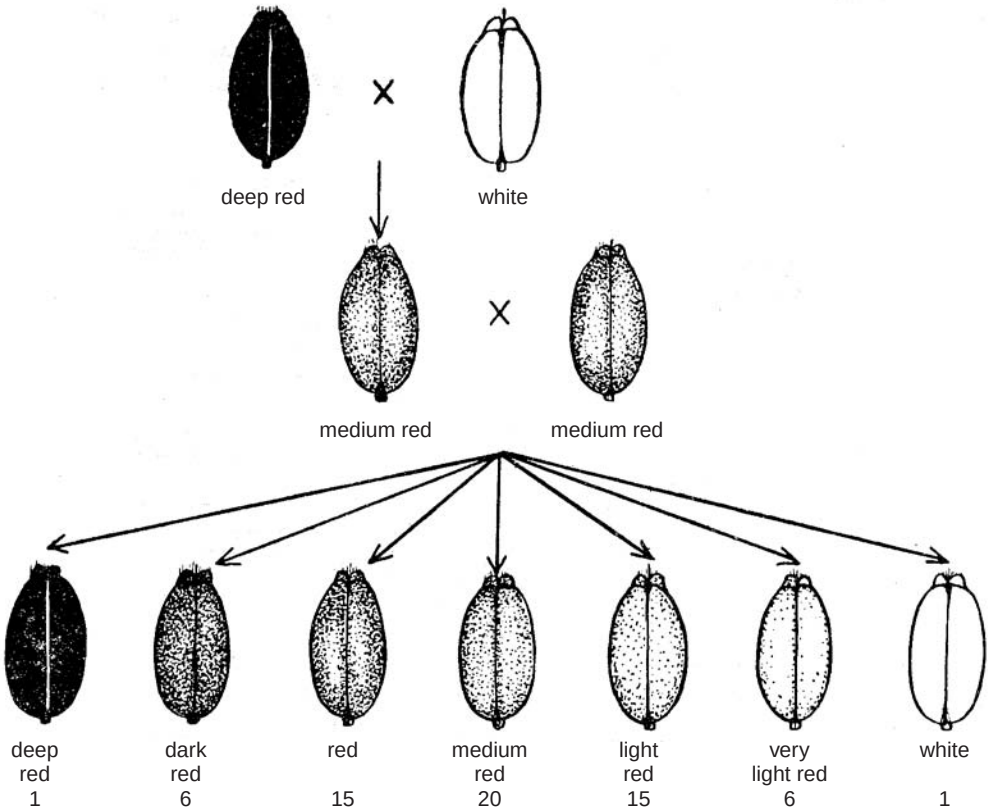


Fig. 4.1 Cross made by Nilsson-Ehle between red and white varieties of wheat.

There were all gradations from the deep red of one parent to pure white of the other parent so that the plants could be divided into 7 different colour classes in the ratio of 1 : 6 : 15 : 20 : 15 : 6 : 1. Nilsson-Ehle could distinguish 6 phenotypic classes with varying intensities of red as follows: 1 deep red (like one parent), 6 dark red, 15 red, 20 medium red, 15 light red, and 6 very light red (Fig. 4.1). Only one out of 64 plants produced completely white kernels and another one of the 64 had red kernels identical to the parents in the first cross.

Nilsson-Ehle postulated three pairs of genes controlling grain colour in wheat with genes for red (*ABC*) dominant over genes for white (*abc*). It also appeared that all alleles contributed equally in the production or absence of red pigment. Each of the three gene pairs when considered singly in crosses segregated in the expected Mendelian fashion producing F_2 progeny of 3 red and 1 white. When the genes were considered two at a time, F_2 segregated in the ratio 15 red to 1 white. All the three genes considered together produced an F_2 ratio of 63 red to 1 white, and segregated in a manner typical for a Mendelian trihybrid cross.

The cross is best explained by assuming that each red gene contributes a small degree of red colour to the wheat kernel resulting in the range of red phenotypes actually observed. The proportion of plants showing deep red kernels (*AABBCC*) like one parent, and white (*aabbcc*) like the other parent is very small as expected.

The intermediate phenotypes appear with greater frequency. Moreover, the intermediate degrees of red appear to suggest some kind of blending due to the continuous range of phenotypes. Yet it can be explained by Mendelism if we assume that many genes are controlling the trait. The term polygenes was used by Mather for genes with small effects on a character that supplement each other to produce noticeable quantitative changes. These are also called *additive* effects.

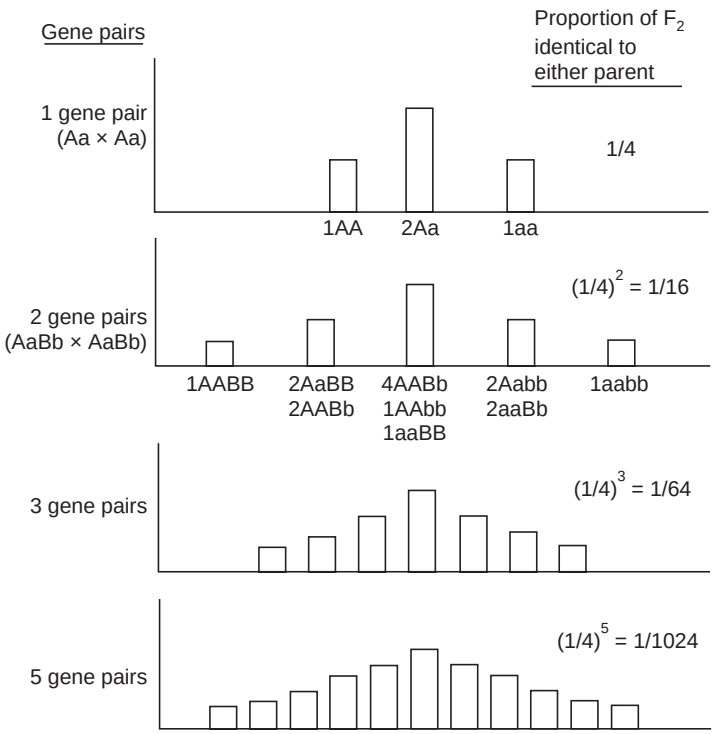


Fig. 4.2 Graphs showing relation of polygenes with F_2 phenotypes.

The small quantitative effects of polygenes can also result from environmental fluctuations. But the two can be distinguished by performing controlled breeding tests. The more difficult task is to determine the number of genes involved in a cross. This can be estimated from the frequency of F_2 genotypes which are identical to the parents. Continuous variations in quantitative characters are due to increased numbers of segregating gene pairs in the F_1 heterozygote. The resulting distribution of phenotypes in F_2 can be plotted on a graph. The parental phenotypes and the intermediate types can be read off from the graphs as shown in Fig. 4.2. However, as will be explained later, this method is applicable only when 2, 3 or 4 gene pairs are involved but not more.

The term polygene is often used interchangeably with multifactorial or multiple factor inheritance. But strictly speaking polygenes should refer to conditions determined by a large number of genes each with a small effect acting additively (Fraser, 1976). Multifactorial inheritance is determined by a combination of genetic and environmental factors. Mostly it is difficult to ascertain whether or not environmental factors are involved, or whether all the genes controlling a trait have small and additive effects.

COROLLA LENGTH IN TOBACCO

East in 1910, 1916 performed experiments on flower length in *Nicotiana longiflora*. He took two homozygous, pure breeding lines differing in flower length, one having 41 cm the other 93 cm for average length of corolla. The minor variations in flower length within a strain were apparently caused by the environment, but the major differences between the two strains were no doubt genetic. East crossed the two lines and obtained an F_1 progeny with flowers intermediate in length to both parents. The F_1 plants were presumed to be genetically uniform and the minor variations observed were attributed to the environment. On selfing F_1 plants the F_2 progeny showed a wide range of phenotypic variation. The F_2 plants were also selfed to get an F_3 progeny. Corolla lengths in F_3 could be classified into different groups; the mean corolla length in each group seemed to correspond to the plant used as F_2 parent. This indicates that differences in corolla lengths between the F_3 groups are genetically based. The variations within a group are caused by the environment.

For determining the number of genes involved in the cross the proportions of genotypes identical to the parents are estimated. If 2 pairs of genes are participating, the parental phenotypes produced by homozygous genotypes would be present in 1/16 of the progeny. With 3 pairs of genes only 1/64 individuals would be identical to the parents; in the case of four genes only 1/256 show the parental phenotypes. Thus if we observe the extreme phenotypes of the parents occurring in 1 out of 16, 1 out of 64 or 1 out of 256 offsprings, it would indicate that 2, 3 or 4 pairs of genes respectively are involved.

Beyond 4 genes it is difficult to recognise a parental phenotype in F_2 generation. This is due to the fact that with an increasing number of genes, the cross between two F_1 heterozygotes produces a very large number of different genotypes, so that the proportion of individuals homozygous for all the genes is very small and difficult to detect. For instance, if 5 gene pairs are involved there would be 1024 different gamete combinations and 243 different genotypes in F_2 (Fig. 4.2). If 10 gene pairs participate, they would produce more than 59000 different genotypes!

If we make the assumptions that polygenes assort independently, and that they produce identical additive effects then we can say that the work of Nilsson-Ehle and East could

demonstrate that quantitative traits follow the same rules of inheritance as qualitative traits. While making this statement, the influence of some genetic phenomena such as linkage, epistasis and dominance have also not been considered. The effect of environment is regarded as negligible.

SKIN COLOUR INHERITANCE IN MAN

Variations in human skin colour exist within a population, and to a much greater degree *between* different populations. The latter have been analysed by many workers and the races have been designated as white, black, yellow and red on the basis of skin colour. The Africans and the Europeans which show both extremes of black and white skin have been extensively studied. The children and grandchildren from marriages between negroes and whites give rise to hybrid groups showing quantitative variations in skin pigment. Investigations of this nature were conducted at the beginning of this century by Davenport and Danielson (1913) and later by Gates, Curt Stern and others.

In all the cases studied for marriages between whites and negroes the F_1 was found to be intermediate in skin colour to both parents and was called a mulatto. According to Davenport crosses between two mulattos produced shades of colour falling in five distinct classes. Davenport therefore considered two gene pairs to control skin colour inheritance in man. Later on Gates assigned 3 genes, and Stern 4 to 6 genes for this trait. Using the modern technique of reflectance spectrophotometry, Harrison and Owen in 1964 have estimated the number of genes for skin colour to be between 3 and 4.

Assuming that 4 genes determine skin colour, a child with white European skin would be expected in one birth out of every 256, and the same probability is true for negro skin colour.

TRANSGRESSIVE AND REGRESSIVE VARIATION

Punnett while studying inheritance of body size in two varieties of chickens crossed a large Golden Hamburg with a small Sebright Bantam. The F_1 chickens were intermediate in size and were allowed to inbreed. In the F_2 generation were surprisingly present some chickens larger than the Hamburgs, and some that were smaller than the Bantoms. Punnett explained these results by assuming four pairs of genes for body size. The parents used in the cross did not represent the extreme classes so that the Hamburgs had the genotypes *A A B B C C d d* and the Bantams *a a b b c c D D*. The F_1 heterozygotes (*A a B b C c D d*) were intermediate in size as expected. The F_2 chickens that turned out to be larger than the Hamburg parent were those that were homozygous for all the genes for large size (*A A B B C C D D*). The F_2 chickens smaller than the Bantam parent had the genotype *a a b b c c d d.* Such a phenomenon in quantitative inheritance by which some of the F_2 phenotypes exceed the parental extremes is known as *transgressive variation*.

A quantitative trait may show an entirely different phenomenon known as *regressive variation*. Due to this the phenotype of the offspring has a tendency to exhibit the mean of the population rather than the parental extremes. For example, human stature is a trait controlled by many genes and shows continuous variation. The children of two tall parents are sometimes shorter than either of them because their height tends to approach the mean of the population. For the same reason short parents can have children taller than themselves.

QUESTIONS

1. Fill in the gaps:
   (*a*) Quantitative traits are due to ..... variation in a population.
   (*b*) Nilsson-Ehle's cross between red and white varieties of wheat produced ..... phenotypic classes of red in F_2.
   (*c*) Experiments with pure lines of ..... demonstrated that continuous variation was due to both heredity and environment.
   (*d*) ..... variations formed the basis of Darwin's Theory of Evolution.
2. Write short notes on: (*a*) polygenes, (*b*) transgressive variation, (*c*) Galton, (*d*) qualitative trait, (*e*) Davenport.
3. Mention some quantitative characters in animals and humans that are likely to be controlled by polygenes.
4. Describe a method by which the number of genes controlling a character can be determined.

SELECTED READINGS

Davenport, C.B. 1913. Heredity of Skin Colour in Negro-white Crosses. Carnegie Inst. Washington Publ. No. 188.

East, E.M. 1910. A Mendelian Interpretation of Variation that is Apparently Continuous. *Amer. Nat.* 44 : 65.

Edwards, E.A. and Duntley, S.Q. 1939. Pigments and Color of Living Human Skin. *Amer. J. Anat.* 65 : 1.

Falconer, D.S. 1961. Introduction to Quantitative Genetics. Ronald Press, New York.

Fisher, R.A. 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. Roy. Soc. Edinb.* 52 : 399.

Fisher, R.A. *et al.,* 1932. The Genetical Interpretation of Statistics of the Third Degree in the Study of Quantitative Inheritance. *Genetics* 17 : 107.

Gates, R.R. 1949. Pedigrees of Negro Families. Blakiston, Philadelphia.

Gilbert, N. 1961. Polygene Analysis. *Genet. Res.* 2 : 96.

Harrison, G.A. and Owen, J.J.T. 1964. Studies on the Inheritance of Human Skin Colour. *Ann. Human Genetics* (Lond.) 28 : 27.

Johannsen, W. 1909. Elements der exacten Erblichkeitslehre. G. Fisher, Jena, East Germany.

Levine, L. 1971. Papers on Genetics, A Book of Readings, Toppan, Japan.

Mather, K. and Jinks, J.L. 1971. Biometrical Genetics. Chapman and Hall, London.

Milkman, R.D. 1960. The Genetic Basis of Natural Variation. 11. Analysis of a Polygenic System in *Drosophila Melanogaster. Genetics* 45 : 377.

Nilsson-Ehle, H. 1909. Kreuzungsuntersuchungen an hafter und Weizen. Lunds Univ. Aarskr. 2, 5, 2 : 122.

Punnett, R.C. 1923. Heredity in Poultry. Macmillan, New York.

Sunderland, E. 1956. Hair Colour Variation in the United Kingdom. *Ann. Human Genetics*, (Lond.) 20 : 312.

Weiner, J.S. 1951. A Spectrophotometer for Measurement of Skin Colour. *Man* 51 : 253.

5

Multiple Alleles

WHAT ARE MULTIPLE ALLELES?

After Mendel first advocated the existence of two factors for each character, it was demonstrated in many organisms that a gene consists of a pair of alleles. Each member of the pair of alleles is said to occupy an identical position or *locus* on each of the two homologous chromosomes in diploid cells of an organism. In Mendel's experiment the gene controlling height of pea plants has both its alleles designated either as T and T or T and t, or t and t. Since there are always only two alleles they can also be denoted as T^1 and T^2. Similarly the gene determining flower colour (R and r) can be denoted by alleles R^1 and R^2. Sometimes more than two alternative alleles or multiple alleles are present in different individuals of a population. When there are *multiple alleles*, a gene is denoted by more than two alleles such as T^1, T^2, T^3, T^4 and R^1, R^2, R^3, R^4, ... and so on. Now there are only two homologous chromosomes in a diploid cell, and at one particular site of a gene or locus, only one allele can be present. Therefore, in one diploid cell only *two* alleles are present at a particular locus. In other members of the population, due to two or more mutations, the same locus on two homologous chromosomes could have two *different* alleles. In this way it is possible to detect a number of alleles for one gene from their different expressions in different individuals. Such a system in which one gene has more than two allelic states at the same locus in different members of the population is known as a multiple allele system.

T.H. Morgan in 1910 described the first case of multiple alleles of a gene controlling eye colour in *Drosophila* during his studies on mutation. In a vial of flies with normal red eyes he found a fly with white eyes which had arisen due to a mutation in the gene which produces red colour in normal flies. By performing genetic experiments, the position of this gene was determined to be on the X chromosome, the exact location being 1.5 units from the left end of this chromosome. This location is identical to the position occupied by the gene which produces red colour in eyes. Later on some other flies were discovered with eye colour resembling the biological stain eosin. The gene for eosin colour was also found to be located at 1.5 units on the X chromosome. When crosses were made between eosin flies and red, and between eosin and white it turned out that the gene for eosin was an allele to both red and white genes. This proves that genes producing red, white and eosin eyes in different flies are all alleles of each other. In other words the gene controlling eye colour in *Drosophila* has multiple alleles. Later on a series of alleles producing eye pigments in different shades and intensities between red

and white were found; each shade has a different name such as wine, blood, coral, cherry, apricot, honey, pearl and ivory, and a few more. It also means that the different alleles become less and less efficient in producing the same kind of biochemical product. Sometimes phenotypes produced by different alleles are not markedly different so that some of them appear close to the normal red. Such alleles with similar effects are referred to as *isoalleles*. There are wild-type isoalleles for genes expressing the wild phenotype. The mutant gene for white eye colour in *Drosophila* is composed of a series of multiple isoalleles W^1, W^2, W^3 etc. These are known as *mutant isoalleles*.

The Himalayan rabbit is a classic example for illustrating multiple alleles, first studied by Sturtevant in 1913. The wild type rabbit has grey fur and is called agouti; rabbits with all white fur and pink eyes are albino; the Himalayan has white fur on the body, but its feet, tail, ears and tip of nose are black (Fig. 5.1). When crosses were made between agouti and albino, and between agouti and Himalayan, both albino and Himalayan were found to behave as recessive to agouti. When Himalayan and albino were crossed, F_1 were all Himalayan, and in F_2 3 Himalayan : 1 albino were produced. Clearly, Himalayan, agouti and albino all result from different alleles of the gene that controls fur or coat colour in rabbits. If C denotes the dominant allele that produces the wild agouti type of fur, then albino would have the genotype cc. The cross of agouti (CC) and Himalayan ($c^h\ c^h$) would give an F_2 consisting of 3 agouti (CC, Cc^h) and one Himalayan ($c^h\ c^h$). Similarly, the cross between Himalayan and albino (cc) would give an F_2 of 3 Himalayan (c^hc^h, c^hc) and one albino (cc). There is also a fourth allele for fur colour in rabbits called chinchilla (c^{ch}) which behaves like Himalayan.

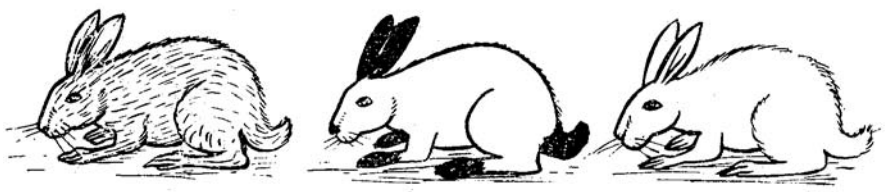


Fig. 5.1 The Himalayan rabbit.

SEXUAL INCOMPATIBILITY IN PLANTS

Sometimes bisexual plants produce functional male and female gametes, yet self pollination does not lead to fertilisation and there is no seed setting. The pollen grains may fail to germinate on the stigma; or, even if pollen tubes are formed their growth may be inhibited. Such a condition where pollen grains fail to fertilise ovules of the same plant is known as self-incompatibility or self sterility. *Nicotiana* (tobacco) was one of the first plants in which self-sterility was noticed.

Breeding tests have established that self-incompatibility is controlled by multiple alleles of a single gene and may be of two types—gametophytic or sporophytic. The gametophytic self incompatibility (*GSI*) is controlled by the genotype of the pollen grains. This type is quite common in families like Liliaceae, Gramineae and Solanaceae. Sporophytic self-incompatibility (*SSI*) is controlled by the somatic or sporophytic cells of the plant which produce pollen grains, as observed in families Cruciferae and Compositae.

East and Mangelsdorf in 1925 explained the behaviour of sterility genes. The gene controlling incompatibility is designated S and its multiple alleles are represented by a series

S_1, S_2, S_3, S_4, S_5, ... and so on. As already explained earlier only two alleles are present in the diploid cells of a plant. In the *GSI* system pollen grains produced by an S_1S_2 plant will have either S_1 or S_2 allele and are not able to pollinate a pistil with S_1S_2 genotype. But if S_1 or S_2 pollen grains land on an $S_3\,S_4$ pistil, fertilisation takes place (Fig. 3.2 *a*, *b*). In plants belonging to the SSI system if even one of the alleles present in the sporophytic tissue of the male parent is present in the pistil of the female parent, then all the pollen grains produced by the male parent fail to effect fertilisation of that pistil. Thus, if the male parent has the genotype S_1S_2, it would be incompatible with plants having S_1S_2, S_1S_4, S_1S_5, S_2S_3, S_2S_4, S_2S_5 genotypes, but would be compatible with S_3S_4, S_3S_5, S_4S_5 plants (Fig. 5.2*c*). A very large number of alleles of the incompatibility genes are known to be present in other plants. *Oenothera* (evening primrose) has 37 different self sterility alleles and red clover has more than 200.

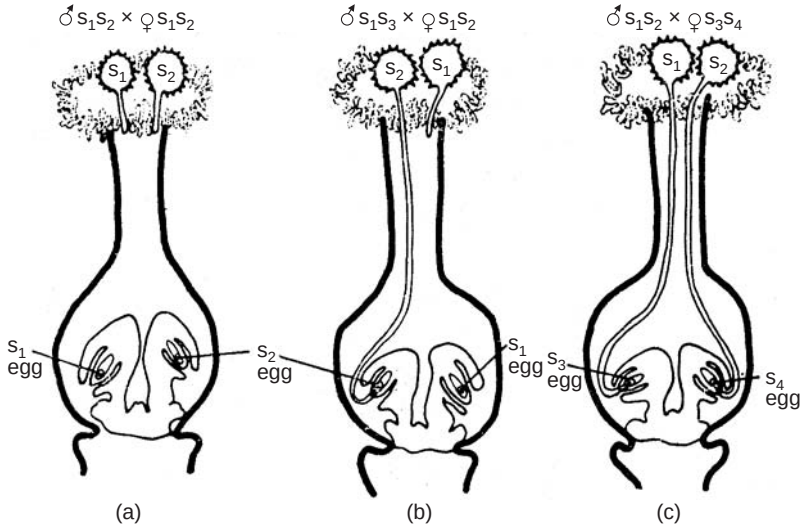


Fig. 5.2 Diagrammatic representation of incompatibility envisioned for plants of the GSI and SSI systems.

BLOOD GROUP ALLELES IN MAN

The *ABO* blood group system in man provides an excellent illustration of a multiple allelic system. If blood from two persons is allowed to mix outside the body, sometimes it will mix well, at other times there is clumping (agglutination) of red blood cells. The reason for this was explained by K. Landsteiner in 1900. He took blood samples from different people, separated their red cells from the colourless serum or plasma, and then mixed them up in various ways. He found that the red cells derived from one sample would mix smoothly with the serum of the same person. But when red cells of one person are mixed with the plasma of another person, sometimes the two would mix well, at other times there would be clumping of red cells. On the basis of these reactions Landsteiner discovered four blood types in man, namely *A*, *B*, *AB* and *O*.

Later on Landsteiner found that human erythrocytes carry on their surface two distinct types of proteinaceous antigens *A* and *B* which can stimulate the production of specific antibodies. Persons with type *A* blood have the *A* antigen (also called agglutinogens), type *B*

persons have *B* antigen, while persons with *AB* blood have both *A* and *B* antigens. Blood type *O* persons have neither *A* nor *B* antigen. These antigens stimulate production of specific serum antibodies (agglutinins) when human blood is injected into an animal such as the rabbit or guinea pig. There are also natural antibodies in the blood so that persons with type *A* blood have anti-*B* antibodies, and *B* group persons have anti-*A* antibodies. Blood of *AB* type does not have antibodies in its plasma. Persons with *O* type blood have both anti-*A* and *B* antibodies and are often called universal donors as they can donate blood to persons of *A*, *B* and *AB* types. The occurrence of natural antibodies in human blood remains a scientific curiosity which research has not been able to explain so far.

Bernstein in 1924 found that the blood group antigens are controlled by an autosomal gene designated *I* (isohaemaglutinin) which has 3 alleles I^A, I^B and I^O. Allele I^A produces type *A* blood and is dominant over the allele I^O which in the homozygous state (I^OI^O) produces type *O* blood. An *A* type individual can have the genotype I^AI^A or I^AI^O. Similarly allele I^B produces blood group *B*, and is dominant over the allele I^O. An individual with blood group *B* may be homozygous I^BI^B or heterozygous I^BI^O. *AB* type persons have the genotype I^AI^B where both alleles are codominant and equally expressed (explanation in Chapter 2).

Serologically, blood group *A* is distinguishable into two subgroups *A* and A_2. The number of blood groups therefore increases to 6—A_1, A_2, *B*, A_1B, A_2B and *O*. The genes controlling blood groups contain information for certain characteristics of the sugar molecule present along with glycoprotein on the surface of the erythrocyte. The sugar molecule determines antigenic specificity. The immunologically active components of both *A* and *B* contain 5 identical sugars. Of these *D* galactose and *N*-acetyl-*D*-galactosamine occupy terminal positions of the carbohydrate chains in *A* and *B* components respectively. These two sugar molecules differ from each other only at the position of carbon No. 2 which in *D* galactose is occupied by a hydroxyl group, and by *N*-acetylamino group in the sugar *N*-acetyl-*D*-amino-galactosamine. This difference in the arrangement of atoms at the position of the second carbon in the sugar molecule is responsible for *A* and *B* antigenic specificities.

The blood group genes are inherited in the Mendelian fashion in the case of *A*, *B* and *O* groups. The *AB* group however, results from codominance of *A* and *B* alleles. Thus blood groups of offspring can be determined from their parent's genotypes in the following way.

* *Genotypes of parents*	*Genotypes of offspring*	*Blood groups of offspring*
AA × *AO*	*AA, AO*	*A*
AO × *AO*	*AA, AO, OO*	*A, O*
AA × *BB*	*AB*	*AB*
AA × *AB*	*AA, AB*	*A, AB*
AO × *AB*	*AA, AO, BO, AB*	*A, B, AB*
AO × *BB*	*AB, BO*	*AB, B*
AO × *BO*	*AO, BO, AB, OO*	*A, B, AB, O*
AO × *OO*	*AO, OO*	*A, O*
OO × *OO*	*OO*	*O*
OO × *BB*	*BO*	*B*

* For sake of convenience the genotypes are represented simply by the symbols *A*, *B*, and *O* instead of I^A, I^B, I^O.

MULTIPLE ALLELES AND COMPLEX LOCI

So far we have considered multiple alleles as a series of alternative forms of a gene present at the same locus. The earlier concept of genetic recombination implies crossingover *between* genes, *not* within a gene. The idea was based on breakage and exchange of segments of paired homologous chromosomes, resulting in new linear arrangements of genes. Genetically it could be observed in the test-cross progeny of a heterozygote and a homozygous recessive; the frequency of recombinant phenotypes was used for gene mapping (Chapter 8). The two genes involved in recombination must obviously occupy different loci.

In 1942, Oliver showed that in the case of lozenge eye alleles in *Drosophila*, crossing over could occur between alleles present in the same locus. But because at that time the gene was thought to be an indivisible unit, his view was not accepted.

Lewis found some other mutants in *Drosophila* which behaved in the same ways as the lozenge eye mutants. Since they could produce recombinant progeny in test crosses, due to crossing over between two mutant sites in a locus, it became doubtful that they were true alleles. The term *pseudoalleles* was coined for such alleles.

The earlier concept of genes arranged like beads-on-a string had to be revised. A gene is a sequence of nucleotides in *DNA* that controls a specific gene product. The different mutations of the gene may be due to changes in single nucleotides at more than one locations in the gene. Crossing over could take place between the altered nucleotides within a gene. Since the mutant nucleotides are placed so close together, crossing over is expected with a very low frequency. When several different genes which affect the same trait are present so close that crossing over is rare between them, the term *complex locus* is applied to them. The term multiple alleles can also be redefined in the following way. Within the nucleotide sequence of DNA that represents a gene, multiple alleles are due to mutations at different points within the gene.

QUESTIONS

1. In rabbits, chinchilla coat colour (c^{ch}) is dominant over Himalayan (c^{ch}). The gene for albino coat (c) is recessive to both chinchilla and Himalayan. Further, all the three genes are recessive to the full coat colour agouti (C). What phenotypic ratios would be obtained from the following crosses:
   (a) $c^{ch}c \times c^{h}c$, (b) $C^{c} \times c^{h}c^{ch}$, (c) $c^{h}c^{ch} \times c^{h}c^{ch}$.
2. A woman of blood group A, and whose father was group O, marries a man of blood group AB. Which blood groups could *not* be expected to occur among their children?
3. If one child in a family has blood group AB and another child O, what can you say about the genotypes of their parents.
4. A plant of *Nicotiana* containing self sterility alleles S_2S_4 is used as female parent for the crosses described below. If pollen from (a) S_1S_2, (b) S_2S_3, (c) S_2S_4, (d) S_1S_4, (e) S_3S_4 plants is used for pollinating the stigma of the female plant (S_2S_4), what genotypes would the offspring resulting from each cross have?

SELECTED READINGS

Baur, E. 1912, Vererbungs- und Bastardierungsversuch mit *Antirrhinum*. II. Faktorenkoppelung. *Zts. ind. Abst.-Vererb.-Lehre* **6**:201.

Carlson, E. 1959. The Comparative Genetics of Complex Loci. *Quart. Rev. Biol.* **34**:33.

Castle, W.E. Heredity of Coat Characters in Guinea Pigs and Rabbits. Carnegic Inst. Wash. Publ. 23.

Emerson, R. A. 1911. Genetic Correlation and Spurious Allelomorphism in Maize, *Annu. Rep. Nebr. Agr. Sta.* **24**:52.

Green, M.M. 1963. Pseudoalleles and Recombination in *Drosophila*. In Methodology in Basic Genetics. Ed. Burdette, W.J. Holden-Day, San Francisco.

Hurst, C.C. 1906. Mendelian Characters in Plants and Animals. *Rep. Conf. Gen. Roy. Hort. Soc.* London, P. 114.

Lewis, D. 1979. Sexual Incompatibility in Plants. Edward Arnold, London.

Lewis, E.B. 1948. Pseudoallelism in *Drosophila Melanogaster. Genetics* **33**:113.

Punnett, B.C. 1912. Inheritance of Coat Colour in Rabbits. *J. Genetics* **2**:221.

Shull, G.H. 1911. Reversible Sex Mutants in *Lychnis dioica. Bot. Gaz.* **42**:329.

Sturtevant, A.H. 1913. The Himalayan Rabbit Case, with Some Considerations on Multiple Allelomorphs. *Amer. Nat.* **47**:234.

Thompson, J.S. and Thompson, M.W. 1973. Genetics in Medicine. Saunders. Philadelphia.

6

Cell Division and Chromosome Theory of Inheritance

The period that follows the formation of a cell by division of its mother cell until the time when the cell divides again to form two daughter cells is called the cell cycle. The cycle consists of four phases, *G*1, *S*, *G*2 and *M*. The first three phases (*G*1, *S* and *G*2) comprise interphase while *M* constitutes cell division (mitosis or meiosis). The *S* phase lasts about 6–8 hours in mammalian cells, *G*2 about 3–4 hours, while the length of *G*1 is variable. Whereas *DNA* synthesis is restricted to the *S* (synthetic) phase, protein synthesis takes place throughout interphase. A cell entering mitosis or meiosis has double the quantity of *DNA* and chromosomal proteins. The cell organelles such as mitochondria and ribosomes are assembled throughout interphase in the cytoplasm and passed on to the two daughter cells.

MITOSIS

Cell division in the somatic or body cells of diploid organisms takes place by mitosis (Fig. 6.1) in successive stages described below.

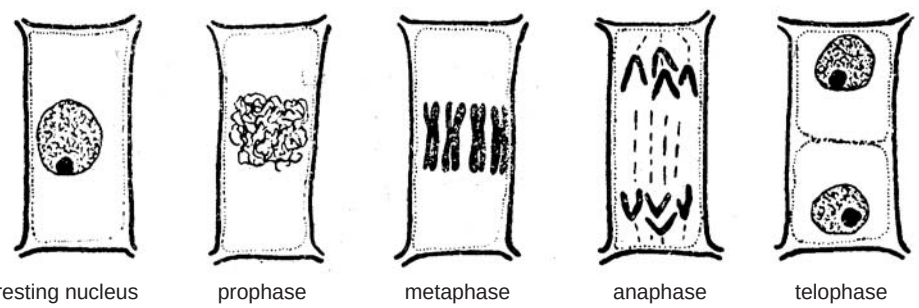


Fig. 6.1 Diagrams showing successive stages of mitosis. For convenience only 4 chromosomes are represented.

Prophase

The *DNA* in the diffuse chromatin of the resting nucleus in interphase has been duplicated in *S* phase preceding this cell division. The extended state of interphase chromatin allows transcription and replication of *DNA*, but is not suitable for division into two daughter cells. Therefore, prophase, the first stage of cell division shows contraction and condensation of chromatin into shorter, thicker fibres. By mid-prophase, the nucleolus starts to disappears and nuclear membrane breaks down, so that chromatin lies free in cytoplasmic space. Shortening of chromatin fibres continues to yield thicker, somewhat rod-like chromosomes. The breakdown of the nuclear envelope involves the enzyme **Cdk kinase** which is activated just before initiation of mitosis, in *G*2 phase of cell cycle. The inner face of the nuclear envelope is lined by a layer of fibrillar proteins of the cytoskeleton (called **intermediate filaments**), termed **nuclear lamina**. The enzyme *Cdk* kinase phosphorylates the lamin filament molecules, causing disassembly of nuclear lamina. The entire nuclear envelope that surrounds the condensing chromatin then breaks up into small vesicles which disperse into the cytoplasm. The nuclear pore complexes in the nuclear envelope also dismantle during fragmentation of the nuclear envelope.

Metaphase

The dissolution of the nuclear envelope at the end of prophase is described by some authors as the **prometaphase** stage. At metaphase, maximum condensation of chromatin fibres has been achieved giving rise to distinct rod-like chromosomes. Metaphase represents the most condensed state of chromatin in a cell. The molecular mechanisms responsible for chromosome condensation are still poorly understood. The process seems to involve the *DNA* untangling enzyme topoisomerase II (Chapter 14). The thick rod-like chromosomes begin to align themselves in the centre of the cell on what is conventionally referred to as the equatorial plate or the **metaphase plate**.

Structure of Metaphase Chromosome: Each chromosome consists of two **chromatids** and a region of central constriction called **centromere** or **primary constriction**. Evidence has established that each chromatid consists of a single duplex *DNA* molecule. The centromere region contains many copies of highly repeated *DNA* sequences (details in Chapter 19). A small nodule-like structure called the **kinetochore** is present at the outer surface of the centromere in each chromatid. The kinetochore functions as the site of attachment of **microtubules**, a bundle of fibres making up a spindle fibre, all the fibres together constituting the mitotic spindle apparatus. Microtubules attached to the kinetochore are called chromosomal or **kinetochore microtubules**. The centromere divides longitudinally, and the two centromeres pull the chromatids apart to the two poles by means of spindle fibres.

The division of the centromere into two at metaphase of mitosis is a key event that segregates accurately, half the genetic material for one daughter cell and half for the other daughter cell. This is ensured by the fact that the two chromatids contain duplicated *DNA* acquired from *DNA* synthesis in *S* phase of cell cycle preceding mitosis. Alignment of metaphase chromosomes in the equatorial region of the cell is followed by beginning of separation of chromatids to opposite poles. Misalignment of chromosomes in equatorial region arrests cells at metaphase and failure to segregate genetic material to the two daughter cells.

The protein encoded by a **gene MAD2** is normally localised at the kinetochores of prometaphase chromosomes and misaligned metaphase chromosomes, but is *not present* on chromosomes that have become properly aligned at the metaphase plate. Cells that possess mutant copies of a gene *MAD*2 fail to become arrested at metaphase when their chromosomes are misaligned. The presence of *MAD*2 at the kinetochores seems to provide a “wait” signal

that delays progression into anaphase. As each chromosome becomes aligned at the metaphase plate, its kinetochore loses all of its *MAD*2 molecules. It is only after *MAD*2 protein is absent from all of the chromosomes that anaphase can begin.

Microtubule Organisation and Centrosome: In most cells microtubules extend outward from a microtubule-organising center, which in animal cells is called the **centrosome**. Centrosomes are absent in plant cells. During mitosis, microtubules extend out from the duplicated centrosomes to form spindle fibres. Thus the centrosome seems to play a key role in determining the intracellular organisation of microtubules, but details of its function are not known. The centrosomes contain a pair of **centrioles**, oriented perpendicular to each other, and surrounded by pericentriolar material. Centrioles are cylindrical structures consisting of nine triplets of microtubules, similar to the basal bodies of cilia and flagella.

During spindle formation, microtubules emerge from the pericentriolar material around a centrosome and form a star-like **aster**. These are referred to as **astral microtubules**. The pericentriolar material acts as the nucleating site for microtubules of the aster. The fast growing ends of the microtubules, denoted **plus ends**, are away from the centrosome. The microtubule initiating activity appears to be stimulated by the *Cdk* protein. Following aster formation, the two centrosomes separate and move to opposite poles, while microtubules stretching between them increase in number and elongate. These are called **polar microtubules**. The two centrosomes establish two poles of the mitotic spindle. After mitosis, one centrosome is distributed to each daughter cell. Centrosomes are not essential components of mitotic spindles in all cells. Many animal cell types do not have centrosomes, nor do higher plants. The minus ends of the microtubules are thought to be gathered into a cluster at each spindle pole through the activity of motor proteins described later.

Anaphase

The two sister chromatids of each chromosome split apart and start moving towards opposite poles. There is rapid degradation of an *inhibitory protein* that acts a proteinaceous "glue" holding the two chromatids together, that facilitates the onset of anaphase. The degradation of anaphase inhibitory proteins occurs in response to activity of the mitotic *Cdk*s. The separation of sister chromatids requires activity of topoisomerase II. The chromatids (actually chromosomes now) start moving towards the poles, accompanied by shortening of microtubules attached to their kinetochore. Shortening of the microtubule results from the loss of subunits at the kinetochore (minus end) during anaphase. When the chromosomes reach the poles, there is elongation of the mitotic spindle, resulting in a simultaneous movement of the spindle poles further away from each other. This is accompanied by the addition of tubulin subunits to the plus ends of the polar microtubules. The movement of chromosomes to the poles has been observed to be completed within 2 to 60 minutes. The forces that bring about movement are discussed later in this chapter.

Telophase

Chromosomes at the poles organise into a mass of chromatin at each pole that marks the beginning of telophase. The process of formation of nuclear envelope begins when membranous vesicles start fusing with one another to produce a double-membraned envelop surrounding the chromatin. The nucleolus reappears. Thus, two daughter nuclei that are identical in genetic constitution to the parent nucleus are formed. Cytoplasmic membranous organelles such as Golgi complex, endoplasmic reticulum reform in each daughter cell.

Cytokinesis

In animal cells division of the cytoplasm of the parent cell is initiated at late anaphase as an indentation of the cell surface that appears as a band around the cell. The band deepens to form a furrow, the plane of the furrow lying in the same plane as the equatorial or metaphase plate on which chromosomes were aligned. Further deepening of the furrow splits the parent cell into two daughter cells. In plants a cell plate consisting of polysaccharides starts depositing in the central region of the parent cell. This gives rise to the first wall or middle lamella between the two cells. Later on a primary wall is deposited towards the inside of the middle lamella of each daughter cell.

MEIOSIS

The division which takes place in cells of the germ line is called meiosis (Fig. 6.2). It results in four products of a parent cell with half the amount of genetic material because *DNA* is duplicated only once and there are two cell divisions, meiosis I and II.

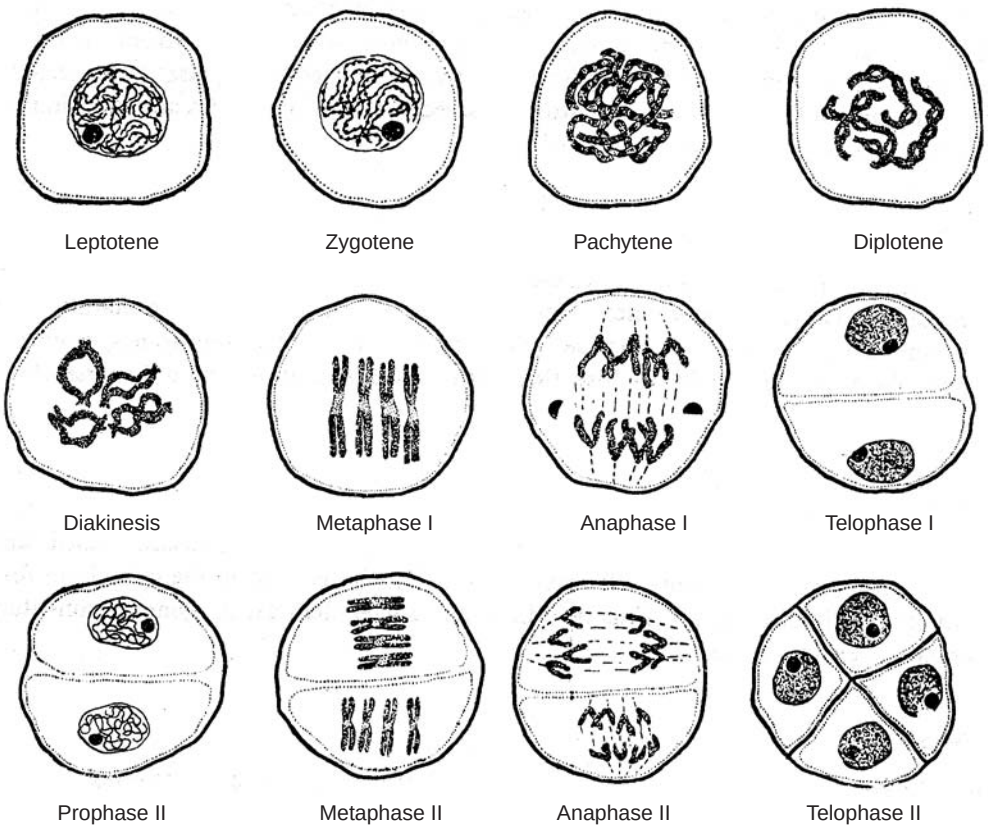


Fig. 6.2 Diagram illustrating stages of meiosis. Only four chromosomes are shown.

Prophase I is of very long duration and consists of five substages. *Leptotene* shows very long, thin thread-like chromosomes. *Zygotene* marks the initiation of pairing of homologous chromosomes. By some force of attraction identical partners are drawn towards each other.

Condensation and shortening of chromosomes is visible. By *pachytene* pairing is complete and stabilised. Due to the intimate nature of pairing or synapsis and continued shortening of chromosomes, thick ribbon-like bivalents are formed. Pairing is exact so that chromomeres and centromeres of homologues lie against each other. The number and arrangement of bead-like chromomeres, position of centromeres and arm length are distinct for each bivalent and allow mapping of pachytene chromosomes. The nuclear membrane and nucleolus disappear. At *diplotene* a force of repulsion between the paired chromosomes tends to draw them apart. They are held together at positions of chiasmata where genetic crossing over or exchange of segments takes place. Continued condensation of chromosomes through *diakinesis* gives rise to the short thick cross-shaped configurations of chromosomes.

By *Metaphase* I maximum condensation has been achieved giving rise to short rod-like bivalents which move towards the centre of the cell and align in the equatorial plane. Their centromeres become attached to spindle fibres running to opposite poles and homologues separate. Due to separation of homologues, half the number of chromosomes will reach one pole and one half of the other pole. Consequently Metaphase I is referred to as the *reductional* division.

The beginning of separation of homologous is also the beginning of *Anaphase* I. When an entire set of chromosomes reaches either pole we call it *Telophase*. The nuclear membrane and nucleolus are reorganised to form two daughter nuclei. The newly formed nuclei go through a short rest period or interkinesis before entering the second meiotic division. Prophase II of meiosis is initiated by condensation of chromosomes and is completed rapidly as in mitosis. *Metaphase* II shows the equatorial alignment of rod like chromosomes which have reached their maximum limit of contraction. The centromere of each chromosome *now* divides longitudinally. The daughter centromeres are attached to spindle fibres and separate to the poles as in mitosis. Thus in meiosis, metaphase II is an equational division. *Anaphase* II and *telophase* II are completed as in meiosis I and in mitosis.

At the end of meiosis II four products are formed each with half the amount (haploid) of genetic material that was contained in the original parent nucleus.

Occurrence of Meiosis

In sexually reproducing diploid plants meiosis takes place in the anther and the ovule. One or more specialised cells of the germ line known as archesporium divide mitotically to produce the sporogenous cells. In the case of male, *sporogenous* cells multiply and increase in number. After a certain stage they stop dividing, are ready to enter meiosis and are called *microspore mother cells.* The four products of meiosis are united in a tetrad but later separate as uni-nucleate *microspores*. Thickenings deposited on the microspore wall produce uninucleate **pollen grain** which soon becomes binucleate and a large vegetative cell and a small generative cell are organised. In many plants pollen grains are shed in this stage. The nucleus of the generative cell divides usually after pollen germination has begun on the stigma resulting in two male gametes. One male gamete fertilises the egg to form a zygote: the second male gamete fuses with the secondary nucleus in the embryo sac and gives rise to the endosperm which provides nutrition to the growing embryo.

Similarly in the case of female, the *sporogenous cell* in the ovule enlarges to become a *megaspore mother cell* which divides meiotically to form a tetrad of four megaspores. Usually three megaspores degenerate and one enlarges into the embryo sac. By three mitotic divisions the nucleus of the megaspore forms 8 nuclei (haploid) of which one organises as the egg.

In the male in higher animals (mammals and man) the *spermatogonial* cells in the testis increase in number by mitosis. When they are ready to divide meiotically they are called

primary spermatocytes. During meiosis II they are called *secondary spermatocytes*. After meiosis they organise into the elongated *spermatids* which finally produce sperms after undergoing some changes (head, middle piece and tail are formed; some histone proteins are changed; motility is acquired).

In the female of mammalian species including humans, and in contrast to the males meiosis takes place during the embryonal life of an individual. During the first few months after conception in human beings, the primary oocytes in the foetal ovary start undergoing meiosis. After meiosis I the primary oocyte produces a secondary oocyte which alone will produce the functional ovum, and a small first polar body. The polar body gives rise to two more polar bodies: eventually all three polar bodies degenerate.

A unique feature of ovarian meiosis is that it stops at about metaphase II stage in the secondary oocyte. In this condition the secondary oocytes (about 400 in number) remain suspended for 40 to 50 years of life after birth. At puberty ovulation starts when one oocyte gets released from the ovary wall each time and is discarded. If an oocyte gets a chance to meet a sperm, then before fertilisation it completes the remaining stages of meiosis (anaphase II to telophase II). Out of the four resulting cells three are polar bodies which will degenerate; the remaining cell enlarges and functions as ovum during fertilisation.

Genetics of Meiosis

There are two special features of meiosis: production of haploid gametes containing recombined genetic material of the two parents; the process of genetic exchange and recombination in homologues. Some aspects of recombination are discussed in Chapters 8, 19 and 22. Genetic events in meiosis have also been studied from mutants. Mutations affecting meiosis lead to abnormalities in genetic recombination or in chromosome segregation. Many such mutants have been isolated in the lower eukaryotes like yeast, *Neurospora*, and a few higher plants. The work on yeast mutants has shown that recombination is regulated at two levels: control of overall frequency of crossing over; and controls which influence the frequency of crossing over only in particular regions. Some mutations are known to affect the subsequent stages of chromosome segregation. There are also mutations which suppress initiation of meiosis.

Detailed studies of meiotic mutants have been carried out in *Drosophila*. Meiosis is unusual in normal *Drosophila* males due to absence of crossing over. The females show cross-overs in all chromosome pairs except number IV. Mutations in male meiosis affect chromosome segregation. One such effect is *nondisjunction* due to which chromosomes fail to disjoin and move to opposite poles either at meiosis I or II. Most meiotic mutants of *D. melanogaster* interfere with meiosis I either in male or in female, but not in both. This implies that control of meiosis I is different in male and in female flies. Some mutations affecting meiosis I in *Drosophila* males have been investigated. The mutant segregation distorter (*SD*) is due to a dominant gene on chromosome II. The heterozygous males (*SD*/ ±) transmit the mutant *SD* gene to 50% of the progeny; the homozygous (*SD*/*SD*) males are sterile. Another mutant in males called recovery disrupter (*RD*) is due to an *X*-linked gene. It causes fragmentation of *Y*-bearing sperm resulting in excess of female flies in the progeny. A few more mutations in males such as *mei-S*8 and *mei*-081 cause nondisjunction at meiosis in *Drosophila* females such as *c*(3) *G*, and *mei*-218 either reduce recombination or interfere with chromosome segregation. Those that affect recombination also cause increased nondisjunction of all chromosomes.

The Spindle Apparatus

Electron microscopy has shown that spindle fibres are made up of bundles of parallel filaments called microtubules. The microtubules are assembled from cytoplasmic proteins namely α and β tubulin, and have an outside diameter of 24 nm, a central lumen of 15 nm across, and variable

length. In cross section a microtubule appears circular in outline, the circle itself being composed of about 13 smaller circles. These small circles represent cross sections of long strands called protofilaments which are assembled from the globular protein subunits, the α and β tubulin present in the cytoplasm. One molecule each of α and β tubulin become associated to form a dimer. The dimers are arranged in linear order to form protofilaments (Fig. 6.3). Each dimer appears to have a specific binding site for colchicine and another for vinblastine (Weisenberg, 1972), both of which inhibit spindle formation by preventing the assembly of microtubules.

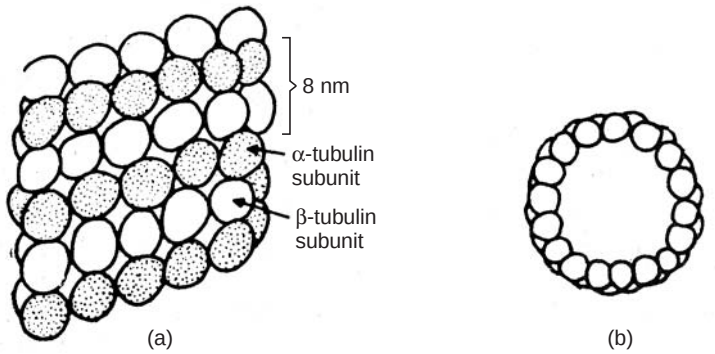


Fig. 6.3 Cross sectional and side views of spindle microtubules.

The microtubules are supposed to be present in the form of a cytoplasmic network in a resting cell. During cell division they become organised as spindle fibres. They are disassembled after cell division. It seems that for normal mitosis there must be a state of equilibrium between microtubules and free subunits of tubulin present in the cytoplasm. Low temperatures disturb the equilibrium and dissociate microtubules.

The alkaloid colchicine obtained from the corms of a liliaceous plant *Colchicum autumnale* binds to tubulin and prevents formation of spindle fibres. Due to the resulting failure of chromosome movement, cell division becomes arrested at metaphase. Such cells may either degenerate, or the duplicated chromosomes may form a nucleus which is polyploid. The effect of colchicine can be reversed by depriving the cells of colchicine. A few other chemicals such as nitrous oxide, acenaphthene, chloral hydrate, vinblastine and podophyllotoxin have the same effect as colchicine.

Forces Required For Chromosome Movement

The forces powering chromosome movement have not been understood. A variety of different molecular motors have been identified in a wide variety of species. All of the motors thought to be involved in chromosome movement are microtubule motors that include some kinesin-related proteins and cytoplasmic dynein.

Kinesin and motor proteins **dynein** move along microtubules in opposite directions, kinesin toward the plus end and dynein toward the minus end. Kinesin translocates along microtubules in only a single direction, toward the plus end. The kinesin molecule is 380 kDa in weight, consists of two heavy chains of 120 kDa each, and two light chains, 64 kDa each. The heavy chains have long α-helical regions wound around each other in coiled coil manner. The amino-terminal globular head domains of the heavy chains are the **motor domains** of the molecule. They bind to both microtubules and ATP, the hydrolysis of ATP providing the energy required for movement. The tail portion of the kinesin molecule consists of the light chains in association with the carboxy-terminal domains of the heavy chains. This portion of kinesin is

responsible for binding to other cell components, such as organelles and membranous vesicles that are transported along microtubules by the action of kinesin motors.

Dynein is a very large molecule, up to 2000 kDa, consisting of two or three heavy chains, complexed with various light polypeptides. In dynein too, the heavy chains form globular ATP-binding motor domains that are responsible for movement along microtubules. The light chains of this molecule bind to organelles and membranous vesicles. All members of the dynein family move toward the minus ends of microtubules.

Regulation of the Cell Cycle

Most eukaryotic organisms duplicate cells by following a more or less similar cell cycle. Since diverse organisms follow a similar pattern for cell duplication, it implies that the cell cycle is under genetic control. Disruption of the genetic controls leads to uncontrolled cell proliferation, as seen in malignancy. Regulation of the cell division cycle involves extracellular signals from the environment as well as internal signals that exert their effect on processes during different cell cycle phases (*G*1, *S*, *G*2 and *M*). In addition, cellular processes such as cell growth, DNA replication and cell division must be coordinated for progression of cell cycle. This is accomplished by a number of control points that check and regulate progression through the different phases of cell cycle.

An important cell cycle regulatory point occurs late in phase *G*1 and controls progression from *G*1 to *S*. This regulatory point was first found in yeast (*Saccharomyces cerevisiae*), where it is referred to as **START** (Fig. 6.4). After passing START, cells become committed to go through one cell division cycle. In yeast, the passage through START is controlled by external signals such as availability of nutrients. If there is shortage of nutrients, yeast cells become arrested at START and enter a nondividing resting stage. START is also the point at which cell growth is coordinated with DNA replication and cell division. Budding yeasts which produce progeny cells of different sizes, cell size is monitored by a control mechanism. Accordingly, each cell must reach a minimum size before it can pass through START.

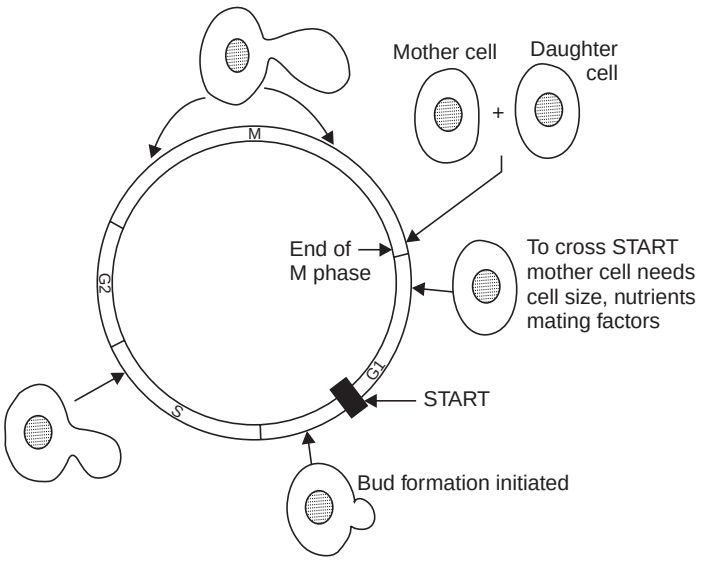


Fig. 6.4 Cell cycle of yeast is regulated at a point START in late G1. After mother cell crosses START, bud formation begins, and is completed after mitosis. The size of the bud reflects the position of the cell in cell cycle.

In eukaryotes, cell proliferation is regulated at the *G*1 phase of cell cycle called restriction point. In contrast to yeasts, the passage of mammalian cells through cell cycle is regulated by extracellular growth factors that signal cell proliferation, instead of availability of nutrients. When the appropriate growth factor is present, cells pass the restriction point and enter *S* phase. Once it has passed through the restriction point, the cell becomes committed to proceed through *S* phase and complete the cell cycle, even in absence of further growth factor stimulation. Progression through cell cycle stops at the restriction point if appropriate growth factors are not available in *G*1. Thus cells become arrested at a quiescent stage called *G*0 (*G* zero). Cells in *G*0 are metabolically active, but have reduced rates of protein synthesis. It has been noted that many animal cells remain in *G*0 unless induced to proliferate by appropriate growth factors or other extracellular signals. A good example is afforded by skin fibroblasts that remain arrested in *G*0. When they are required to repair damage resulting from a wound injury, they are stimulated to divide. The proliferation of fibroblasts is signalled by platelet-derived growth factor released from blood platelets during clotting.

An example of cell cycle control in *G*2 is provided by vertebrate oocytes. Oocytes can remain arrested in *G*2 for very long periods of time, until they are triggered by hormonal stimulation to proceed to *M* phase. In human female, oocytes become arrested in *G*2 during fetal life for several decades until stimulated to complete the meiotic cell cycle.

Checkpoints in Cell Cycle

The coordination between the different phases of the cell cycle depends on a system of checkpoints as well as feedback controls that prevent entry into the next phase of the cell cycle until events of the previous phase have been completed. To ensure that incomplete or damaged chromosomes are not replicated and passed on into daughter cells, there are several cell cycle checkpoints. One well defined checkpoint occurs in *G*2 which prevents initiation of mitosis until DNA replication is completed. The *G*2 checkpoint senses unreplicated *DNA*, then generates a signal that leads to cell cycle arrest. Thus the *G*2 checkpoint prevents the initiation of *M* phase before completion of *S* phase. The cells remain in *G*2 until the entire genome has been replicated. The checkpoint then releases inhibition of *G*2, allowing the cell to progress into mitosis, and completely replicated chromosomes are passed on into daughter cells. Progression into the cell cycle is also arrested at the *G*2 checkpoint in response to *DNA* damage caused by radiation. In this case, arrest in *G*2 allows time for repair of damage.

DNA damage can also arrest cell cycle progression at a checkpoint in *G*1, that allows repair of damage to take place before the cell enters *S* phase. Thus, replication of damaged DNA is prevented. In eukaryotes, arrest at the *G*1 checkpoint is mediated by the action of a protein called **p53** which is rapidly induced in response to damaged DNA (Fig. 6.5). The gene encoding *p*53 is frequently mutated in human cancers. Mutations in this gene lead to loss of function of *p*53 that prevents *G*1 arrest, in response to DNA damage. Thus, damaged DNA is replicated and passed into daughter cells without being repaired. Through damaged DNA the daughter cells inherit an increased frequency of mutations and instability of the genome which contribute to cancer development. Mutations in *p*53 gene are most common genetic alterations in human cancers.

The third cell cycle checkpoint occurs toward the end of mitosis. This checkpoint ensures alignment of chromosomes on the mitotic spindle, so that a complete set of chromosomes is distributed into the daughter cells. If one or more chromosomes fail to align correctly on the spindle fibres, the checkpoint leads to arrest at metaphase. The chromosomes do not separate at the equator until a complete complement of chromosomes has been organised for accurate distribution to the two poles.

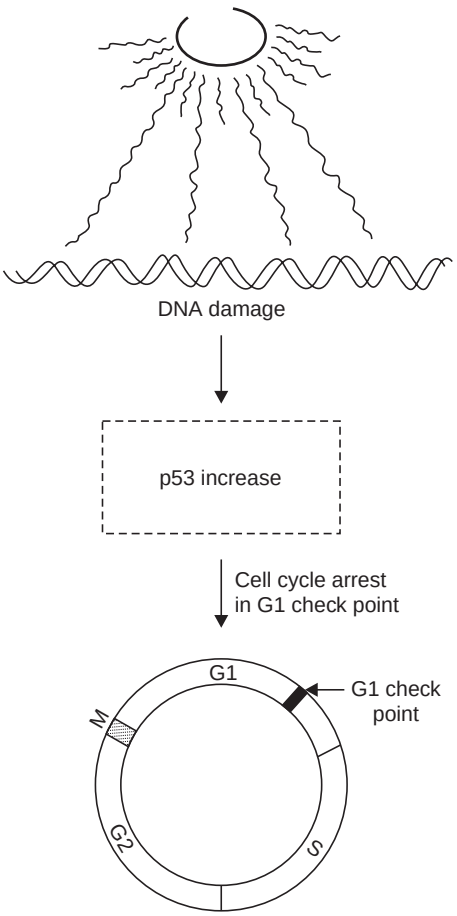


Fig. 6.5 Involvement of p53 in G1 arrest induced by DNA damage. When DNA is damaged by irradiation or any other means, p53 levels increase and signal cell cycle arrest at the G1 checkpoint.

Mechanisms for Regulation of Cell Cycle

Studies on molecular mechanisms that control the progression of mammalian cells through the division cycle have revealed that the cell cycle is controlled by a set of protein kinases which are responsible for triggering the major cell cycle transitions.

*Cdc*2 *and Cyclin*

Studies on frog oocytes indicated that the oocytes that are arrested in *G*2 phase of cell cycle could be stimulated to enter into the *M* phase of meiosis by hormonal stimulation. Later investigations showed that oocytes arrested in *G*2 could be induced to enter *M* phase by microinjection of cytoplasm taken from oocytes that had been hormonally stimulated. Thus, a cytoplasmic factor present in hormonally stimulated oocytes would allow oocytes that had not been exposed to hormone, to progress from *G*2 to *M*. This cytoplasmic factor was called **maturation promoting factor (MPF).** Further studies showed that *MPF* is not specific to oocytes, but appeared to act as a general regulator of the transition from *G*2 to *M* in somatic cells as well.

Studies on temperature-sensitive mutants of yeast *Saccharomyces cerevisiae* that were defective in cell cycle progression also contributed to understanding of regulation of cell cycle. These mutants called *cdc* (cell division cycle mutants) were remarkable as they showed growth arrest at specific points in the cell cycle. For example, the mutant designated *cdc*28 replicates normally at the permissive temperature, but at the nonpermissive temperature there is arrest of cell cycle at START, thus indicating that the *cdc*28 protein is necessary for passage of cells through the regulatory point START in *G*1. Further investigations on *cdc*2 brought to light two important points. First, that *cdc*2 encodes a protein kinase, indicating the role of protein phosphorylation in cell cycle regulation. Second, a gene related to *cdc*2 was identified in humans and shown to function in yeasts, implying that the cell cycle regulatory activity of this gene is conserved.

Studies on protein synthesis in sea urchin embryos provided further insights into cell cycle regulation. After fertilisation, sea urchin embryos go through rapid cell divisions. However, the entry of embryonic cells into *M* phase requires new protein synthesis. Two proteins were then identified that accumulate during interphase and are degraded at the end of each mitosis. These proteins were called **cyclin A** and **cyclin B**. It seemed that cyclins might be able to induce mitosis by controlling entry and exit from *M* phase. It was then demonstrated that cyclins control *G*2 to *M* transition.

Subsequent investigations on *MPF*, the regulator of cell cycle gave interesting results when it was shown that *MPF* is composed of two subunits, *cdc*2 and cyclin *B*. Cyclin *B* is required for the catalytic activity of the *cdc*2 protein kinase. Thus, *MPF* activity is controlled by the periodic accumulation and degradation of cyclin *B* during cell cycle progression. Further studies have demonstrated regulation of *MPF* by phosphorylation and dephosphorylation of *Cdc*2 protein. Cyclin *B* synthesis takes place in *S* phase, and it forms complexes with *Cdc*2 protein throughout *S* and *G*2. During this time, *Cdc*2 is phosphorylated at two regulatory positions, one on threonine-161 (required for *Cdc*2 kinase activity), the other of tyrosine-15 catalysed by a protein kinase called Wee1 that inhibits *Cdc*2 activity resulting in accumulation of inactive *Cdc*2/cyclin *B* complexes during *S* and *G*2. The transition from *G*2 to *M* takes place by activation of the *Cdc*2/cyclin *B* complex brought about by dephosphorylation of threonine-14 and tyrosine-15 by a protein phosphatase called *Cdc*25. After activation the *Cdc*2 protein kinase phosphorylates a variety of target proteins for entry into *M* phase. Furthermore, *Cdc*2 activity also stimulates degradation of cyclin *B* by ubiquitin-mediated proteolysis. Degradation of cyclin *B* inactivates *Cdc*2, causing the cell to exit mitosis, undergo cytokinesis and enter interphase.

The Cyclins

The characterisation of the *Cdc*2/cyclin complex have provided insights into the regulation of the cell cycle. Both *Cdc*2 and cyclin *B* are found to be members of large families of related proteins, with different members of these families controlling distinct phases of the cell cycle. As already stated, in the case of yeasts, *Cdc*2 controls passage through START and entry into mitosis. By associating with distinct cyclins, *Cdc*2 is able to phosphorylate different substrate proteins required during specific phases of the cell cycle.

In higher eukaryotes, cell cycles are controlled not only by multiple cyclins, but also by multiple *Cdc*2-related protein kinases. Referred to as **Cdks** (**cyclin-dependent kinases**). The original member of this family, *Cdc*2, is known as *Cdk*1 with others following up to *Cdk*8. Members of the *Cdk* family associate with specific cyclins to accomplish progression through different stages of the cell cycle. Briefly, progression from *G*1 to *S* is regulated by *Cdk*2 and *Cdk*4 in association with cyclins *D* and *E*; *Cdk*4 and *Cdk*6 control progression through restriction

point in *G*1 in association with cyclins *D*1, *D*2, and *D*3; *Cdk*2 and cyclin *E* complexes are required for *G*1 to *S* transition as well as initiation of *DNA* synthesis in *S*; *Cdk*2 with cyclin A control progression of cells through *S* phase; *Cdc*2 complexed with cyclin *B* plays a role in transition from *G*2 to *M*.

The activity of *Cdks* during cell cycle is regulated by at least 4 molecular mechanisms. The first level of regulation involves formation of *Cdk*/cyclin complexes. Second, activation of *Cdk*/cyclin complex by phosphorylation of a *Cdc* threonine residue at position 160. The third involves inhibitory phosphorylation of tyrosine residues near the *Cdk* amino terminus. Fourth, binding of inhibitory proteins called **Cdk inhibitors** or **CkIs** to *Cdk*/cyclin complexes. The combined effects of these multiple mechanisms of *Cdk* regulation accomplish control of cell cycle progression in response to both checkpoint controls and to the large number of extracellular stimuli that regulate cell proliferation.

The D-Type Cyclins

The proliferation of animal cells is regulated by a variety of extracellular growth factors that control progression of cells through the restriction point in late *G*1. In absence of growth factors, cells are not able to progress through the restriction point, become quiescent or enter *G*0. When stimulated by growth factor, the cells can re-enter the cell cycle. The study of *D*-type cyclins has provided an important link between growth factor signalling and cell cycle progression. Cyclin-*D* synthesis is induced in response to growth factor stimulation and continues as long as growth factors are present. Growth factors must necessarily be present through *G*1 to allow complexes of *Cdk* 4, 6/cyclin *D* to drive cells through the restriction point. But if growth factors are removed prior to *G*1, the concentration of *D*-type cyclins falls rapidly. The cells are then unable to pass through *G*1 and *S*, become quiescent and enter *G*0. Thus, *D*-type cyclins play a role in growth factors control of progression of cells through *G*1.

The insights gained on growth factors and cyclin *D* led to a most important finding that defects in cyclin *D* regulation contribute to the loss of growth regulation that is characteristic of cancer cells. Indeed, many human cancers have been found to arise as a result of defects in cell cycle regulation, whereas many other cancers result from abnormalities in the intracellular signalling pathways activated by growth factor receptors. Therefore, mutations which result in the continuous unregulated expression of cyclin *D*1 lead to development of several human cancers, including lymphomas and breast cancers. Furthermore, a key substrate protein of *Cdk*4, 6/cyclin *D* complexes is found to be frequently mutated in a variety of human tumors. This protein is called **Rb** since it was first identified as the product of a gene responsible for retinoblastoma, a childhood eye tumor. It was later found that if *Rb* is rendered nonfunctional due to mutation, it could result in a variety of human cancers, besides retinoblastoma. Rb seems to be the prototype of a **tumour suppressor gene**, a gene whose inactivation leads to development of tumors.

Subsequent investigations have revealed that **Rb** plays a significant role in coupling the cell cycle machinery to the expression of genes required for cell cycle progression and DNA synthesis. Rb's function is regulated by changes in its phosphorylation as cells traverse the cell cycle. For example, *Rb* becomes phosphorylated by *Cdk* 4, 6/cyclin *D* complexes as cells pass through the restriction point in *G*1. When *Rb* is underphosphorylated (present in *G*0 or early *G*1), *Rb* binds to members of the **E2F** family of transcription factors, which regulate expression of several genes involved in cell cycle progression, such as gene encoding cyclin *E*. Details of *Rb*'s action indicate that *Rb* acts as a molecular switch that converts *E*2*F* from a repressor to an activator of genes required for cell cycle progression.

Inhibitors of Cell Cycle

Agents that cause damage to DNA result in arrest of cell cycle. Cell contacts between cancerous cells in culture arrest division (contact inhibition). A number of extracellular factors inhibit cell division, frequently mediated by regulators of the cell cycle machinery, usually by induction of *Cdk* inhibitors. For example, arrest of cell cycle in response to DNA damage is mediated by protein *p*53. Protein *p*53 a transcriptional regulator that stimulates expression of the *Cdk* inhibitor known as *p*21. The *p*21 inhibits several *Cdk*/cyclin complexes. Following DNA damage, *p*53 induces synthesis of *p*21 and brings about cell cycle arrest. Protein *p*21 can also directly inhibit DNA replication.

A well characterised extracellular inhibitor of animal cell proliferation is the polypeptide factor *TGF*-β, that inhibits the proliferation of a variety of types of epithelial cells by arresting cell cycle progression in *G*1. This action of *TGF*-β is mediated by induction of *Cdk* inhibitor *p*15 which blocks *Cdk*4 activity. Then *Rb* is not phosphorylated and the cell cycle becomes arrested in *G*1.

THE CHROMOSOME THEORY OF INHERITANCE

In order to understand the role played by chromosomes as carriers of hereditary material, we must turn the clock backward to some landmarks in the history of genetics. The earliest records date back to the work of the 18th Century plant hybridisers. One of them Kölreuter had some theoretical knowledge on the basis of his practical experience in hybridisation work on *Nicotiana*. He crossed *Nicotiana rustica* with *N. paniculata* and found that for all 13 characters studied by him, the resulting hybrid was intermediate between the two parents. The results of his reciprocal crosses were the same. Thus he was first to suggest that the hereditary contribution of the two parents to their offspring was equal.

The later half of the 19th century then covered a few more milestones in genetics. Oscar Hertwig in 1876 studied fertilisation in sea urchins while Strasburger (1877, 1884) and Schmitz (1879) made similar observations in plants. Hertwig noted the presence of two nuclei in the fertilised egg and concluded that fertilisation involved fusion of nuclei from two parental gametes. Working with seed plants, Strasburger in 1884 could clearly show that in the orchid *Orchis latifolia* the pollen tube travels downward through the pistil and enters the embryo sac. But the existence of a hereditary substance inside the nucleus was first postulated by August Weismann working as Professor at the University of Freiburg. He called the substance germplasm. The next problem was to understand the continuity of the germplasm *i.e.* its transmission from parent to offspring. It was Schneider in 1873 who first demonstrated the continuity of germplasm (nucleus) through cell division by observing condensing chromosomes and their movements in dividing eggs of the flatworm *Mesostomum*.

Walter Flemming in 1878 was first to study mitosis in detail and coined the term. He could observe metaphase chromosomes longitudinally split in half and their movement apart to each of the two daughter nuclei. In this way a parent cell could pass on two identical groups of chromosomes to its two daughter cells. Another evidence for the role of chromosomes in heredity came when E. Van Beneden and Weismann showed that gametes contained only half the number of chromosomes present in somatic tissues and that the somatic number was restored at fertilisation. In 1883 Wilhelm Roux went a step further by suggesting that cell division not only divides the quantity of nuclear material, but also its properties or individual qualities (hereditary determinants of a trait).

The cytological studies on chromosomes conducted till 1900 constitute the first or premendelian phase in the development of the Chromosome Theory of Inheritance. But after the rediscovery of Mendelism, all efforts were directed towards determining the relationship between Mendelian factors and chromosomes. In Mendel's crosses, the F_2 progeny segregated in typical ratios as expected on theoretical grounds. But the cytological basis of Mendelism was not understood until the behaviour of chromosomes during cell division was known. In 1902 Correns provided evidence that segregation of Mendelian factors occurred during meiotic division. In 1901 Montgomery made the important observation that during meiosis maternal chromosomes paired only with paternal chromosomes. He studied a very favourable material *Ascaris megalocephala* var. *univalens* which has only two chromosomes. Obviously when there is pairing it must involve a maternal and a paternal chromosome.

In 1902 McClung could associate a specific heritable trait in grasshopper with a specific chromosome. In the same year Theodor Boveri studied multipolar mitoses in sea urchin embryos and concluded that the individual chromosomes are qualitatively distinct from each other and carry different hereditary determinants. Boveri's conclusion was further strengthened by the observations of Walter Sutton in 1903 who could demonstrate morphological differences between the 23 chromosomes of the grasshopper *Brachystola*. Sutton was a graduate student of Wilson at Columbia University and is credited for demonstrating a parallel between meiotic behaviour of paired chromosomes and the behaviour of pairs of Mendelian factors. He could explain Mendel's principle of segregation by showing cytologically that in meiosis one member of a pair of homologous chromosomes goes to one daughter cell, the other to the second daughter cell. Mendel's second principle of independent assortment found cytological proof from the fact that members of one pair of homologous chromosomes move to the poles independently of the members of another pair. In this way the final mixture of chromosomes (paternal and maternal) at a pole is different from one cell to another. In other words, segregation of paired homologues at Metaphase I occurs at random. This phenomenon was also demonstrated explicitly by Carothers, a student of McClung, in grasshopper in which one pair of homologous chromosomes is such that one member is larger than the other partner. Moreover there is one unpaired chromosome in the grasshopper. She observed separation of the two unequal chromosomes to the two poles in about 300 cells and found that the unpaired chromosome passed to one pole with the larger homologue in about 50% of cells, and with the smaller homologue in the remaining 50% cells. From this she inferred that different chromosome pairs assort independently.

Boveri's work on sea urchins was confirmed later by Blakeslee in 1922 while working with *Datura* (Jimson weed). The normal diploid chromosome number in *Datura* is 24 which form 12 pairs in meiosis. Some plants however have 25 chromosomes (designated trisomics today). An interesting feature of the 25th chromosome is that it could be identical to any one of the 12 pairs so that *Datura* plants could be classified into 12 types. Blakeslee found that the shape and size of the fruit capsule was different in all the 12 groups of plants. This proved that the 12 chromosomes differed from each other qualitatively and each produced a morphologically different capsule.

The work of Boveri and Sutton provided excellent correlation between Mendelian factors and chromosomes and became known as the Sutton-Boveri theory. Their experiment led to the discovery of the following characteristic features of chromosomes:

(*a*) Continuity of chromosomes from one cell division to the next,

(*b*) Qualitative differences between individual chromosomes,

(*c*) Pairing of maternal with paternal chromosomes,

(*d*) Segregation of chromosomes at random *i.e.* independent assortment of chromosomes.

Sutton in 1903 found that the number of factors obeying Mendel's law were more than the number of chromosome pairs in the cell. This means that there were many genes on a single chromosome and Mendel's law of independent assortment could not be applied to them (due to phenomenon of linkage described in Chapter 8). T.H. Morgan in 1909 showed that in *Drosophila* the gene for white eye colour was linked to the sex chromosome. He crossed a white-eyed male fly with normal red-eyed female and obtained an F_1 progeny of red-eyed flies only. The F_1 red-eyed flies when mated amongst themselves produced F_2 progeny in the ratio of 3 red to 1 white. But the striking feature was that all the F_2 white-eyed flies were males. In this way Morgan demonstrated that the gene for eye colour was present on the *X*-chromosome. The following year he showed that the genes for yellow body colour and for miniature wings were also carried on the *X*-chromosome. By performing dihybrid crosses (involving eye pigment and body colour) he could show that crossing over could change positions of linked genes (incomplete linkage).

C.B. Bridges, a student of Morgan provided support to the chromosome theory from his studies on nondisjunction. When a white-eyed female is crossed with a red-eyed male, normally the F_1 consists of red-eyed females and white eyed males. Bridges found that some-times red-eyed males and white-eyed females were also present. He found a cytological explanation by which the two *X*-chromosomes failed to separate at Metaphase I of meiosis. Thus both *X*'s passed together into 50 per cent of resulting eggs; the remaining 50 per cent of eggs did not receive an *X*-chromosome, a phenomenon known as *primary nondisjunction*. If an egg with two *X*-chromosomes is fertilised by a normal *Y* carrying sperm, the resulting zygote has *XXY* chromosome constitution and is female. Bridges could demonstrate cytologically that such a female indeed had *XXY* chromosomes thus establishing the validity of the chromosome theory of inheritance.

QUESTIONS

1. A preparation of metaphase cells showed 13 chromosomes. Which explanation/s given below could be correct:
   (*a*) The cells are at metaphase I of meiosis;
   (*b*) The cells are in mitosis; they belong to an individual with *XO* sex chromosome constitution;
   (*c*) Impossible to explain; the cell has an uneven number of chromosomes;
   (*d*) The cells are at metaphase II of meiosis.
2. How many chromatids in a zygote nucleus would be:
   (*a*) of maternal origin?
   (*b*) of paternal origin?
3. Distinguish between:
   (*a*) mitotic metaphase and meiotic metaphase stages;
   (*b*) mitotic prophase and mitotic telophase.

4. A diploid cell has 6 chromosomes. One pair of homologous chromosomes has the centromeres very near the telomere, the second pair has it near the middle. The third pair is abnormal in that one chromosome is short and without a centromere, the other is longer than normal and dicentric. Draw a figure to show the anaphase stage of mitosis in such a cell.
5. Two daughter cells resulting from mitosis have the same amount of genetic material. Do they also have the same genotype?
6. Point out the similarities and differences between mitosis and meiosis.
7. What are the microscopically visible differences between paired homologues at pachytene and at metaphase I of meiosis?

SELECTED READINGS

Bridges, C.B. 1916. Nondisjunction as Proof of the Chromosome Theory of Heredity. *Genetics* **1** : 1.

Carbon, J. 1984. Yeast Centromeres: Structure and Function. *Cell* 37 : 351–353.

DuPraw, B.J. 1970. DNA and Chromosomes. Holt, Reinhart and Winston, New York.

Dustin, P. 1980. Microtubules. *Sc. Amer* **243** : 58.

Hexter, W. and Yost, H.T. 1976. The Science of Genetics. Prentice Hall, N.J.

Luyx, P. 1970. Cellular Mechanisms of Chromosome Distribution. *Int. Rev. Cytol.* **32** Suppl. 2.

Moens, P.B. 1973. Mechanisms of Chromosome Synapsis at Meiotic Prophase. *Int. Rev. Cytol*. **35** : 117.

Nicklas, R.B. 1971. Mitosis. Adv. *Cell Biol*. **2** : 225.

Schulman, I. and Bloom, K.S. 1991. Centromeres: An Integrated Protein/DNA Complex Required for Chromosome Movement. *Annu Rev. Cell Biol*. 7 : 311–336.

Voeller, B. 1968. The Chromosome Theory of Inheritance. Appleton-Century-Crofts, New York.

White, M.J.D. 1973. The Chromosomes. Chapman and Hall. London.

Willard, H.F. 1990. Centromeres of Mammalian Chromosomes. *Trends Genetics* 6 : 410–416.

7

Chi-Square Test and Probability

Mendel's 3 : 1 ratio is approximate and is not based on the study of 4 plants. Instead, a very large number of plants (in hundreds and thousands) were observed for inheritance of a character and the ratios were calculated. Thus in the cross between tall and dwarf plants, out of a total of 1064 plants, 787 were tall and 277 were dwarf, giving a ratio of 2.84 : 1 which is close to the 3 : 1 ratio. Similarly when plants with green pods were crossed with plants bearing yellow pods, the 580 plants of F_2 progeny showed segregation into 428 green and 152 yellow plants giving a ratio of 2.82 : 1. When Mendel's experiment gave a result that deviated considerably from the expected ratio, he repeated the experiment for accurate results. But it is not always possible to repeat experiments. The chi-square test is a statistical method of determining whether the deviation from the expected result is significant. Putting it in another way, what are the chances that results based on a small sample represent a true sample of a very large population ? When we use a mathematical test to determine how an observed ratio deviates from an expected ratio we say we are determining the *goodness of fit* or *chi-square.*

CHI-SQUARE TEST

For determining chi-square or goodness of fit, the size of the population must be considered. Suppose in one cross of tall and dwarf plants, out of the 100 plants of F_2, 70 are tall and 30 dwarf instead of 75 and 25 as expected from a 3 : 1 ratio. Obviously there is a deviation of 5 from the normal. In a second cross where 1000 plants of F_2 progeny showed the same numerical deviation of 5, there were 745 tall plants and 255 dwarf ones. In the third cross the 1000 plants of F_2 generation appeared in the same proportion as in the first cross (70 : 30 or 7 : 3) so that 700 plants were tall and 300 dwarf. The chi-square formula given below will show whether the differences in the observed results of the three crosses are significantly different from the 3 : 1 ratio or not (Table 7.1)

$$\chi^2 \text{ (chi-square)} = \sum \left(\frac{d^2}{e} \right)$$

where d represents deviation from expected ratio, e the expected ratio, and Σ is the sum. The smaller the chi-square value the more likely it is that deviation has occurred due to chance.

TABLE 1. Chi-square analysis of F_2 progenies of three crosses between tall and dwarf pea plants

No. of plants in F_2 population	*Cross no. 1* 100		*Cross no. 2* 1000		*Cross no. 3* 1000	
	Tall	*Dwarf*	*Tall*	*Dwarf*	*Tall*	*Dwarf*
Observed	70	30	745	255	700	300
Expected (*e*)	75	25	750	250	750	250
Deviation (*d*)	5	5	5	5	50	50
d^2	25	25	25	25	2500	2500
d^2/e	0.33	1	0.033	0.1	3.3	10
$\chi^2 = \sum \left(\frac{d^2}{e} \right)$	1 + 0.33 = 1.33		0.033 + 0.1 = 0.133		3.3 + 10 = 13.3	

Now to find out whether the differences between expected and observed results are due to chance alone or not, we must be familiar with two more concepts namely *degree of freedom* and *level of significance.*

Degrees of Freedom: The number of degrees of freedom is calculated as the number of classes whose value is required to describe the outcome from all classes. The concept of degrees of freedom is important in experiments and genetic ratios because one must consider the total number of observed individuals in the experiment as a fixed or given quantity. This fixed quantity is composed of one or more classes some of which are variable. In the experiment between tall and dwarf pea plants there are only two classes, tall and dwarf. As soon as the number of one class is set, the other can be determined. Thus when two classes are scored, there is one degree of freedom. In an experiment where three classes are scored, there are two degrees of freedom, and so on. The rule states that for the kind of genetic experiments described, the degrees of freedom are equal to one less than the number of classes.

Level of Significance: In the experiment described the actual ratio departs from that which is expected. We must now determine how significant is this discrepancy so that we can decide to accept or reject the results. Small discrepancies are not significant; large discrepancies are significant and lead to rejection of a result or hypothesis. Therefore values are assigned to these two kinds of discrepancies—the large discrepancies are the largest 5% and small discrepancies are remaining 95%. On this basis if the discrepancy lies in the large class it is significant and the result may be discarded. The 5% frequency value that enables us to reject the result is called the 5% *level of significance.* The level of significance can be changed. If 5% is too high we can decide on a low level of significance say 1%. In this case it is not so easy to reject a result. Contrarily, if we decide on a high level of significance say 10%, it is easier to reject a result. Usually the accepted level of significance is between the two extremes, that is 5%.

Chi-square: After determining the degrees of freedom in an experiment and deciding on the level of significance, the actual size of the discrepancy between expected and observed is found by chi-square. Statisticians have prepared tables that relate the number of degrees of freedom with the probability that particular groups of chi-square values will be found (Table 2). For a more detailed table refer to Table IV in Fisher and Yates, 1963.

We can now examine the results of the experiment described in Table 1. The chi-square values of the first two crosses are 1.33 and 0.133. Both are acceptable discrepancies because these values are smaller than the chi-square value for one degree of freedom given as 3.84 in Table 2. The results of the first two crosses therefore may be considered to be consistent with Mendel's hypothesis, the difference between expected and observed being due to chance.

TABLE 2. Values of chi-square corresponding to 5% and 1% levels of significance and up to 10 degrees of freedom

Degrees of freedom	*Level of significance*	
	5%	*1%*
1	3.84	6.64
2	5.99	9.21
3	7.82	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21

PROBABILITY

In the experiments described above the F_2 generation segregates in a particular ratio. The numbers of individuals in each ratio result from *chance* segregation of genes during gamete formation, and their chance combinations to form zygotes. Since these are chance events, accurate predictions about the results cannot be made. This is especially true in cases where the progeny is limited to a small number such as in experiments in animal breeding and even more so in human pedigree studies. It is hardly possible to predict the appearance of a certain phenotype or genotype. But what we can say is that there is a certain probability of occurrence of a given genetic event.

In a general way we can say that the probability or chance that an event will occur can be defined as the proportion of times in which that event occurs in a very large number of trials. If there are x trials and the event occurs on the average y times during the x trials, then the probability is expressed as y/x. The probability would be between zero and one. The closer the probability is to zero, the less chance is there for the event to occur. When probability is one, the event is certain to occur.

Rules of Probability

The multiplication rule: While considering probabilities it is important to note that the inheritance of certain genes such as for height of pea plants (tall or dwarf), flower colour (red or white), seed texture (round or wrinkled) are independent events. If we consider each gene separately, the probability of any one F_2 plant being tall is 3/4, and of its being dwarf 1/4.

Similarly there is a probability of 3/4 that an F_2 plant will bear red flowers and 1/4 that the flower will be white. Now what is the probability that an F_2 plant would be both tall and coloured? Assuming that the inheritance of each gene is an independent event, the probability that a plant be coloured and tall is equal to the product of their individual probabilities, *i.e.*, 3/4 × 3/4 = 9/16. Thus there are 9 chances out of 16 that an F_2 plant be coloured and tall. Stated in a general way, when the probability of an event is independent of that of another event, and the occurrence of one does not influence the occurrence of the other, probability that both events will occur together is the product of their individual probabilities.

The Addition Rule: The probabilty that *one* of several mutually exclusive events will occur is the *sum* of their individual probabilities. This law is applicable when different types of events cannot occur together. If one occurs the other is *excluded.* When a coin is tossed there are two alternatives, either heads or tails will appear. If the probability for heads is 1/2, for tails 1/2, then the probability that either heads or tails will appear is 1/2 + 1/2 = 1.

Probability and Human Genetics

In some recessive genetic traits in humans such as phenylketonuria (PKU), albinism and others, the birth of an affected child indicates that the parents are heterozygous carries for that trait. Due to the recessive nature of the gene the parents are healthy and normal. Such parents could be anxious to know what chances exist for any of their future children to show the genetic defect. Since the trait is recessive they could expect affected children and normal children in the ratio 1 : 3. But since human families are limited in size, this ratio can be misleading. All we can say is that the chance or probability of having an affected child is 1/4 at each birth. Moreover, even after the birth of an affected child, the same probability (1/4) exists for an affected child in all future pregnancies. The probability for a normal child at each birth is 3/4.

The rules of probability can be applied for predicting the ratio of boys and girls born in a family. Since the human male produces an equal number of *X* and *Y* sperm, the chance for a boy at any birth is 1/2, and for a girl also is 1/2. From the probability of each single conception it is possible to calculate the probability of successive births *together.* For example, what is the probability that the first two children born in a family will *both* be males. To find this out we must determine the product of the separate probabilities at each conception, that is 1/2 × 1/2 = 1/4. Now consider a different question. What is the probability that the third child in the family in which the first two are males, will also be a male ? For the answer we must remember that the sex of any child is *independent* of the sex of the other children; therefore the probability for the third child to be male is 1/2.

BINOMIAL EXPANSIONS

There are many situations in genetics where we would like to know the probability that a combination of events will occur. For instance it may be required to determine the probability that two offsprings of a mating between *Aa* and *aa* parents will have a particular genetic constitution, namely 2 *Aa*, 2 *aa*, and 1 *Aa* and the other *aa*.

The occurrence of a particular genotype in a single offspring is an independent event, as it is not influenced by the genotype of any other offspring. The probability that 2 *Aa* offspring will be formed from this mating is therefore equal to the product of their separate probabilities.

$$Aa = 1/2 \times 1/2 = 1/4 \quad \text{or} \quad 25\%$$

Thus the probabilities for each sequence of two offsprings are as follows:

1st offspring	2nd offspring	Probabilities
Aa	*Aa*	1/2 × 1/2 = 1/4
AA	*aa*	1/2 × 1/2 = 1/4
aa	*Aa*	1/2 × 1/2 = 1/4
aa	*aa*	1/2 × 1/2 = 1/4

Thus the probability that both offspring are *Aa* is 1/4, that one is *Aa* the other *aa* is 2/4, that both are *aa* is 1/4. In other words the pattern for this distribution is 1 : 2 : 1. This also represents the coefficients of raising two values of the binomial p and q to the power of 2:

$$(p + q)^2 = 1p^2 + 2pq + 1q^2$$

or if we substitute *Aa* for p and *aa* for q

$$[(Aa) + (aa)]^2 = 1(Aa)\,(Aa) + 2(Aa)\,(aa) + 1(aa)\,(aa)$$

Note that the probability for a complementary event such as the probability that both offspring will not be *aa* is 1 – the probability of the particular event, or 1 – 1/4 = 3/4. If the probabilities are calculated for the different combinations of genotypes possible among 3 children of the mating *Aa* × *aa*, the frequencies of each combination will be found to correspond with raising a binomial to the third power.

Probability that 3 offspring are *Aa* = 1/8

Probability that two are *Aa* and one *aa* = 3/8

Probability that two are *aa* and one *Aa* = 3/8

Probability that 3 offspring are *aa* = 1/8

or $$(p + q)^3 = 1p^3 + 3p^2q + 3pq^2 + 1q^3$$

or $$[(Aa) + (aa)]^3 = 1(Aa)\,(Aa)\,(Aa) + 3(Aa)\,(Aa)\,(aa) + 3(Aa)\,(aa)\,(aa) + 1(aa)\,(aa)\,(aa)$$

The probability for each particular combination of offspring can therefore be determined by the binomial coefficient for that combination relative to the total number of possible combinations. In general we can say that when p is the probability that a particular event will occur and q or $1 - p$ is the probability of an alternative form of that event so that $p + q = 1$, then probability for each combination in which a succession of such events may occur is described by the binomial distribution.

MULTINOMIAL DISTRIBUTIONS

In the examples described above there are only two possible alternatives for each event. Sometimes a genetic cross or mating can produce three types of offspring. For instance *Aa* × *Aa* will produce *AA*, *Aa* and *aa* offspring in the ratio 1 : 2 : 1 so that their probabilities are 1/4 for *AA*, 1/2 for *Aa* and 1/4 for *aa*. In such cases an additional term is added to the binomial to represent the third class of offspring. We now have a trinomial distribution $(p + q + r)^n$, where p, q and r represent probabilities of *AA*, *Aa* and *aa* respectively.

For determining probabilities of trinomial combinations we use the formula

$$\frac{n}{w \times y}\ p^w\ q^x\ r^y$$

where w, x and y are the numbers of offspring of each of the three different types, and p, q and r are their probabilities respectively. In a mating of $Aa \times Aa$ where only *four* offspring are produced, the probability of having exactly 1 *AA* homozygote, 2 *Aa* heterozygote and 1 *aa* homozygote would be

$$\frac{4}{1 \cdot 2 \cdot 1}\ (1/4)^1\ (1/2)^2\ (1/4)^1 = (24/2)\ (1/64) = 3/16$$

Other multinomial distributions can similarly be worked out. For example when a dihybrid cross is made in which four phenotypes can appear with frequencies *p, q, r* and *s*, the formula would be $\frac{n}{w \times yz}\ p^w\ q^x\ r^y\ s^z$.

As an illustration, consider the cross between smooth and yellow pea plants with wrinkled green plants described in Chapter 1. When the F_1 heterozygote is self-fertilised, 19 offspring are obtained of which 8 are smooth yellow (probability 9/16), 5 are smooth green (probability 3/16), 4 wrinkled yellow (probability 3/16) and 2 wrinkled green (probability 1/16). The probability of having the 19 offspring in exactly this ratio (9 : 3 : 3 : 1) would be

$$\frac{8}{8 \cdot 5 \cdot 4 \cdot 2}\ (9/16)^8\ (3/16)^5\ (3/16)^4\ (1/16)^2 = 0.000574.$$

QUESTIONS

1. In a plant heterozygous for striped leaves (*Ss*) what is the probability that a pollen grain will receive the *S* gene?
2. In cattle hornless condition is dominant over horned. What is the probability that the matings of two heterozygous cattle should result in:
   (*a*) a calf with horns;
   (*b*) a hornless calf.
3. At each conception the probability for a child being male is 1/2, and of female also 1/2. If a couple have three girls, what is the probability that their fourth child would be a boy?
4. In a plant heterozygous for purple flower colour and round seed (*PpRr*) what is the probability that a pollen grain will contain:
   (*a*) the *P* allele;
   (*b*) *P* and *R* alleles;
   (*c*) *P* and *r* alleles;
   (*d*) the *r* allele.
5. *Drosophila* has 4 chromosomes. Considering any three genes at random, they will have equal chance to be located on any of the four chromosomes. What is the probability that all three should be present on the same chromosome?

SELECTED READINGS

Bailey, N.T.J. 1959. Statistical Methods in Biology. John Wiley, New York.

Dunn, O.J. 1964. Basic Statistics, A Primer for the Biomedical Sciences. John Wiley, New York.

Kempthorne, O. 1957. An Introduction to Genetic Statistics. John Wiley, New York.

Strickberger, M.W. 1976. Genetics. Macmillian, New York.

Winchester, A.M. 1977. Genetics. Houghton, Mifflin, Boston.

Fisher, R. A. and Yates, F. 1963. Statistical Tables for Biological, Agricultural and Medical Research, 6th ed. Oliver and Boyd, Edinburgh.

8

Linkage, Recombination and Gene Mapping

In Mendel's dihybrid experiments the F_1 double heterozygote always showed independent assortment of the two pairs of genes at the time of gamete formation. In fact the law was justified only because the two genes were not linked to each other. This was true because each gene was located on a different chromosome. Mendel studied seven characters and the pea plant has only seven pairs of chromosomes. It followed therefore that in Mendelian inheritance one character must be located on one chromosome.

Sutton in 1903, working on the Chromosome Theory of Heredity pointed out that the number of characters which obeyed Mendel's laws when studied singly in monohybrid crosses, was much more than the number of chromosome pairs to which they could be assigned. This means that there must be many genes located on the same chromosome. If we focus attention on any two genes located on the same chromosome and perform a dihybrid cross, we cannot expect the two genes to assort independently as in Mendelism. On the contrary the two genes must be linked and show a tendency to be inherited together. If this is true, then the two phenotypes controlled by these genes must also appear together in an individual more often than if there was independent assortment. Stating it in another way, genes on the same chromosome are linked and tend to be transmitted together in a single unit. Obviously then the 9 : 3 : 3 : 1 ratio typical for a dihybrid Mendelian cross would be expected to become modified for two genes that are not assorting independently due to linkage.

Indeed Bateson, Saunders and Punnett (1905, 1906) found results of a dihybrid cross in sweat peas different from those expected in independent assortment. The experiment involved two characters, flower colour (purple vs. red) and pollen shape (long vs. round). When two varieties of sweet peas, one purple long the other red round were crossed, the F_1 progeny did *not* appear in the expected 9 : 3 : 3 : 1 ratio. Instead, plants with purple long and those with red round combinations were more frequent than expected. It will be noticed that these are the same combinations that were present in the parents. The new combinations purple round and red long which were not present in the parents were observed in F_2 with lesser frequency (Fig. 8.1). The fact emerges that when genes are linked, parental combinations occur more frequently than recombinations.

P:	Purple long	×		Red Round
	PPLL			*ppll*
F_1:		Purple Long	×	self
		PpLl		
F_2:	Purple long	Purple Round	Red long	Red Round
Observed	1,528	106	117	381
Expected	1,199	400	400	133
	(9)	(3)	(3)	(1)

Fig. 8.1 Dihybrid cross in sweet peas showing linkage.

The same cross was repeated by Bateson and Punnett in a different way. This time the parents had a different combination of characters namely Purple Round and Red Long. The F_1 plants were all Purple Long. The F_2 again showed the parental combinations (Purple Round and Red Long) in a higher frequency, and the recombinations (Purple Long and Red Round) in a lower frequency than expected.

Bateson and Punnett applied the testcross method to the above mentioned crosses in sweet peas. They crossed the F_1 double heterozygote (Purple Long *PpLl*) with the double recessive parent (red round *ppll*). In Mendelian inheritance such a cross indicates that F_1 plants are producing four types of gametes in equal frequency, which combine with the single type of gamete from the recessive parent to produce F_2 progeny of four types in the ratio of 1 : 1 : 1 : 1. Instead, the actual ratio observed by Bateson and Punnett was 7 : 1 : 1 : 7.

The above experiment clearly indicates that the two genes for flower colour and pollen shape are located on the same chromosome. The genes are said to be linked. During meiosis linked genes tend to pass *en bloc* to the same gamete and are responsible for the appearance of parental combinations in the resulting progeny. The small number of recombinations in the progeny are due to the fact that some amount of recombination does take place between the linked genes at the pachytene stage of meiosis. More about this will be discussed later. Incidentally, none of the dihybrid crosses performed by Mendel had linked genes. Genes for all seven characters studied by him showed random assortment. It was later demonstrated cytologically that sweet pea has seven pairs of chromosomes. It was also shown genetically that the gene for each of the seven characters studied by Mendel was located on a separate chromosome. Had Mendel studied some more characters that were linked, he would not have been able to interpret his results and his Principle of Independent Assortment may not have been formulated.

Bateson coined the term *coupling* for referring to the situation where two dominant alleles of a gene are both present in one parent and the two recessive alleles in the other. Thus both dominant genes pass together into one gamete in one parent and both recessives together in the gamete of the other parent as in the cross purple long (*PPLL*) and red round (*ppll*). The term *repulsion* was applied when the parents were heterozygous such as in the cross *PpLl* and *PpLl*. In this case the two dominant genes come from the different parents so that they are said to be in *repulsion*. In other words in the case of two linked loci in a double heterozygote if the two dominant alleles are on one chromosome and the two recessives on the other (*PL/pl*) the linkage relationship is said to be in the *coupling* phase. But when the dominant allele of

one locus and the recessive allele of the other occupy the same chromosome (*pl/pL*) linkage is said to be in *repulsion* phase. The terms coupling and repulsion are now of historical interest only.

MORGAN'S WORK ON DROSOPHILA

In the early decades of the twentieth century T.H. Morgan and his associates A.H. Sturtevant, H.J. Muller, C.B. Bridges and a few others at Columbia University, New York were actively engaged in studies on Chromosome Theory of Heredity with a very suitable material *Drosophila.* From their studies on mutants in *Drosophila* they could assign several genes to chromosomes. For their extensive researches on *Drosophila*, T.H. Morgan became the first geneticist to be awarded Nobel Prize in Medicine in 1934.

Some of the experiments performed in Morgan's laboratory indicated linkage because the genes did not assort independently. In *Drosophila* the normal fly has grey body and long wings. There are some recessive mutations due to which the body colour of the fly is black and the wings are underdeveloped or vestigeal. It was proposed by Bridges that in linkage experiments the wild normal type allele be represented by the sign +, and mutants recessive to wild type by abbreviated initials in small letters such as *b* for black, and *v* for vestigial. Since linked genes are present on the same chromosome, they are represented by their symbols above and below a horizontal line, genes on one homologue being above the line, those on the other below the line.

Bridges crossed a wild type female fly with grey body and long wings with a male having black body and vestigial wings. The F_1 was wild type with grey body and long wings. The females from F_1 were testcrossed with double recessive males *i.e.* black vestigial. Now if the genes are assorting independently, the test cross would yield all four combinations in the ratio of 1 : 1 : 1 : 1. But the actual results showed that the parental types (grey long and black vestigial) were more and the recombinations (grey vestigial and black long) were less than expected (Fig. 8.2). Obviously the genes were linked. One more striking feature was revealed by these experiments, that is, the absence of crossover in the *Drosophila* male. Due to this when a male fly was used as a double recessive parent in a testcross, there were recombinants present in the F_2 progeny. Reciprocal testcrosses where female flies were employed as double recessives, recombinants were absent in the F_2 progeny. The crosses provide excellent illustrations of linkage and absence of crossing over in *male Drosophila.* The *female* silk moth (*Bombyx*) is another example where crossing over does not take place.

The *Drosophila* school performed numerous experiments in their efforts to understand the chromosome theory of heredity. They had accumulated a large number of mutants for various characters. It was possible to isolate stocks of flies carrying two mutant genes for different characters on the same chromosome. Interestingly, they also found some hereditary characters which were linked to the sex chromosome *X* in females. Characters such as white eyes, yellow body, miniature wings and a few others were therefore said to be sex-linked.

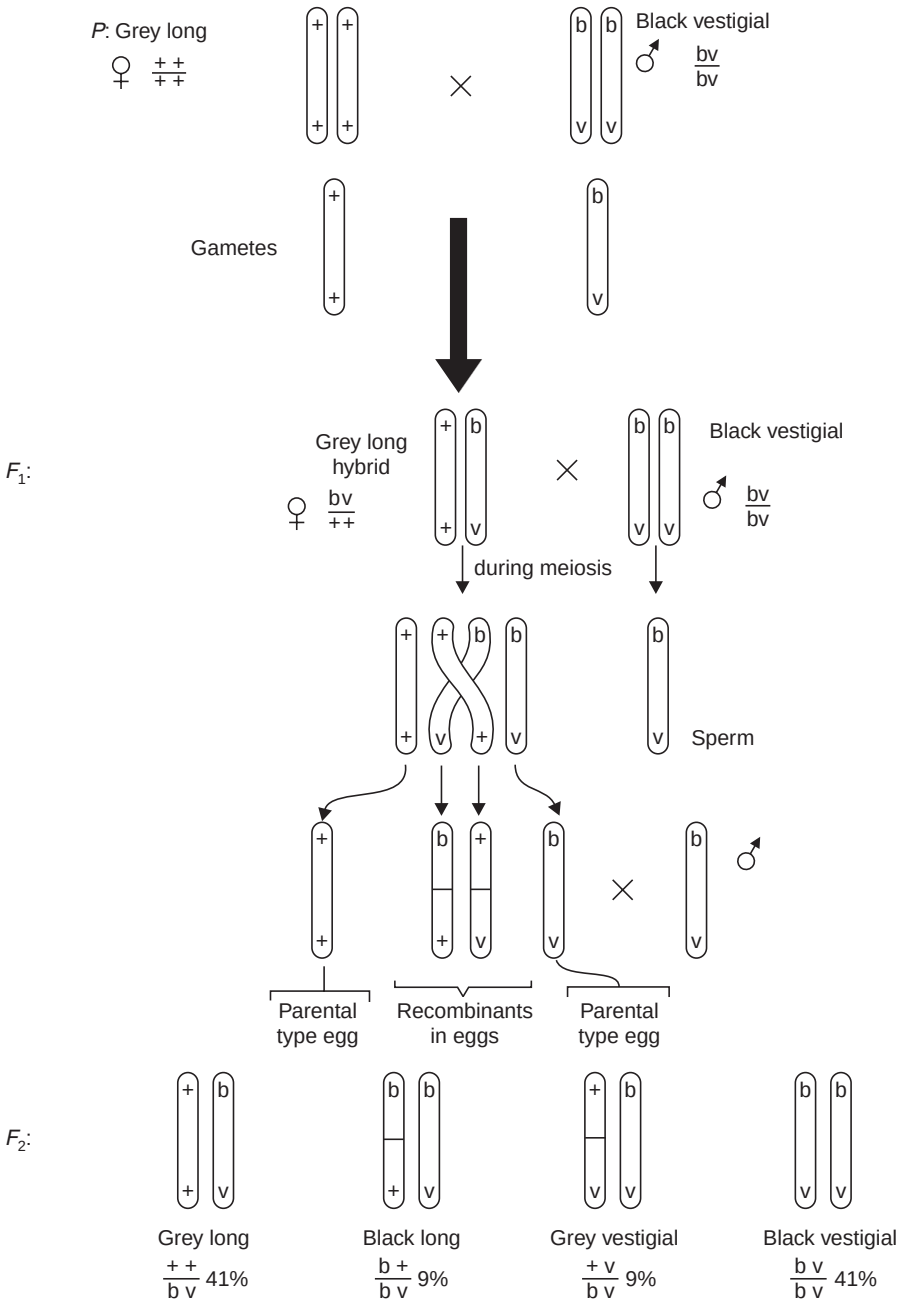


Fig. 8.2 Cross made by Bridges demonstrating linkage in *Drosophila.*

In all *D. melanogaster* was found to have four linked groups of genes. *Drosophila* has four pairs of chromosomes (Fig. 8.3). Each linkage group in *Drosophila* was associated with one pair of chromosomes and the linkage groups were numbered 1, 2, 3 and 4. Further, the number of genes in each linkage group was found to be proportional to the size of chromosomes.

Thus, the tiniest dot-like chromosomes are found to carry the smallest linkage group of only about 12 genes whereas the larger chromosomes each had more than 150 linked genes.

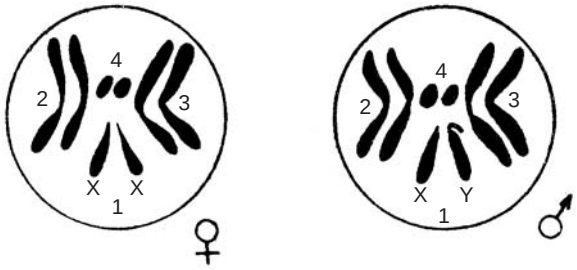


Fig. 8.3 The chromosome complement of the cell of *Drosophila.*

CROSSING OVER

If linkage is complete, there should be all parental combinations only and no recombination. But actually there is no absolute linkage, thus allowing for some recombination. How does recombination take place ? At the beginning of the 20th century, Janssens had made cytological observations on meiotic chromosomes in salamanders. He found that chromosomes showed cross-shaped configurations and suggested that they represented a break and exchange of chromosome segments. A few years later, Morgan supplemented his genetical studies on *Drosophila* with cytological observations and explained linkage on the basis of breakage and exchange in synapsed chromosomes. He could thus account for the greater frequency of parental combinations and also why linkage was not absolute so that recombinant types occurred in F_2 progeny. Morgan termed the cross-shaped configuration observed by Janssens as chiasma. The term *crossing over* referred to the actual exchange of segments between homologous chromosomes and could take place due to breakage and reunion in the paired homologues. *Recombination* is a genetic outcome of breakage and exchange of segments. It cannot be observed cytologically, but can by inferred genetically from experiments. *Crossing over* is the process of exchange of genetic segments which cannot be observed cytologically but can be estimated genetically from the frequency of recombinants in the F_2 progeny.

Chiasma Frequency

Every pair of homologous chromosomes usually forms at least one chiasma somewhere along its length. There is a characteristic average number of chiasmata for each type of chromosome. In general, the longer the chromosome the greater the number of chiasmata. Moreover, the further apart two genes are located on a chromosome, the greater the chance for a chiasma to occur between them. The percentage of crossover (recombinant) gametes formed by a given genotype indicates the frequency of chiasma formation between the genes in question. When chiasma forms in one cell between two gene loci, only half of meiotic products will be of crossover type. Therefore, chiasma frequency is twice the frequency of crossover products.

$$\text{Chiasma \%} = 2 \times \text{crossover \%}$$

$$\text{Crossover \%} = 1/2 \text{ Chiasma \%}$$

In other words, if chiasma forms between the loci of genes A and B in 30 % of the tetrads (paired homologous chromosomes) of an individual of genotype AB/ab, then 15% of the gametes

will be recombinant (*Ab* or *aB*), and 85% will be parental (*AB* or *ab*). Further, the map distance between *A* and *B* would be 15 units apart.

It is noteworthy that the *Drosophila* male shows complete linkage due to absence of crossing over. In the cross between a normal red-eyed long-winged fly and purple vestigial, the F_1 hybrids are all red-eyed and long-winged. If heterozygotes from the F_1 progeny are used as *male* parents and backcrossed with purple vestigial females, only two phenotypes appear in the progeny: the homozygous purple vestigial and the heterozygous red-eyed long-winged. The recombinations are absent demonstrating absence of cross over and presence of complete linkage in *Drosophila* male.

If a reciprocal testcross is performed using F_1 heterozygotes as females and purple vestigial as male, recombinants appear in the progeny.

Frequency of Crossing Over

In experiments on linkage, the proportions of parental phenotypes and the new combinations can be counted. From the percentage of recombinants the amount of crossing over can be calculated. In the cross between purple long and red round sweet peas described earlier, the sum of the new combinations (106 + 117) = 223 when divided by total progeny (1528 + 106 + 117 + 381) = 2132 and multiplied by 100 indicates 10.4% recombination or frequency of crossing over in F_1 gametes. In this cross the parental types equal 89.6%.

The fact that recombinants occurred in F_2 indicates that the distance between genes for flower colour and pollen shape allowed crossing over to take place between one parental chromosome and its homologue from the other parent.

Now if distance between two linked genes is more, there are greater chances of chiasma formation between them resulting in higher percentage of recombinants in the progeny. Contrarily, if distance between linked genes is less, chiasma formation between them will be less, and corresponding reduction in the number of recombinants in the progeny. Thus it is possible to locate genes on chromosomes on the basis of their crossover frequency. The distance between genes is measured in *map units.* According to Sturtevant one map unit is equal to 1% crossing over. In other words, if one gamete out of 100 gametes carries a crossover chromosome for two linked genes, we say that the two genes are one map unit apart.

The correlation between crossing over and distances between genes may not be true for all genes on a chromosome. This is because chiasma formation does not occur at random throughout the length of a chromosome. It occurs with different frequency in different parts of the chromosome. In the two large chromosomes of *Drosophila* there is less crossing over near the centromere and towards the ends, but more in the middle of the two arms. This would suggest that genes near the centromere are closer when actually they are further apart than the crossover percentage indicates.

After determining positions of several linked genes in *Drosophila*, Morgan hypothesised that genes occurred in linear order along the length of the chromosome. The position of each gene was called the *locus.*

THE THREE-POINT CROSS

When two genes are mapped by performing a cross, it is called a two-point cross. The three-point cross involves three pairs of linked genes, is a valuable method for determining positions of genes in relation to each other and for mapping distances between genes. In 1926, Bridges and Olbrycht used this method for mapping three recessive sex-linked genes in *Drosophila*:

scute *sc* (without bristles), echinus *ec* (rough eyes), and crossveinless wings *cv* (absence of transverse veins). The cross involved mating of a scute crossveinless fly with an echinus fly, and then testcrossing female F_1 heterozygotes with the recessive hemizygous male showing all three recessive phenotypes (Fig. 8.4).

In the data of Bridges and Olbrycht clearly the first group of progeny $\frac{(\ +\ +\ ec\ )}{(\ sc\ cv\ +\ )}$ represents the parental combinations, and the second two groups $\frac{(\ +\ cv\ +\ )}{(\ sc\ +\ ec\ )}$ and $\frac{(\ +\ cv\ ec\ )}{(\ sc\ +\ +\ )}$ are recombinations.

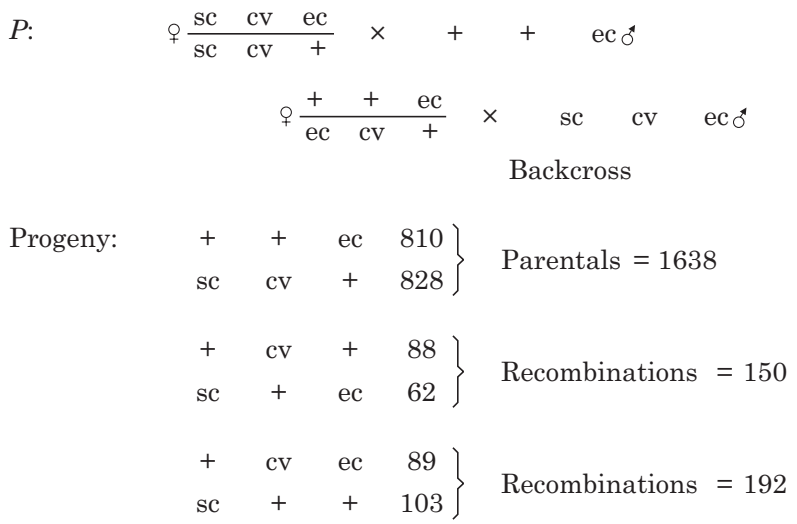


Fig. 8.4 The three point cross in *Drosophila*. Data from Bridges and Olbrycht (1926).

Since chiasma formation takes place between linked genes, in order to determine crossover percentage in a three point cross, the genes must be analysed two at a time, ignoring the third gene each time. Thus we can examine three combinations: *cv* – *ec*, *cv* – *sc*, and *ec* – *sc* in each group of progeny. Considering genes *cv* – *ec*, these are found to be present in progeny of the first group in the same way as in the parent. That is *cv* is present in its normal wild form (+) on one homologue along with *ec*; in the other homologue *cv* is present along with the wild form of *ec*. Thus in the first group of progeny *cv* and *ec* are present as *cv* + and *ec* + . Since this is the way they are present in the parent, we can infer that there is no recombination between *cv* and *ec*. The second group of progeny again shows *cv* + on one homologue and *ec* + on the other as in the parents. Therefore, the second group of progeny also does not represent recombination between genes *cv* and *ec*. In the third group of progeny both *cv* and *ec* are present on one homologous chromosome whereas the other homologue has wild forms (+ +) of both *cv* and *ec*. This arrangement of *cv* and *ec i.e.*, + and + is different from that in the parental combination and has arisen due to crossing over between the loci for *cv* and *ec*.

The data in Fig. 8.4 shows that there were 192 flies in the third group of progeny. We can calculate the percentage recombination between *cv* and *ec* by dividing 192 by the total progeny (1638 + 150 + 192 = 1980) and multiplying by 100. The amount of crossing over between *cv* and *ec* is thus found to be 9.7%.

Let us now consider genes *sc* and *cv*. In the first group of progeny they are present as in the parental combinations, that is ++ on one chromosome and as *sc cv* on the other. In the second group of progeny their arrangements are different, *sc* + being together in one chromosome and *cv* + together on the other homologue. Their changed positions indicate recombination between *sc* and *cv*. In the third group of progeny also *sc* and *cv* are present differently than in parental combinations. Thus all the 150 members of second group of progeny and 192 of the third group represent recombination between *sc* and *cv*. The recombination percentage between *sc* and *cv* is calculated as described earlier, *i.e.,*

$$\frac{150 + 192 \times 100}{1980} = 17.3\%$$

Analysing the genes *ec*–*sc* in the same way we find that in the first group of progeny they are present as in the parental combinations. In the second group they are present as ++ on one chromosome, and as *sc ec* on the other *i.e.* they represent recombination. In the third group of progeny they are present as in the parental combinations. Therefore, percentage recombination between *sc*–*ec* is $\frac{150 \times 100}{1980}$ = 7.6%. We have thus found that percentage recombination between scute and crossveinless is 17.3, between scute and echinus 7.6 and between crossveinless and echinus 9.7. The recombination percentage also represents crossover percentage and map distance between the genes.

With the data available, it is now possible to map the genes. The chromosome is drawn as a line and the two genes showing lowest recombination frequency are marked first (Fig. 8.5*a*). In this case they are scute and echinus 7.6 map units apart. Next mark *cv* and *ec* which are 9.7 units apart by indicating *cv* either on the left or right of *ec*. If we mark *cv* 9.7 map units on the left of *ec*, then as seen in Fig. 8.5*b,* the distance between *cv* and *sc* would equal 9.7 minus 7.6 equal to 2.1 units. This does not agree with the data on recombination percentage found experimentally. But if we mark *cv* on the right of *ec* 9.7 units apart, then it indicates a map distance between *sc* and *ev* equal to 9.7 plus 7.6 *i.e.*, 17.3 which is identical to the experimental data.

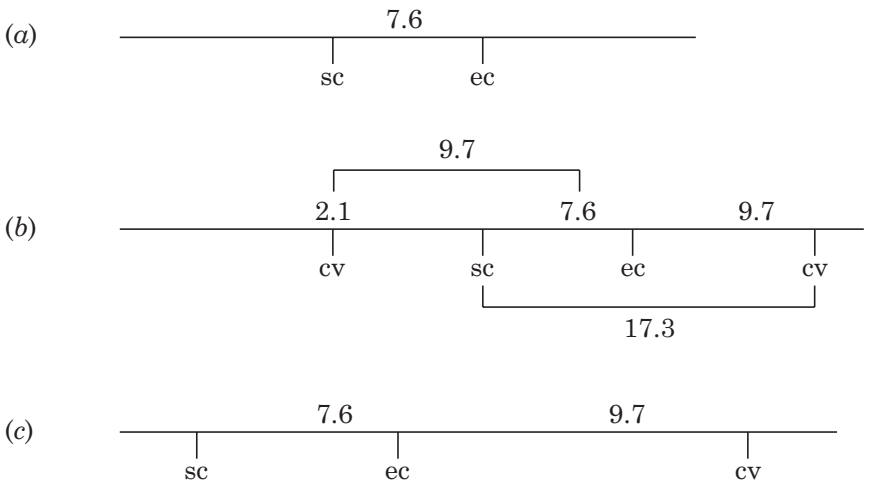


Fig. 8.5 Mapping genes involved in a three-point cross on a chromosome.

DETECTION OF LINKAGE

Whether there is linkage or independent assortment between genes can be detected by backcrossing the dihybrid or heterozygote with the recessive parent and different results would be obtained. This method allows expression of all recessive genes in the heterozygote. The number of parentals and recombinations (noncrossovers and crossovers) in the resulting progeny determines the presence of linkage (Fig. 8.6).

P: AaBb × aabb

Gametes: AB, Ab, aB, ab ab

F_1: 1/4 AaBb: 1/4 Aabb: 1/4 aaBb: 1/4 aabb

Above cross indicates independent assortment

P: AB/ab × ab/ab

Gametes: AB, ab ab

F_1: 1/2 AB/ab : 1/2 ab/ab

Same cross as above but with linked genes.

Deviations from 1 : 1 : 1 : 1 in testcross progeny indicate linkage.

Fig. 8.6 Cross illustrating the backcross method of detecting linkage.

DOUBLE CROSSING OVER

In a three-point cross involving three genes *A*, *B* and *C* there are eight possible combinations of genes, namely *ABC*, *AbC*, *ABc*, *abC*, *aBc*, *Abc*, *abc*. Sometimes one or more of the expected combinations do not appear in the progeny. This is due to two crossovers occurring simultaneously in two regions (Fig. 8.7).

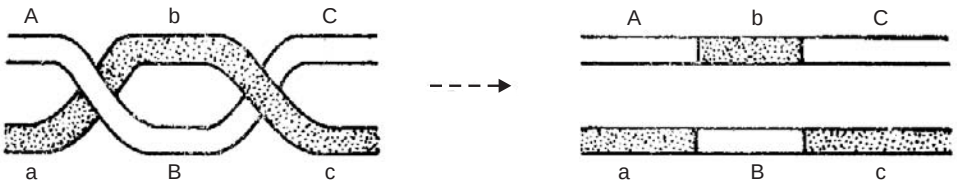


Fig. 8.7 The result of double crossovers between genes A and B and C and D.

When a single crossover occurs, two genes are exchanged resulting in the formation of a crossover gamete. But if at the same time a second crossover also takes place between the next two genes, the original combination of genes is restored on each chromatid resulting in a parental combination. It happens then that in such a cross a double crossover is represented as a noncrossover, giving a recombination frequency lower than the actual. This also reveals that Sturtevant's statement that 1% crossing over equals one map unit is not always justified. Clearly, crossover percentage is not always equal to recombination percentage. When there are double crossovers between same two chromatids, the number of recombinants in the progeny is less than the number of crossover gametes. The crossover percentages are important for mapping

genes accurately. Some geneticists prefer to use *morgan* for map units, one morgan being equal to one per cent recombination frequency and one centimorgan equal to 0.01 morgan.

Double crossovers cannot occur between genes that are located close to each other. In *Drosophila* it has been found that double crossovers cannot occur between genes closer than 10 or 15 map units apart. Moreover, the class of progeny that occurs least frequently represents the double crossovers. It also indicates greater map distance between two genes. Maximum frequency of double crossovers can occur between gene loci at each end of the chromosome. In any case, more than 50% recombination cannot be expected between two genes because only two of the four chromatids in a paired meiotic bivalent are involved in a crossover.

Interference

Sturtevant pointed out that certain parts of chromosomes were more liable to exchange segments than others. Thus if we consider three hypothetical genes *A*, *B* and *C* the probability of a crossover occurring between *A* and *B* may be 10%, and between *B* and *C* may be 15%. But what is the probability that two crossovers between *A* and *B* and between *B* and *C* should occur simultaneously ? We know that the probability of two chance events occurring simultaneously is equal to the product of the individual probabilities. In the hypothetical cross stated above, the probability that two crossovers occur between *A* and *B*, and *B* and *C* would be 10% × 15% or 0.1 × 0.15 = 0.015 = 1.5%. It has also been found experimentally that the actual percentage of double crossovers is a little less than that expected theoretically. This is due to *interference*, a term coined by Muller. Accordingly, the occurrence of one crossover reduces the chance of a second crossover in its neighbourhood.

Although some explanations have been put forward for interference, both at the cytological and molecular levels, none is considered satisfactory. Muller further proposed the terms *coefficient of coincidence* to describe the strength or degree of interference. The coefficient of coincidence is equal to the ratio of the observed percentage of double crossovers to the expected percentage of double crossovers. The extent of interference is different between different pairs of genes. The value of coincidence falls and the value of interference rises when the distance between genes decreases. Based on the coefficient of coincidence interference can be described to range from absolute (no double crossovers) to partial (doubles less frequent than expected), none (doubles equal to expected frequency) or negative (doubles more frequent than expected).

CYTOLOGICAL BASIS FOR CROSSING OVER

At the zygotene stage of meiosis homologous chromosomes come together and start pairing. By pachytene pairing is stabilised, and each ribbon-like chromosome actually consists of two homologues paired (synapsed) close to each other called bivalents. Each homologue in a bivalent consists of two identical sister chromatids. Chromatids belonging to two different homologues in a bivalent are called nonsister chromatids. Due to presence of four chromatids, the pachytene bivalent is sometimes called a *tetrad*. Crossing over takes place between nonsister chromatids and involves breakage and reunion of only two of the four chromatids at a given point on the chromosomes. Figure 8.8 illustrates how one homologous pair of chromosomes goes through meiosis to form four gametes. Two of the gametes receive a chromosome with genes linked in the same way as in the parental chromosomes (*ABC* and *abc*). These gametes represent noncrossovers or parental types and are produced from chromatids that were not involved in

crossing over. The other two gametes, (*ABc*, *abC*) represent the recombinant or crossover types and were produced after crossing over and recombination between the originally linked genes.

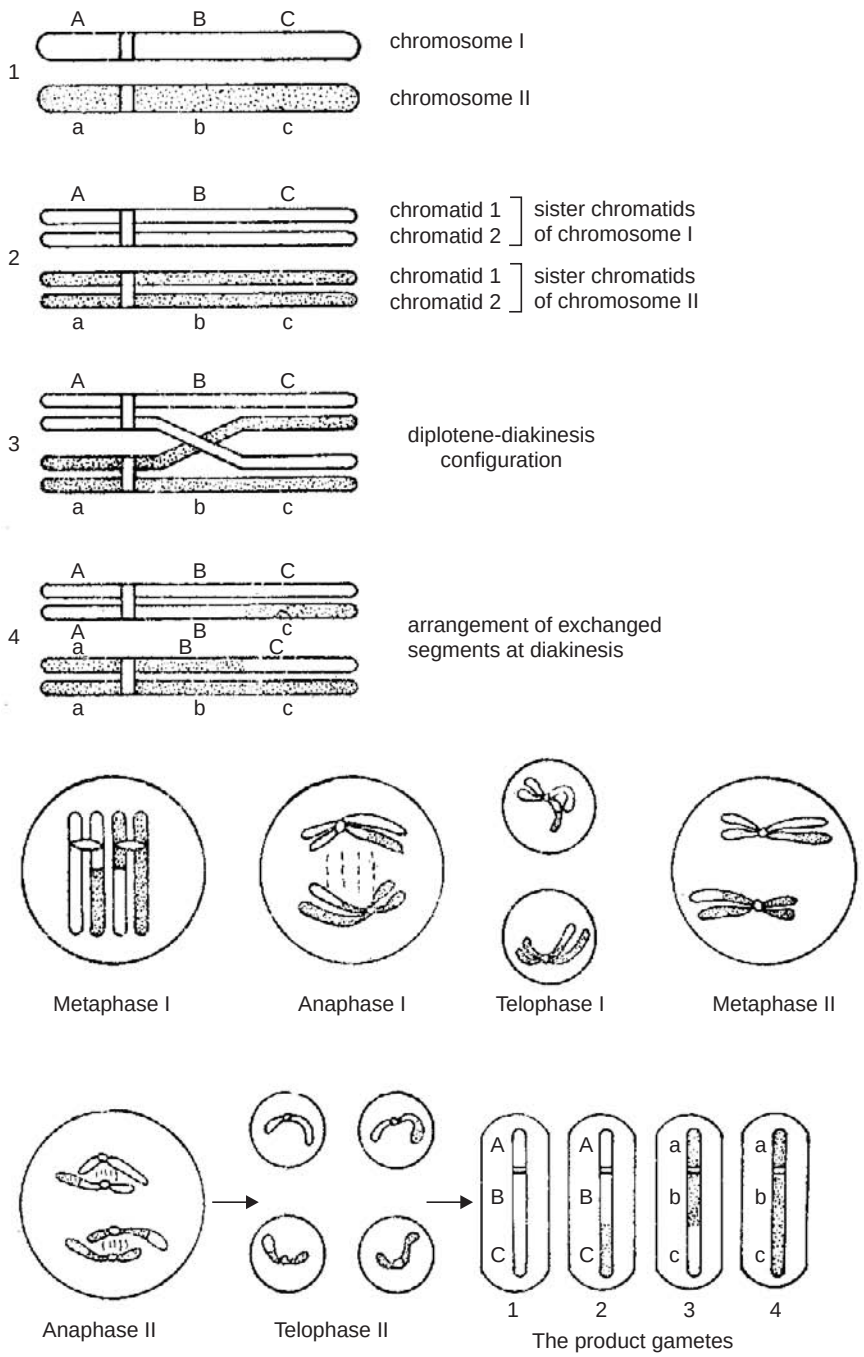


Fig. 8.8 Diagrams illustrating the cytological basis of crossing over.

SEX LINKAGE

When genes are carried on sex chromosomes they are said to be sex-linked. Inheritance of eye colour in *Drosophila* demonstrates this phenomenon. There is a white eye gene which behaves as a recessive to the normal red eye gene. In a cross between a red-eyed fly and a white-eyed fly, eye colour in the progeny depends upon which parent has red eyes and which one white. If a red-eyed female fly is crossed with a white eyed male, the F_1 consists of all red-eyed male and female flies (Fig. 8.9A).

The F_1 flies on inbreeding gave in F_2 both red and white-eyed males, and only red-eyed females. There were no white-eyed females in the progeny. But when the F_1 red-eyed flies were backcrossed with white-eyed males, the progeny consisted of both red and white-eyed males, as well as red and white eyed females (Fig. 8.9).

In the reciprocal cross in which a female fly with white eyes is crossed with a red-eyed male (Fig. 8.9B), the F_1 progeny shows both red and white-eyed flies in equal proportions. But the striking feature is that all the red-eyed flies are females, all white-eyed ones males ! The F_1 red-eyed flies on inbreeding gave the same results as F_1 reds of the above cross. But when F_1 white-eyed males were crossed with white females, they gave only white-eyed males and females.

Morgan explained his results by assuming that the gene for eye colour is carried on the *X* chromosome, *i.e.,* the gene is sex-linked. It was shown cytologically that in *Drosophila* and some other animals the male is heterogametic (*XY*) and female homogametic (*XX*). Since the *Y* carries no gene, the single *X* in the male expresses even the single recessive gene it carries as there is no dominant allele of the same gene on the *Y* chromosome to mask it. Males are therefore said to be *hemizygous.* In explanation of his cross Morgan proposed that the F_1 red-eyed females were heterozygous for white eye character. The white gene was recessive so that the flies were red-eyed. White eyes could appear in female flies that carried both genes for white (homozygous). The male flies, due to hemizygous condition could be red-eyed if the single *X* chromosome carried one red gene, or white-eyed due to a single allele for white eyes on the *X* chromosome. The results of Morgan's other crosses between red-eyed and white-eyed flies agreed with his explanations, thus providing support for his hypothesis.

Sex-Linkage in Poultry

Spillman in 1909 found sex-linked inheritance of feather colour in poultry. In the Plymoth Rock strain of chickens one variety called *barred* has black and white pigments in feathers alternating to give a barred appearance. The other variety of chicken was black or nonbarred. When a nonbarred female was crossed to a barred male, the F_1 chickens were all barred. The F_1 males and females were mated with nonbarred chickens to give a testcross progeny consisting of 4 classes: barred males, nonbarred males, nonbarred females, barred females. When barred females were crossed to nonbarred males, the testcross progeny yielded only barred males and nonbarred females (Fig. 8.10). In poultry male is homogametic (*XX*) and females heterogametic (X_0).

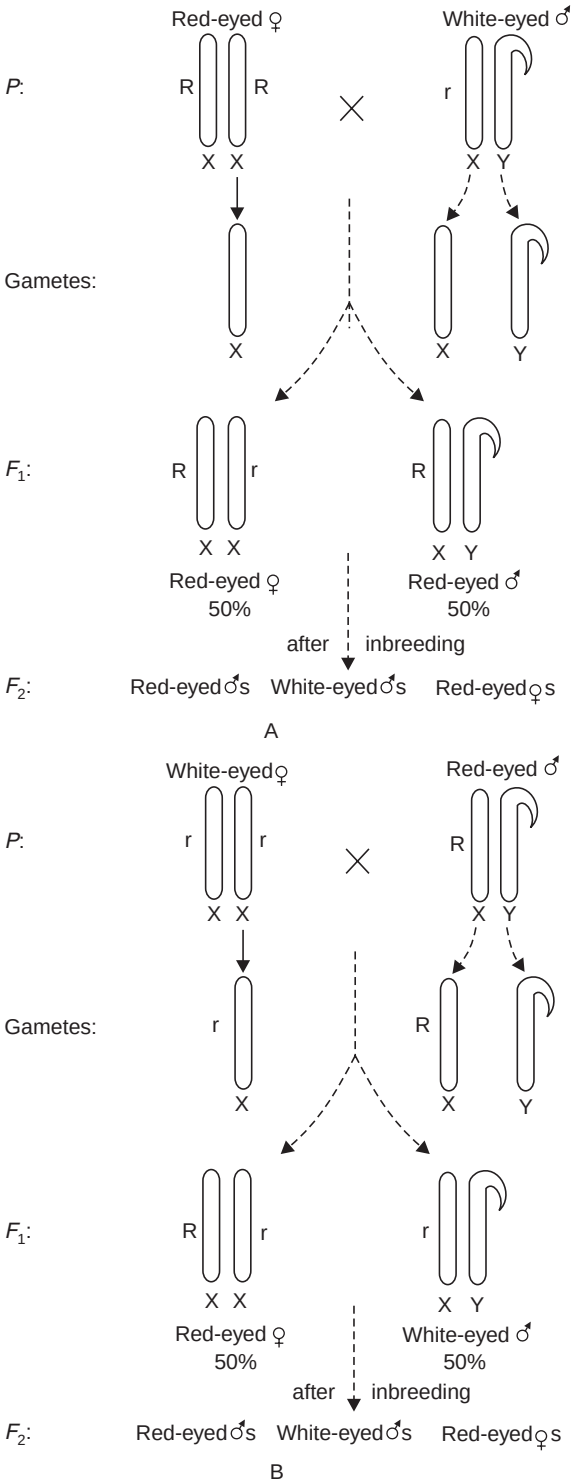


Fig. 8.9 A, B Inheritance of eye colour in *Drosophila* demonstrating absence of crossing over in male and sex-linkage.

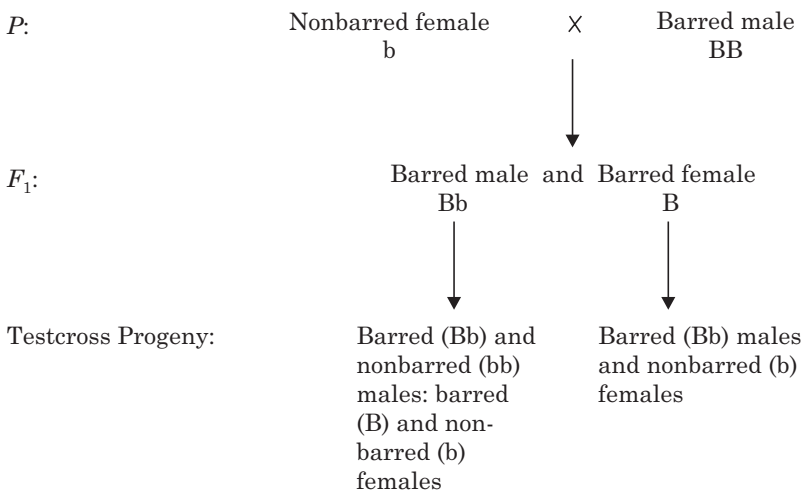


Fig. 8.10 Cross illustrating sex-linkage in poultry.

RECOMBINATION IN NEUROSPORA

The fungus *Neurospora crassa* has some advantages for study of crossing over. In the vegetative phase of its life cycle, there is growth of filamentous, *coenocytic hyphae* to form the haploid *mycelium*. Sexual reproduction is achieved by coming together of two hyphae of proper mating type and their nuclei fuse to form a diploid zygote (Fig. 8.11). The zygote starts enlarging into an elongated structure called *ascus* and the zygote nucleus starts dividing meiotically. The four products of meiosis are arranged in a single linear row inside the ascus. Each one of these undergoes a mitotic division so that 8 nuclei are formed which develop a wall and become ascospores.

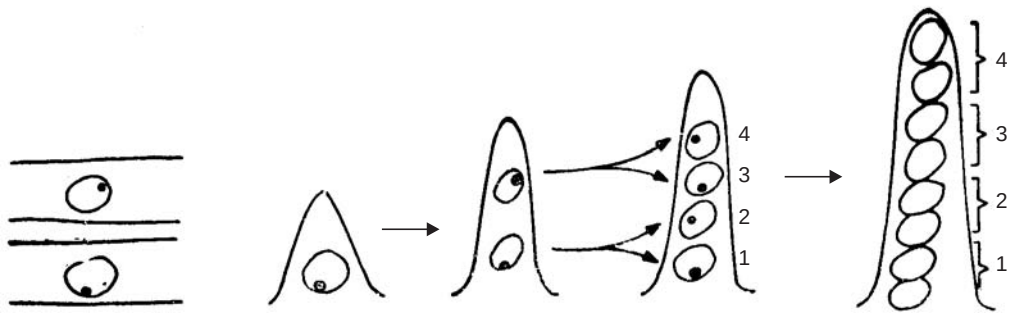


Fig. 8.11 Mating hyphae in *Neurospora* and stages in formation of ascus.

What makes *Neurospora* interesting for a geneticist is the fact that the ascospores are linearly ordered in the same sequence as the chromatids were on the meiotic metaphase plate. It is therefore possible to recover all the four products of a single meiosis and analyse them, as also each chromatid from a tetrad. This is called *tetrad analysis*. *Neurospora* thus presents a direct way of demonstrating recombination which in higher organisms can be inferred on a

statistical basis from progeny counts. A still further advantage is the use of centromere as a marker for determining map distances as explained later in this chapter.

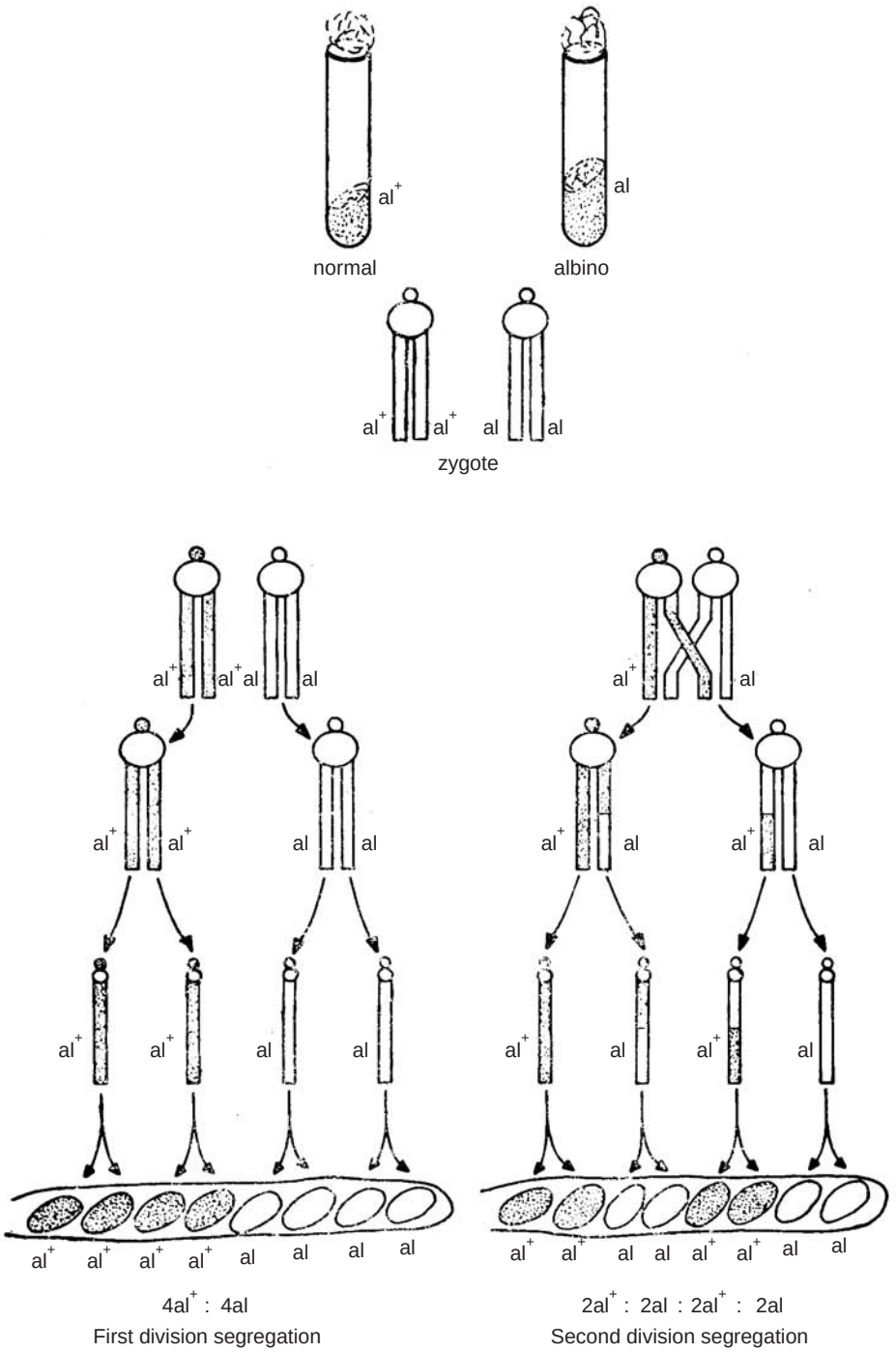


Fig. 8.12 Results of first and second division segregation in *Neurospora.*

When the segregation of a single pair of alleles is analysed, the ascus of *Neurospora* shows two arrangements of the eight ascospores: 4 : 4 ratio resulting from first division segregation; and 2 : 2 : 2 : 2 ratio resulting from second division segregation (Fig. 8.12).

First Division Segregation: The results of a cross between a normal (al^+) and an albino (al) strain of *Neurospora* are shown in Fig. 8.12. If the 8 ascospores are removed from the ascus and each is grown separately, it is found that 4 ascospores produce mycelium of normal type and 4 of albino type. This is due to first division segregation explained as follows: during anaphase I of meiosis in the zygote, one homologue carrying al^+al^+ segregates from the second homologue which has the other allele *alal* and both move to opposite poles. Their products are recovered in a 4 : 4 ratio after meiosis II (and after mitosis) because no *crossing over* has occurred between the gene and the centromere. This situation results when the gene is located close to the centromere. Therefore, if a large number of asci are found to exhibit 4 : 4 ratio, it indicates that the gene locus in question is close to the centromere.

Second Division Segregation: As shown in Fig. 8.12 if crossing over occurs between the said gene and the centromere, then one chromatid of each homologue will carry al^+ and the other al. Therefore al^+ and al will not be able to segregate from each other at anaphase of first meiosis when the two homologues separate to the two poles. It is only at anaphase of second meiosis when centromeres divide and chromatids separate from each other to the two poles that al^+ and al will segregate from each other. This is called second division segregation and produces a ratio of $2al^+ : 2al : 2al^+ : 2al$, which also indicates that crossing over has occurred between the gene locus and centromere.

GENE MAPPING IN FUNGI

In *Neurospora* the centromere is a marker for determining map distances. For detecting linkage and map distances, the frequency of crossing over is determined from the number of asci showing second division segregation. If there is one crossover, the resulting ascus shows 50% of ascospores with parental combinations and 50% with recombinations. Suppose in a cross involving a pair of alleles 30% of asci show second division segregation. This shows that 30% of zygotes had crossing over during meiosis and 70% did not. Since there are four chromatids in each tetrad, the 50% asci have resulted from $30 \times 4 = 120$ original chromatids in meiosis. When there is crossing over only *two* of the four chromatids are involved in an exchange. Therefore only half of the 120 chromatids *i.e.,* 60 are crossover chromatids, the remaining 60 chromatids are noncrossover chromatids. It was also stated above that 70% of zygotes did not have crossing over, which means that $70 \times 4 = 280$ are noncrossover chromatids. The actual number of noncrossover chromatids is larger because the 30% asci showing second division segregation also have 60 noncrossover chromatids. The exact number is therefore $280 + 60 = 340$. Therefore, of the original 100 tetrads or asci 340 are noncrossover chromatids and 60 are crossover chromatids. Since 100 tetrads or asci also mean 400 chromatids, the percentage of crossover chromatids is $\frac{60}{400} \times 100 = 15\%$. From this we can conclude that there was 15% crossing over between the gene and the centromere. We can also say that the gene in question is 15 map units apart from the centromere. Because the centromere itself serves as a marker, in *Neurospora* it is possible to map a single gene pair. It is also called a two-point cross.

It follows that the method of detecting gene linkage in fungi is basically similar to that for diploids. The main feature in all cases consists in comparing the frequency of parental types to recombinant types. If there is a significant reduction in the frequency of recombinant types from the frequencies expected on the basis of independent assortment, we can consider linkage.

Whereas in *Neurospora* the meiotic products occur in an ascus in a linear order (ordered tetrads) this is not true for other fungi in Ascomycetes which have unordered tetrads. Let us examine a cross involving two linked genes + +/*ab* in a fungus having unordered tetrads. The zygote (++/*ab*) undergoes meiosis, and depending upon the occurrence of crossover, the tetrads are of following three types:

(*a*) *Parental ditype (PD)* (Fig. 8.13*a*):

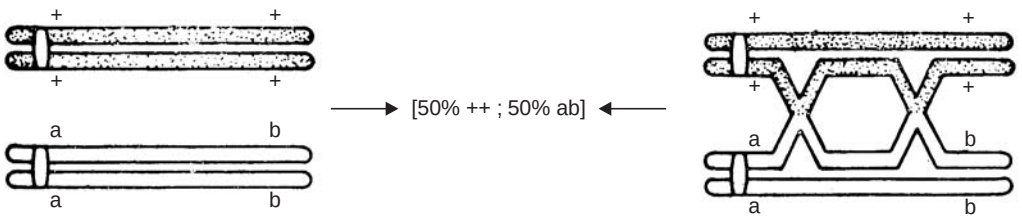


Fig. 8.13(a)

If a crossover does not occur between these two loci, or if a two strand double crossover occurs between them, the resulting meiotic products will be of two kinds, both resembling parental combinations, and appear in equal frequency (1 + +; 1 *ab*). Such a tetrad is called parental ditype (*PD*).

(*b*) *Nonparental ditype* (*NPD*) (Fig. 8.13*b*)

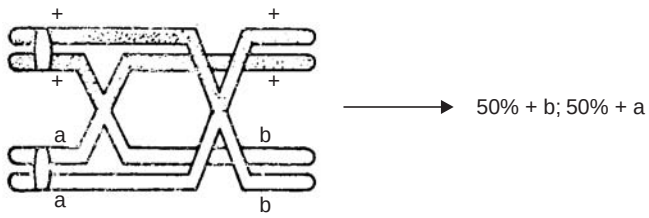


Fig. 8.13(b)

If a four strand double crossover occurs between the two genes, two kinds of products are formed, both being recombinations. Such a tetrad is called a nonparental ditype (*NPD*).

(*c*) *Tetratype* (*TT*) (Fig. 8.13 *c*): This is produced either by a single crossover or a three-strand double crossover (of two types) between the two genes.

Whenever the number of parental ditypes and nonparental ditypes are significantly unequal, linkage between the two genes must be considered. For determining the amount of recombination between the two genes, the following formula is used:

$$\text{Recombination Frequency} = \frac{NPD + 1/2\,TT}{\text{Total number of tetrads}}$$

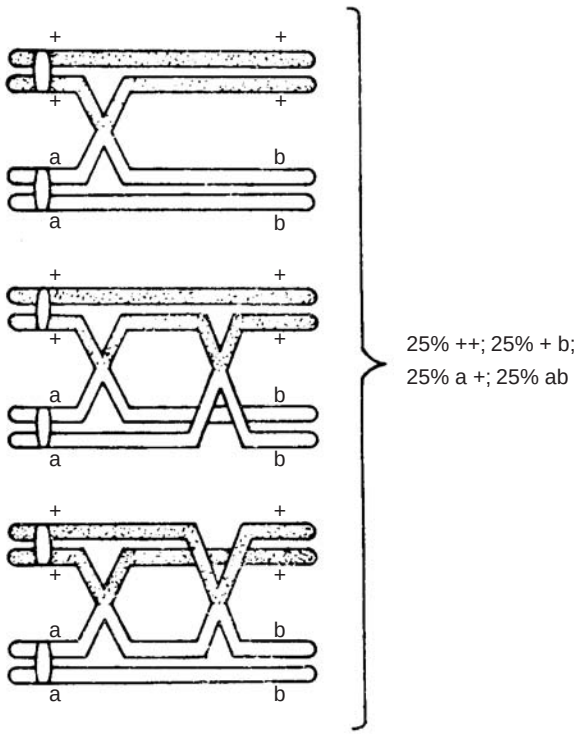


Fig. 8.13(c)

Cytological Proof of Crossing Over

In 1917 Goldschmidt proposed that recombination takes place due to exchange of alleles without exchange of chromosome segments. He assumed that at metaphase genes get detached from the chromosome. Later during meiosis the genes get reabsorbed on the chromosome either in the same or in a different place. In 1930 Winkler put forth his "gene conversion" hypothesis. Accordingly, if gene replication occurs in closely synapsed homologous chromosomes, the wrong allele may get replicated. When this occurs at only one locus it appears like a crossover. Both the above theories presented difficulties in understanding.

Bellings hypothesis: In 1931 John Belling proposed a theory of crossing over based on exchange of chromosome segments in lily plants. Belling studied the morphology of bead-like chromomeres which are arranged linearly on the chromonema. Since the structure and arrangement of chromomeres is identical in a pair of homologous chromosomes, Belling thought they might represent genes. He explained recombination by assuming that the chromomeres were synthesised first, and the chromonemata which were synthesised later, became connected to the chromomeres. Wherever the strands of chromonemata did not get connected with chromomeres in the original linear order, they crossed over and passed through another chromomere resulting in recombination and genetic exchange. The name *copy choice* was later given to Belling's mechanism for recombination. However, the idea was disproved due to lack of evidence from genetic tests (see Chapter 22).

Significance of Crossing Over

Crossing over occurs in living organisms ranging from viruses to man. It constitutes evidence for sexual reproduction in an organism. Its widespread occurrence in organisms ensures exchange of genes and production of new types which increase genetic diversity. This increases phenotypic diversity, which at the species level is responsible for genetic polymorphism. The occurrence of polymorphism is of advantage to a species because it leads to groups of individuals becoming adapted to a wider range of habitats. This increases the potential for evolutionary success.

Factors Affecting Crossing Over

The following external factors can affect the frequency of crossing over:

1. Bridges showed that in *Drosophila*, as maternal age increases, crossing over decreases.
2. H.H. Plough, a student of Morgan found that both low and high temperatures changed the frequency of crossing over.
3. The existence of crossing over factors has been shown in the cytoplasm of females. Consequently, females with reduced recombination frequencies can pass on this trait to their daughters.
4. Calcium and magnesium ions affect crossover frequency. Antibiotics such as mitomycin-*C* and actinomycin-*D* increase crossing over. Similarly *X*-ray irradiation can increase crossover frequency in *Drosophila* females and induce it in males.
5. Heat and shock treatments also change rates of crossing over. Genotypic effects were observed in *Drosophila* by R.P. and E.E. Levine by which a gene on the 3rd chromosome completely suppresses crossing over in homozygous females.

Genotypic Control of Recombination

In some cases at least recombination itself is under the control of genes. In *Neurospora crassa* several recombination genes control frequency of recombination either between genes or within genes. Thus a recessive gene *rec*-1 controls frequency of recombination at the *his*-1 locus. Both *rec*-1 and *his*-1 are on the same chromosome. In maize chiasma formation can be entirely suppressed by a recessive gene *as* present on chromosome 1. There are quite a few examples known of single genes as well as polygenes which can cause variation in frequency and distribution of chiasma in one or more chromosomes.

QUESTIONS

1. A plant heterozygous for three alleles *AaBbCc* is cross pollinated by pollen from a homozygous recessive plant *aabbcc*. Out of the 100 progeny plants raised 45 showed the phenotypes of alleles *A*, *B* and *C*. Does this indicate linkage ? Explain.
2. If all the genes in an organism are mapped, how many linkage groups would be expected to be found in

   (*a*) man; (*b*) *E. coli* (main chromosome only):(*c*) the haploid mycelium of a fungus having 11 chromosomes in each cell; (*d*) the bacteriophage.

3. In *Drosophila* genes for body colour and wing size are linked. In performing testcrosses with F_1 heterozygotes would the results be different if the double recessive parent was a female instead of male? Explain.
4. If two linked loci *C* and *G* are 40 map units apart, whereas *C* and *E* have 8 map units between them, what proportions of the gametes of a heterozygous individual would be expected to have *CG*? With what frequency would *CE* gametes be formed?
5. Explain how double crossovers can increase the frequency of parental combinations. How do they affect map distances?
6. A trihybrid *PpQqRr* is crossed. The F_1 progeny indicate that the trihybrid parent produced the following gametes.

   PQR—25 *pqr*—20
   Pqr—240 *pqR*—230
   PQr—201 *pQR*—250
   PqR—20 *pQr*—28

   (*a*) which loci are linked and which show independent assortment?

   (*b*) what is the genotype of the other parent?

   (*c*) calculate the map distances between the linked genes.
7. If an attached-*X* female of *Drosophila* is heterozygous for both genes *A* and *B*, what kind of gametes would form if (*a*) there was one crossover between *A* and *B*, (*b*) if there was no crossover.
8. In chickens females are *XO* and males *XX* and the gene for barred feathers is carried on the *X* chromosome. What would be the results of the following crosses: (*a*) barred male *x* barred female; (*b*) nonbarred male *x* nonbarred female; (*c*) nonbarred male *x* barred female.
9. You are asked to perform a cross between a pink strain of *Neurospora* and a white strain and then grow all the 8 ascospores in an ascus. What results would you expect if (*a*) first division segregation had occurred; (*b*) second division segregation had occurred?
10. If you are given the percentage of *Neurospora* asci showing second division segregation, how can you calculate the map distance of the gene from the centromere.
11. Given the following map distances between individual pairs of genes construct a map of the following chromosome.

    o–p—10 units *m–p*—15 units
    p–n—13 units *m–o*—8 units
    n–o—3 units

SELECTED READINGS

Barratt, R.W. *et al.,* 1954. Map Construction in *Neurospora crassa. Adv. Genetics* **6 : 1.**

Bridges, C.B. and Olbrycht, T.M. 1926. The Multiple Stock "Xple" and its Use. *Genetics* **11 :** 41.

Creighton, H.S. and McClintock, B. 1931. A Correlation of Cytological and Genetic Crossing over in *Zea* Mays. *Science* **17 :** 492.

Fincham, J.R.S. 1970. Fungal Genetics. *Annu. Rev. Genetics* **4 :** 347.

Fincham, J.R.S. and Day, P.R. 1971. Fungal Genetics. Blackwell Scientific Publ. Ltd.

Fu, T.K. and Sears, E.R. 1973. The Relationship between Chiasmata and Crossing-over in *Triticum aestivum. Genetics* **75 :** 231.

Henderson, S.A. 1970. The Time and Place of Meiotic Crossing Over. *Annu. Rev. Genetics* **4 :** 295.

Mather, K. 1951. Measurement of Linkage in Heredity. Wiley, New York.

Morgan, T.H. 1911. Random Segregation Versus Coupling in Mendelian Inheritance. *Science* **34 :** 384.

Muller, H.J. 1916. The Mechanism of Crossing Over. *Amer. Nat.* **50 :** 193.

Sears, E.R. 1976. Genetic Control of Chromosome Pairing in Wheat. *Annu. Rev. Genetics* **10 :** 31.

Stern, C. 1931. Zytologisch—genetische Untersuchungen an Beweise für die morganische Theorie des Faktorenaustauchs. *Biol. Zentralblatt* **51 :** 547.

Sturtevant, A.H. 1913. The Linear Arrangement of Six-linked Factors in *Drosophila,* as shown by their mode of association. *J. Exptl. Zool.* **14 :** 43.

Whitehouse, H.L.K. 1973. Towards an Understanding of the Mechanism of Heredity. St. Martin's Press, New York.

9

Extranuclear Transmission of Traits

In Mendelian inheritance the transmission of hereditary traits depends almost entirely on the chromosomes contained in the nucleus and their behaviour during meiosis. One of the exceptions to this regular pattern is the transmission of genetic information from parent to offspring through the cytoplasm and is called extranuclear or cytoplasmic inheritance. The role of cytoplasm in heredity is determined from results of reciprocal crosses in which the sources of male and female gametes are reversed. In chromosomal inheritance it makes no difference in the transmission of a gene whether it comes from male parent, and identical phenotypes are obtained in reciprocal crosses. But if there is extranuclear inheritance, the resulting phenotypes in a reciprocal cross are nonidentical and indicate uniparental transmission, in most cases from the maternal parent. The bigger role of the maternal parent results from the unequal cytoplasmic contributions of the male and female parents as most of cytoplasm of the zygote comes only from the egg. However, a few cases of paternal transmission are also known.

Sometimes confusion can arise in recognising a true case of cytoplasmic inheritance due to certain other modes of transmission which deviate from the regular pattern of nuclear inheritance. In practice, therefore, breeding tests for extranuclear inheritance are so designed as to exclude step by step all possible chromosomal explanations of the observed behaviour as mentioned below:

(*a*) Reciprocal crosses in which identical phenotypes are produced.

(*b*) Sex-linked inheritance, it follows a definite pattern and is easily recognised. It is rare in plants.

(*c*) Dauer (persistent) modifications: Sometimes maternal influence is exerted on the offspring due to nutrition, environment or certain treatments given to the female parent which affect the cytoplasm of the egg. The effects shown by the progeny and their inheritance are transient. They will either disappear in the next generation or will appear with a diminishing effect in succeeding generations. In a true case of cytoplasmic inheritance a trait must persist undiminished through many generations.

(*d*) Maternal effect due to a cytoplasmic trait under the control of nuclear genes. Inheritance of some traits is influenced by cytoplasm in the egg, but the expression of the trait is modified by chromosomal genes. Coiling in snails (dextral and sinistral), inheritance of some intranuclear symbionts and of eye colour in moth *Limnaea* are good examples. Maternal effects can also result

in a developing organism from mRNA that was included in the egg cytoplasm prior to fertilisation. The stored RNA originated from chromosomal genes. Such maternal effects usually last in the offspring for one generation only.

The term extranuclear inheritance is restricted to cytoplasmic factors that exhibit replication and independent transmission. These factors have their own genetic material and the traits controlled by their genes are not transmitted in Mendelian fashion. Because of their location outside the nucleus, these genetic factors are referred to as *plasmagenes* or *cytogenes.* They are most commonly detected from differences in the results of reciprocal crosses. These differences persist as long as the extra-chromosomal factor can perpetuate itself. They are mostly transmitted through the egg cytoplasm of the female parent because the amount of cytoplasm contributed to the zygote by the male parent is negligible. However, a few cases of male cytoplasmic inheritance are known and will be described later.

KILLER TRAIT IN PARAMECIUM

Sonneborn in 1938 discovered certain strains in *Paramecium* which showed a killer trait due to the presence of a cytoplasmic factor called kappa. The killer strain can destroy the sensitive strains growing in culture which do not have kappa by liberating a toxic substance *paramecin.* Killer strains are not killed by their own paramecin.

Paramecium has two kinds of nuclei, a small micronucleus and a very large macronucleus which is highly polyploid and irregular in shape and behaviour during cell division. Only the micronucleus behaves according to Mendelian principles. *Paramecium* has three modes of reproduction. The first is a simple mitotic division called *binary fission.* The second method is *conjugation.* Here two protozoans divide meiotically to form four micronuclei in each cell; out of these, three nuclei degenerate and only one remains which divides by mitosis to produce two genetically identical haploid nuclei in each *Paramecium.* During conjugation only one of the two haploid nuclei is exchanged through a cytoplasmic bridge formed between the two ciliates. The cells then separate as two exconjugants.

The third method of reproduction is called autogamy. Here a single *Paramecium* divides *meiotically* and by the same process that occurs in conjugation, two identical haploid nuclei are formed which fuse to form a diploid organism. As there was no genetic exchange, the diploid *Paramecium* is homozygous.

One noteworthy feature of the sensitive strains is that they are not killed by paramecin while they are in the process of conjugation. This has an advantage because it allows the investigator to perform crosses between the killer and sensitive strains. The two strains can be distinguished morphologically as killers have granular cytoplasm and sensitives are clear. When a cross is made between a killer and a sensitive *Paramecium* (each made homozygous by autogamy), there is exchange of genetic material through conjugation. This is followed by separation of the two genetically identical exconjugants. It is found that killer exconjugants produce only killer *Paramecia* and the sensitive exconjugants only the sensitive *Paramecia.* Obviously, the killer and sensitive traits are not controlled by Mendelian genes (Fig. 9.1).

If the heterozygous (*KK*) killer exconjugant is inbred to another heterozygous killer, it produces three-quarter killer (1 *KK* and 2 *kk*) and one quarter sensitives (*kk*). But if the sensitive exconjugant (*Kk*) is crossed to another heterozygous sensitive, it results in all sensitive progeny

even though their genotypes are in the ratio of 1 *KK* : 2*Kk* : l*kk*. The results suggest nonchromosomal inheritance of killer trait.

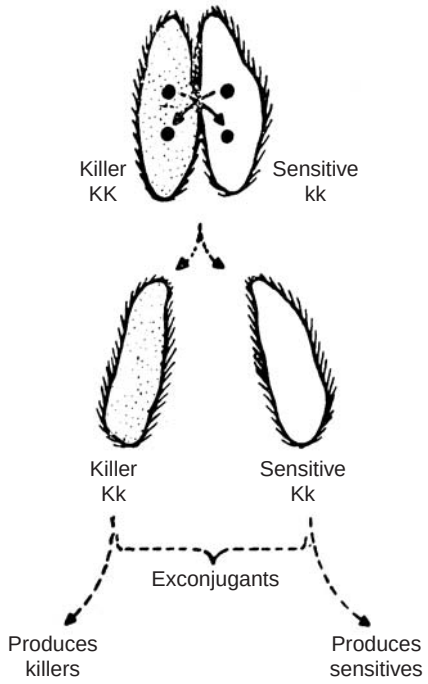


Fig. 9.1 Conjugation in *Paramecium* demonstrating extranuclear inheritance of the Killer trait.

The final proof regarding inheritance of killer trait was obtained by modifying the experiment in the following way. The cross between killer and sensitive was prolonged, allowing enough time for exchange of cytoplasm to take place. In this way, some of the kappa particles could move from killer into the sensitive strain. All the resulting progeny of such a cross consisted of killers thus confirming the cytoplasmic inheritance of kappa particles.

The above experiment can be performed in such a way that *Paramecia* divide very rapidly by controlling the nutrient conditions. Under such conditions, a homozygous killer strain (*KK*) containing kappa particles can produce a few individuals that are sensitive and without kappa particles. The explanation is that kappa particles cannot multiply as rapidly as the cells, and become fewer in number in comparison with the number of *Paramecium* cells. Due to their reduced number kappa particles are not passed on to some members of the progeny at all.

It was found that although kappa particles are transmitted cytoplasmically, yet they require a dominant *K* gene for maintenance. The *K* gene cannot initiate the presence of kappa particles. Kappa are virus-like particles about 0.2 micron in diameter and have ability to reproduce independent of the nucleus. They have their own DNA, can multiply and produce the substance paramecin.

CO_2 SENSITIVITY IN DROSOPHILA

Carbon dioxide is used as an anaesthetic for *Drosophila,* and normal flies can withstand high CO_2 concentrations without any adverse effect. But some strains of *Drosophila* are sensitive to CO_2 so

that within a few minutes they become paralysed and die. Reciprocal crosses between CO_2 sensitive and resistant flies gave differing results. If a CO_2 sensitive female is crossed with a normal (resistant) male, almost all the offspring are sensitive. But the reciprocal cross of a sensitive male with a normal female yields almost all normal offspring indicating extranuclear inheritance of this trait. There is mostly maternal transmission although some transmission through sperm may also take place. Moreover, normal flies can be induced to become sensitive by injecting a cell-free filtrate of haemolymph obtained from sensitive flies.

The causal agent for CO_2 sensitivity has been observed in electron micrographs to be a virus-like particle called *sigma*. In flies that have been induced to become sensitive, *sigma* enters the egg cytoplasm and is transmitted maternally. In the naturally occurring sensitive strains, sigma particles become incorporated into the nucleus (without associating with a specific chromosome) of the egg or some of the sperms and are transmitted in a non-Mendelian manner. They are self-reproducing independent bodies and can mutate.

PLASTID INHERITANCE

Plastids are cytoplasmic organelles which contain *DNA* and duplicate themselves independently of nuclear genes, and are distributed more or less equally to daughter cells during mitosis. While studying leaf pigmentation in *Mirabilis jalapa* (four O'clock plant), Correns (1909) found for the first time that plastids could be transmitted to the offspring through the egg cytoplasm. In a variegated strain of *Mirabilis* the cytoplasm of the zygote contains both green and colourless plastids. During cell division in the zygote to form embryo, these plastids are unequally distributed. Such an embryo on germination produces a variegated plant with three types of branches—those bearing green foliage, colourless foliage and variegated (mixture of green and white) foliage. No matter what the colour of the branch from which the pollen is used for fertilisation, it was found that seeds produced by green branches gave rise only to green plants; and seeds from colourless branches produced colourless seedlings which did not survive due to lack of chlorophyll. But seeds from the variegated branch could produce three types of progeny: plants with green, colourless or variegated branches. This is because the egg cell of a variegated plant will have both kinds of plastids—green as well as white. At the time of cell division some cells will receive only green plastids, some only white, and others will receive both green and white plastids. Likewise the progeny of such a branch could be green, white or variegated (Fig. 9.2).

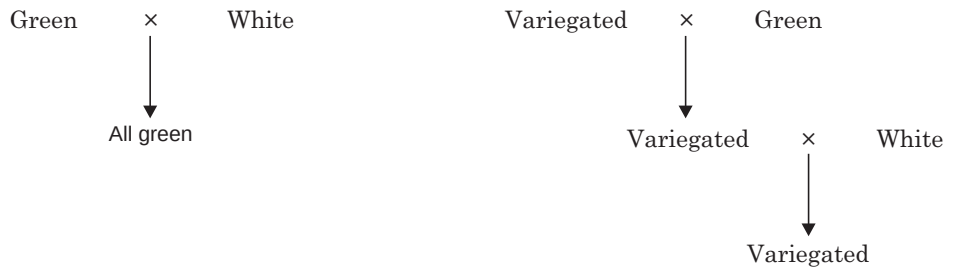


Fig. 9.2 Crosses in *M. jalapa* showing no influence of pollen.

All the progeny inherit maternal characteristics of foliage through egg cytoplasm.

MALE TRANSMISSION IN GERANIUM (PELARGONIUM)

As in *Mirabilis*, plants of geranium could be green, white or variegated. But cytoplasmic inheritance in this case follows a unique pattern. If the egg cell from a white plant is fertilised with pollen from green plants, the progeny consists of green, white and variegated plants. The question arises—when the egg cytoplasm has only white plastids, what is the source of the green plastids in the progeny? The answer revealed an unusual feature of *Pelargonium*, that is the male parent contributes plastids through the cytoplasm of the pollen grain. In *Oenothera* also male transmission of cytoplasmic traits is known.

IOJAP STRAIN OF MAIZE

The *iojap* strain of maize is characterised by green and white stripes on the leaf. The name *iojap* originates from two parental strains *Iowa* which is green and *Japonica,* a striped variety.

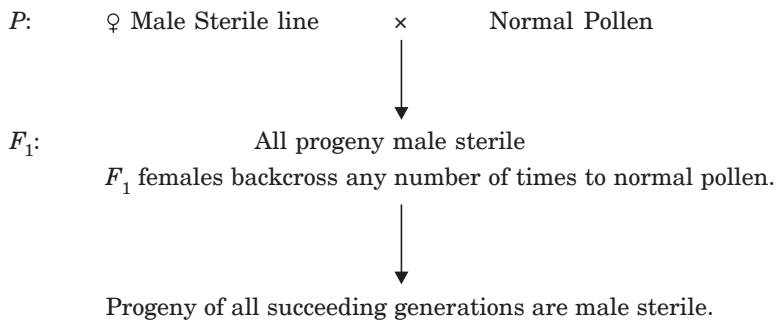
In crosses between *iojap* and green varieties when *iojap* is used as the male parent, the trait is inherited according to Mendelian pattern with F_1 progeny all green, and F_2 segregating into 3/4 green and 1/4 *iojap*. But in the reciprocal cross where *iojap* is used as the female parent, the F_1 plants showed all three phenotypes *viz*. green, white and striped.

P:	Iojap	×	Green	
	iijj	↓	IIjj	
F_1:		Iijj		
Phenotypes:	Green, White and Striped			
F_1	♀ Striped	Iijj	×	IIjj Green ♂
F_2	Genotypes:	Iijj		IIjj
	Phenotypes	Green, White, and Striped		Green, White, and Striped

It was found that the *iojap* gene in the homozygous recessive condition (*ii*) causes some of the plastids to mutate giving rise to colourless plastids. The mixture of green and colourless plastids accounts for the origin of striped plants. Once created, further transmission of the striped character takes place maternally through the egg cytoplasm as evident from the phenotypes of the F_2 progeny of the cross above.

MALE STERILITY IN PLANTS

The failure of pollen formation results in male sterile plants. In some crop plants male sterile mutants occur in which inheritance of male sterility follows one of the following two patterns: those in which the trait is inherited through a single recessive chromosomal gene segregating in Mendelian ratios; secondly, those that show maternal transmission. Rhoades in 1933 described maternal inheritance of male sterility in *Zea mays*.



In the cross shown above, the process of repeated backcrossing results in substitution of all the chromosomes of the fertile line for those of the male sterile line. The experiment therefore, not only shows cytoplasmic maternal transmission, but also confirms that chromosomal genes are not responsible for male sterility.

INHERITANCE THROUGH MITOCHONDRIA

Mitochondria are cytoplasmic organelles containing their own DNA and some enzymes that catalyse the process of oxidative phosphorylation. Mitochondria are not autonomous bodies because they require their own genes and nuclear genes to function normally. The role of mitochondria in cytoplasmic inheritance was first understood when Ephrussi and his colleagues (1955) studied petite mutants (petite meaning small) in yeast *(Saccharomyces cerevisiae).* When a cell divides, approximately equal numbers of mitochondria pass into daughter cells. They can originate from pre-existing mitochondria, and can probably divide transversely.

Several mutant strains of yeast are referred to as petite mutants because they form smaller colonies on agar as compared to normal yeast. The growth rate of petites is slow due to absence of respiratory enzymes, cytochromes *a*, *b* and *c*, and due to deficiency in some dehydrogenases present in normal mitochondria.

The petite strains are of three types. In the first type called *segregational petites* when an individual is crossed with a normal strain, a 1 : 1 ratio of normal: petite results after segregation. This suggests Mendelian inheritance and the petite strain has originated due to a mutation in nuclear genes.

The second and third types of *petite* strains arise when normal yeast cells are treated with the acridine dye euflavin or with ethidium bromide dyes which are known to intercalate in double stranded DNA. The petites so formed do not segregate regularly when crossed with normal yeast, and fall under two types—neutral and suppressive. When a neutral petite is crossed with a normal type, all the progeny in the next and successive generations are normal and the petite trait never reappears. If a haploid cell of the *neutral petite* strain is mated to a haploid vegetative cell of a normal strain, a diploid zygote is formed. The diploid cells produced by this zygote reproduce asexually, sometimes may even divide meiotically to form spores, but the petite trait is never visible in the progeny. The genetic explanation of the trait is that there is a cytoplasmic factor (p^+) which is present in the normal strains of yeast but absent in petites (p^-).

The *suppressive petites* when crossed to normal cells of yeast show the petite trait in the progeny but in non-Mendelian ratios. Suppressive petite mutants are found to have mutant DNA in their mitochondria. The mutant mitochondria replicate and transmit the mutant phenotype to the progeny cells.

Mitochondrial mutants are also known in *Neurospora, Paramecium* and *Trypanosoma.*

THE GENETICS OF MITOCHONDRIA AND CHLOROPLASTS

The presence of organelle-specific DNA in mitochondria and chloroplasts was identified with certainty over three decades ago. The occurrence of a specific chloroplast DNA was first demonstrated by Chun *et al*. in 1963, and of mitochondrial DNA a year later by Luck and Reich (1964). It is now fairly well established that organelle DNA is distinct from nuclear DNA in several respects. That it can carry out synthesis of RNA and protein on its own synthesising 'machinery' and replicate itself. There are mitochondrial and chloroplast ribosomes. The genomes of these organelles contain genes for the rRNA species present in both large and small subunits of their respective ribosomes.

Mitochondrial DNA (mt DNA): Mitochondria contain an open or closed circular DNA molecule. DNA from different species shows a wide range in average base composition ranging between less than 20% to more than 50% of $G + C$ content. It has been found that the amount of DNA present in mitochondria is sufficient to code for proteins. Grossman *et al*. (1971) have determined that a minimum base composition of 35% $G + C$ is needed to code for an average protein. Thus regions of mt DNA containing very small amounts of *G* and *C* nucleotides (in cases where the average base composition is less than 30%) may not be able to code for a protein.

Recent evidence indicates that there is one copy of the gene for each of the rRNA species associated with the large and small subunits of mitochondrial ribosomes. The positions of these two genes have been mapped in some lower eukaryotes. In the mold *Neurospora crassa* the two genes for rRNA are adjacent to each other, whereas in yeast *(Saccharomyces cerevisiae)* they are separated. With the use of radioactive isotopes it has been shown that in *N. crassa* and *S. cerevisiae* some of the mitochondrial inner membrane proteins are synthesised by the mitochondrial protein synthesising apparatus.

The existence of mutations have been recorded in the mitochondrial genome. A study of the oligomycin resistance mutations have led to the identification of 3 distinct linkage groups in mtDNA. Recently Tzagaloff *et al*. (1975) have been able to identify a number of mutants with defects in cytochrome oxidase and coenzyme *Q*-cytochrome *c* reductase.

mtDNA has also been studied in a limited number of higher plants. A closed circular DNA with molecular weight ranging between $60 - 80 \times 10^6$ daltons has been seen in spinach, lettuce and pea. However, higher plant mtDNA has not shown the presence of rRNA and tRNA genes. No protein synthesis is recorded for plant mtDNA.

Chloroplast DNA (ctDNA): Circular DNA molecules have been observed in ctDNA of both higher plants and algae. It contains genes for both large and small rRNA species present in chloroplast ribosomes. Hybridisation studies have indicated one gene per chromosome for each rRNA in tobacco, whereas in bean, lettuce, pea, spinach and maize there are two genes per chromosome. Similar studies have demonstrated the presence of genes for about 25 tRNA species each of size about 25,000 daltons.

Radioactive isotope studies *in vivo* have clearly shown that protein synthesis occurs in chloroplasts. One such protein is the large subunit of ribulose diphosphate carboxylase (termed

fraction I protein) which is a major constituent of chloroplast as well as other membrane proteins. The large subunit of fraction I protein and some other membrane proteins are also synthesised by isolated chloroplasts. The small subunit of fraction I protein is coded for by nuclear DNA.

Replication of Mitochondrial and Chloroplast DNA

Both mtDNA and ctDNA replicate within their respective organelles. *In vitro* synthesis of DNA also occurs in the isolated organelles if the four deoxyribonucleotide triphosphates are supplied and is blocked by inhibitors of DNA synthesis such as ethidium bromide, acriflavin and actinomycin-*D*. Whereas nuclear DNA synthesis is limited to the *S* phase of the cell cycle, there is evidence for continuous synthesis of organelle DNA throughout the cell cycle. *In vivo* density label studies have shown that mt DNA in Hela cells and ct DNA in *Chlamydomonas* and *Euglena* replicates semiconservatively. As in bacteria, replication of organelle DNA appears to involve the membrane (called mesosome in bacteria).

EXTRANUCLEAR GENES IN CHLAMYDOMONAS

The unicellular green alga *Chlamydomonas* is haploid with a single nucleus, a chloroplast and several mitochondria (Fig. 9.3). It can reproduce asexually as well as sexually by fusion of gametes of opposite mating types (mt^+ and mt^-). The mating type is controlled by a single nuclear gene. There is a fusion of the two chloroplasts in the diploid zygote. The zygote divides meiotically to produce mt^+ and mt^- cells in the ratio 2:2.

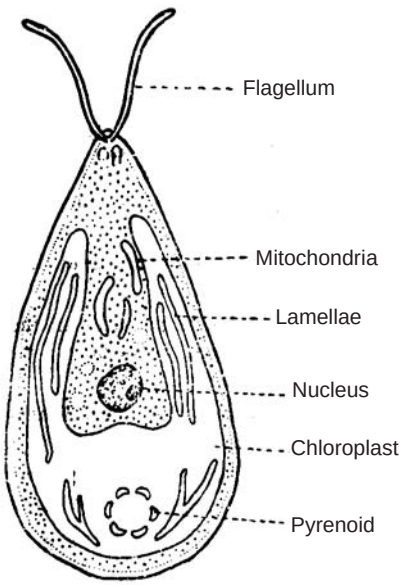


Fig. 9.3 The green alga *Chlamydomonas*.

In 1954 Ruth Sager isolated a mutant resistant to streptomycin (*sm*-2) which showed a different pattern of segregation after meiosis. When a streptomycin resistant strain is crossed with a wild type strain, the resulting progeny depends upon which mating type contributed the

gene for streptomycin-resistance. Thus when the mt^+ parent carries the *sm*-2 allele, the progeny is almost all streptomycin-resistant instead of segregating in a 2 : 2 ratio. In the reciprocal cross when mt^- parent has *sm*-2 allele, *none* of the progeny show streptomycin-resistance. Such uniparental transmission of a trait is typical of extrachromosomal inheritance.

A noteworthy feature of the above crosses is the production of a very small percentage of 'exceptional' zygotes which receive the *sm*-2 allele from the mt^- parent. These zygotes indicate biparental transmission. There is still another class of 'exceptional' zygotes, though rare, which transmit extranuclear genes only from the mt^+ parent. A class of *Chlamydomonas* extranuclear mutants known as 'minutes' also show biparental inheritance. The 'minutes' are induced by treatment with acriflavin and ethidium bromide. The photosynthesis deficient mutants which depend upon acetate for growing in light, and the temperature sensitive mutants in *Chlamydomonas* also exhibit uniparental inheritance. All the four haploid products of meiosis are inherited from the mt^+ parent. Exceptional zygotes showing biparental transmission occur also in these crosses at a frequency of less than 1 per cent.

There is evidence that the inheritance of the extranuclear traits in *Chlamydomonas* is due to chloroplast genes. According to Sager uniparental transmission of chloroplast genes could be due to degradation of chloroplast DNA originating from the mt^- parent in the zygote. When degradation does not occur exceptional zygotes showing biparental inheritance are formed. Such zygotes produce haploid progeny heterozygous for chloroplast genes. Subsequent mitotic divisions in the haploid progeny will show segregation of alleles, the members showing one or the other parental phenotype. When two or more chloroplast genes are considered in a cross, recombinants appear, their frequencies suggesting linkage between chloroplast genes. Sager and her colleagues have extended these studies and mapped positions of chloroplast genes.

QUESTIONS

1. What criteria can be used to distinguish between cytoplasmic inheritance and nuclear inheritance?
2. In maize some male sterile plants were artificially pollinated with pollen from healthy male fertile plants. The seeds were used for raising progeny plants which again turned out to be male sterile. Explain.
3. Distinguish between (*a*) maternal influence and cytoplasmic inheritance; (*b*) chloroplast DNA and mitochondrial DNA.
4. (*a*) The killer trait in *Paramecium* is due to kappa particles in the cytoplasm. What results would be obtained if prolonged matings are allowed between (*i*) two killer strains; (*ii*) two sensitive strains; (*iii*) one killer and one sensitive strain.

   (*b*) Sometimes *Paramecia* of killer strain undergo fast multiplication and produce sensitives in their progeny. Explain.
5. Describe one example of a mutation in mitochondria that produces a visible phenotype in a fungus.
6. (*a*) Extrachromosomal traits are mostly transmitted through the maternal parent. Give reasons.

   (*b*) Is there an extranuclear trait showing (*i*) uniparental transmission through the male gamete? (*ii*) biparental transmission?

SELECTED READINGS

Battey, J. and Clayton, D.A. 1978. The Transcription Map of Mouse Mitochondrial DNA. *Cell* **14 : 143.**

Beale, G. and Knowles, J. 1978. Extranuclear Genetics. Edward Arnold, London.

Cavallier-Smith, T. 1970. Electron Microscopic Evidence for Chloroplast Fusion in Zygotes of *Chlamydomonas reinhardii. Nature* **228 :** 333.

Chiang, K.S. and Sueoka, N. 1967. Replication of Chloroplast DNA in *C. reinhardii* during Vegetative Cell Cycle: Its Mode and Regulation. *Proc. Natl. Acad. Sci. U.S.* **57:** 1506.

Dujon, B. *et al.,* 1974. Mitochondrial Genetics IX. A Model for Recombination and Segregation of Mitochondrial Genomes in *S. cerevisiae. Genetics* **78 :** 415.

Ephrussi, B. *et al.,* 1966. Et udes sur la suppressivite des mutants a deficience respiratoire de la levure. II Etapes de la mutation grande on petite provoquee par le facteur suppressif. *Genetics* **54 :** 1.

Kasamatsu, H. and Vinograd, J. 1971. A Novel Closed Circular Mitochondrial DNA with Properties of a Replicating Intermediate. *Proc.Natl. Acad. Sci. U.S.* **68 :** 2262.

Kirk, J.T.O. and Tilney-Bassett. R.A.E. 1967. The Plastids. W.H. Freeman and Co., San Francisco.

Perlman, P.S. *et al*., 1977. Localisation of Genes for Variant Forms of Mitochondrial Proteins on Mitochondrial DNA of *S. cerevisiae. J. Mol. Biol.* **115 :** 675.

Rhoades, M.M. 1946. Plastid Mutations. *Cold Spring Harb. Symp. Quant. Biol.* **11 :** 202.

Sager, Ruth 1972. Cytoplasmic Genes and Organelles. Academic Press, New York.

Sonneborn, T.M. 1970. Gene Action in Development. *Proc. Roy. Soc. London (B)* **176 :** 347.

Surzycki, S.J. and Gillham. N.W. 1971. Organelle Mutations and their Expression in *Chlamydomonas reinhardii. Proc. Natl. Acad. Sci. U.S.* **68 :** 1301.

Wilkie, D. 1964. The Cytoplasm in Heredity. Methuen, London.

10

The Genetic Control of Sex

Some organisms such as earthworms, flukes, snails and numerous species of higher plants are hermaphrodites or monoecious as they produce both male and female functional gametes in the single individual. Such organisms do not possess special genetic sex determining mechanisms because all the cells have developed from mitotic divisions of the original fertilised egg. But in unisexual or dioecious organisms where the two sexes occur in separate individuals, there are different kinds of genetic sex determining mechanisms present.

Historically speaking, the earlier scientists knew nothing about the origin or inheritance of sex. Until the last quarter of the nineteenth century it was still debatable as to whether or not sex arose as a response of the organism to external stimuli such as temperature and nutrition. Some thought that sex was controlled by internal factors whose nature was obscure at that time. With the advent of the microscope the cytological basis of sex began to be understood. Oscar Hertwig in 1876 observed fusion of two nuclei during fertilisation, one derived from the egg the other from the spermatozoon in the sea urchin *Toxopneustes lividus*. These observations were confirmed by van Beneden (1883) in *Parascaris equorum*, a parasite on the horse, and by Boveri (1890) in the sea urchin *Echinus microtuberculatus*. The existence of definite sex chromosomes was realised in 1891 when Henking observed a chromatin body in the dividing spermatocytes of the plant bug *Pyrrhocorus apterus* which stained more intensely, and lagged behind other chromosomes during anaphasic movement. At its function was unknown Henking called it the '*X*' chromosome. MuClung (1901) was first to suggest that this chromosome was concerned with determination of sex.

After Sutton's proposal in 1903 that Mendelian factors are carried on chromosomes, a number of scientists got engaged in the study of chromosome cytology in various plants and animals. It was established that chromosomes occurred in the paired condition, a pair consisting of two homologues. One pair of homologues did not have identical chromosomes and was referred to as a heteromorphic pair. It was easy to observe segregation of this one heteromorphic pair which later turned out to be the sex-chromosome pair. With further studies on the heteromorphic pair, Wilson (1905) found some variation in insects of the order Hemiptera. In some insects the males had the heteromorphic pair, while in other insects the females had this pair. It was confirmed that there was one pair of chromosomes which was different in the two sexes and

was given the name heterochromosomes by Montgomery in 1904; the rest of the chromosomes which were same in both sexes were called *autosomes*. Wilson (1909, 1911) used the terms sex-chromosomes as well as *X* and *Y* chromosomes for the heterochromosomes. Stevens (1905) called the male dipteran flies which had an unequal pair of chromosomes as *heterogametic* and females with two *X* chromosomes as *homogametic*.

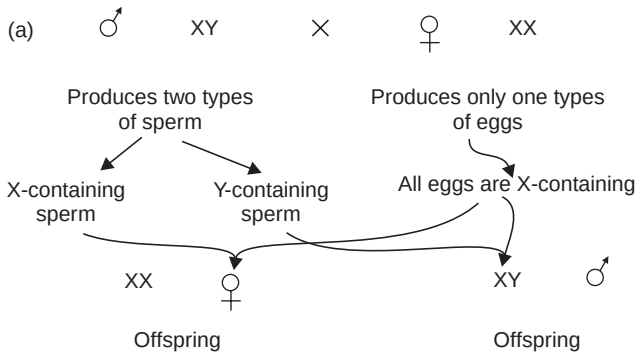
In 1910, T.H. Morgan described the first instance of sex linkage in *Drosophila* where he showed that the gene for white eye colour was carried on the *X* chromosome. He could explain the correlation between segregation of sex chromosomes and the inheritance of sex chromosomes and the inheritance of a sex-linked trait. In the next five years more than 30 genes carried on the *X* chromosome had been described.

The establishment of the chromosomal basis of sex determination brought to light some other types of chromosomal arrangements underlying sex determination which were different from the *X* and *Y* mechanism of dipteran flies. These will be discussed below. Later on it was also found that sex determining mechanisms are not always chromosomal, they may be sometimes genotypic, cytoplasmic or environmental (hormonal).

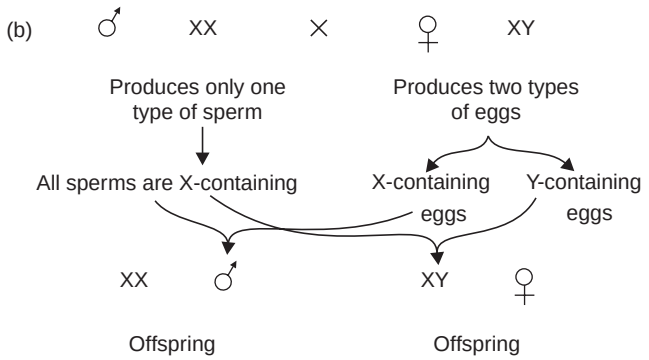
CHROMOSOMAL BASIS OF SEX DETERMINATION

The chromosomal mechanisms underlying sex determination are of the following types:

1. *The XY mechanism.* This is the most widespread type of sex determining mechanism. Here one sex is heterogametic and has *XY* sex chromosomes; the other sex is homogametic and has *XX* chromosomes. Depending upon which sex is heterogametic, the *XY* mechanism is of two types: (*a*) *XY* male type where the male is heterogametic and has *XY* chromosomes, the female is homogametic with *XX*; and (*b*) *XY* female type in which the female is heterogametic with *XY* chromosomes, the male has *XX*. The types of gametes formed and the possible fertilisations are shown diagrammatically below.



Since *X* and *Y* sperms are produced in roughly equal proportions, there is 50% chance of a fertilisation to result in a male offspring, and 50% for female offspring. The *XY* male type is found in *Drosophila*, human being, many animals, and some plants such as *Rumex*, *Coccinea*, *Melandrium*, the moss *Sphaerocarpos* and others.



The *XY* female type is found in some birds, fishes, some insects and only one plant namely *Fragaria elatior*. Here the females are heterogametic and males homogametic.

Some authors like to restrict the *XY* terminology for organisms having male as the heterogametic sex. But when the female is heterogametic they prefer to call it the *ZW* : *ZZ* type of sex determining mechanism. This terminology is useful in some organisms like fishes where both male and female heterogamety can occur in closely related species. It is otherwise not universally applied.

In some Chironomidae, amphibia and fishes, the *X* and *Y* chromosomes are not very different from each other cytologically. Consequently they are homologous and paired for the most part except a short region where the sex determining loci are present (Fig. 10.1*a*). Where the *X* and *Y* chromosomes are cytologically different, they have one or more short homologous segments where pairing takes place (Fig. 10.1*b–d*).

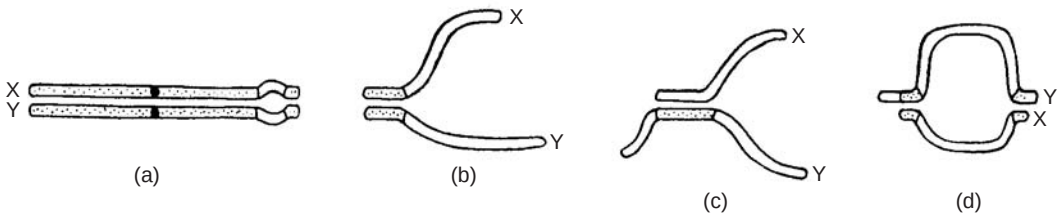


Fig. 10.1(*a–d*) Diagrams showing pairing relationship between *X* and *Y* chromosomes. Blank regions represent differential segments, dotted pairing segments.

For orderly segregation of chromosomal bivalents during meiosis, pairing and chiasma formation are essential. Darlington proposed that in the case of unlike *X* and *Y* chromosomes of the heterogametic sex, the *X* and *Y* chromosomes each consist of two parts: (*a*) a pairing segment in which chiasmata are formed: (*b*) a differential segment where no pairing and no chiasma formation takes place (Fig. 10.1*b–d*). In many bugs and the order Neuroptera possibly there are no definite pairing segments common to both *X* and *Y*.

2. *The XO mechanism.* In some nematodes, spiders and insects the *Y* chromosome is absent altogether so that the sex determining mechanism consists of *XO* : *XX* chromosomes in the two sexes. The letter *O* means zero and denotes absence of a sex chromosome in one sex. The *XO* mechanism is also of two types: (*a*) *XO* male type where the male lacks one sex chromosome resulting in an odd number of chromosomes in the somatic tissues. The females have two *X* chromosomes, therefore an even number of somatic chromosomes. The mechanism was first discovered in the *bug Protenor* where the diploid males have 13 chromosomes and females have 14. Males produce two types of sperms, one with 7 chromosomes, others with 6.

The eggs always have 7 chromosomes. If a sperm with 7 chromosomes fertilises an egg, it results in a female offspring; if a sperm with 6 chromosomes fuses with a egg, the result is a male. (*b*) *XO* female type. In the insect *Fumea* the females have *XO* sex chromosome constitution and the males *XX*. The females produce two types of eggs, some containing an *X* chromosome and some without *X*. All sperms contain an *X,* and on fertilisation can theoretically produce 50% females (fusion with egg without *X*), and 50% males (union of egg with *X*). In human beings some abnormal females born with a clinical disorder known as Turner's syndrome have *XO* sex chromosome constitution. This is an anomalous condition and is discussed in the chapter on human genetics.

Special Characteristics of the X and Y Chromosomes

In most organisms the *Y* chromosome is smaller than the *X*. *Drosophila* is an exception in having a very large *Y* chromosome which at least in 2 species *D. hydei* and *D. neohydei* throws off five pairs of giant loops similar to those in oocyte lampbrush chromosomes. At least five fertility genes have been identified through deletion studies on these loops (Fig. 10.2). In some beetles and a few mammals such as the vole *Microtus agrestis* there are giant *X* and *Y* chromosomes.

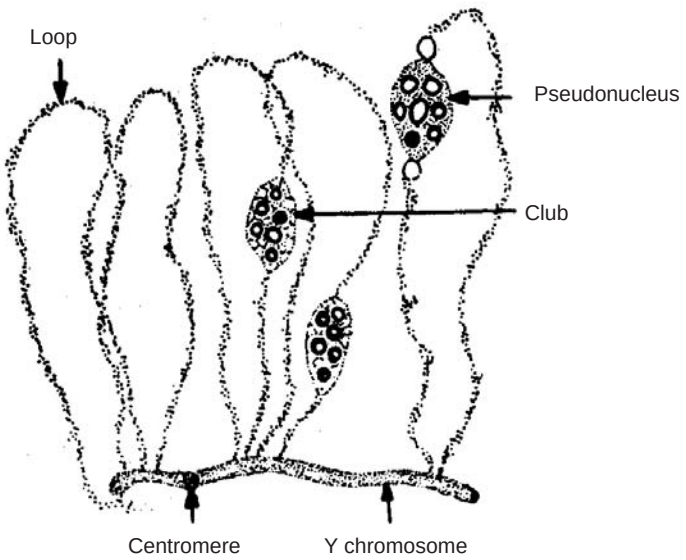


Fig. 10.2 The *Y* chromosome loops in *Drosophila*. Adapted from Hess, 1969.

SEX CHROMOSOMES IN RESTING NUCLEI: BARR BODIES AND LYON'S HYPOTHESIS

In 1949 Murray L. Barr and E.G. Bertram found that the nerve cells of the female cat had a small, dark staining body lying against the nuclear membrane. This body was absent in the neurons of male cats. It was subsequently found to be present in many tissues and organs not only of female cats but of a large number of mammalian females including human beings. It was always absent in males. The name *Barr body* (after the discoverer) or sex chromatin (due to its association with sex) was given to it. In human beings it is easily observed in scrapings from the mouth cavity (Fig. 10.3), and the method is of great value in determining sex of

newborn babies. Sex chromatin is also present in the polymorphonuclear leukocytes of circulating blood in females in the form of a small body called "*drumstick*" attached to one of the lobes of the nucleus. Barr bodies are not present in *Drosophila*.

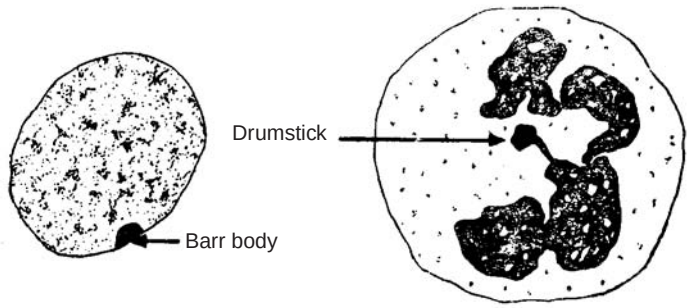


Fig. 10.3 Sex chromatin in a cell of buccal mucosa (Barr body), and in a polymorphonuclear leukocyte (drumstick).

After the association between Barr body and sex was established, scientists turned their attention to individuals with an abnormal number of sex chromosomes such as *XO* (Turner's syndrome), *XXY*, *XXXY* and *XXXXY* and others. It was found that the number of sex chromatin bodies is always *one less* than the number of *X* chromosomes in the cell. A normal female with *XX* chromosomes has one Barr body; normal male (*XY*) has none. Due to this females are said to be *chromatin positive* and males *chromatin negative*. An abnormal female with *XO* chromosomes has no Barr body. Abnormal females with three *X* chromosomes have two Barr bodies; those with four *X* chromosomes have three Barr bodies, and so on. In 1969 Mary F. Lyon proposed that in normal females one of the two *X* chromosomes was inactive and condensed and as such it became visible in interphase nucleus as a deeply staining body. It appears dark due to its ability to take up DNA stain. The second *X* chromosome was active and extended and therefore could not be stained interphase nuclei. Males have a single *X* chromosome which is active and uncondensed. In abnormal individuals with multiple *X* chromosomes, only one *X* is active, all the remaining *X* chromosomes are condensed into a corresponding number of sex chromatin bodies. This is known as *Lyon's hypothesis* and has since been verified in a number of mammalian species.

Some interesting characteristics of the inactive *X* have been revealed by the autoradiographic technique. It has been found that the condensed *X* of mammalian females is late replicating so that during the *S* phase of DNA synthesis it lags behind the active *X*-chromosome. The late replication is associated with the condensed state and inactivation or switching off of genes. The question arises as to which of the two *X*-chromosomes is late replicating and forms the Barr body? Out of the two *X*-chromosomes one is derived from the paternal parent, the other from the maternal parent. It has been found that in some tissues the late replicating *X* may be the one derived paternally, while in other tissues of the same individual the maternal *X* would be late replicating.

The phenomenon of inactivation affects the functioning of the sex-linked genes borne on the two chromosomes. Mammalian females are therefore somatic mosaics for some of their sex-linked genes. The tortoise-shell cat with its mixture of black and white patches is a clear cut example. Such a cat arises from a cross between a black and a white skinned cat. The

heterozygous offspring shows black patches where the *X*-chromosome from the black parent is active; and white patches where the second *X* is active. Hence the tortoise-shell appearance. A similar effect is noticed in the variegated coat colour in mice. Lyon's hypothesis can also be demonstrated by autoradiography of cells of the female mule, where the *X*-chromosome from the donkey parent differs morphologically and is easily identified from the horse *X*. It can be observed that in some cells the donkey *X* is late replicating, in other cells the horse *X* is late replicating.

Another proof that the condensed *X*-chromosome has mostly inactive nonfunctional genes came from studies of the enzyme G-6-PD by Davidson and his colleagues. The enzyme G-6-PD is controlled by genes on the *X* chromosome. The enzyme has two variants, normal and deficient. The enzymatic activity in peripheral blood of normal males and females is the same. A woman heterozygous for the two variants shows the presence of both variants of the enzyme. But if single somatic cells are cultured to form identical daughter cells of a *clone*, it is found that each clone contains either one enzyme variant or the other, but not both. This indicates inactivation of one *X*-chromosome in a clone.

A combined cytological and genetic study of coat colour in mice has confirmed that genes present on the condensed *X*-chromosome are inactivated, whereas genes on the euchromatic *X* are active. In mice a segment of an autosomal chromosome may become translocated to an *X*-chromosome (X^t, Cattanach's translocation); such an *X* can be distinguished cytologically from the normal chromosome (X^n). The translocated autosomal segment carries dominant genes for coat colour. In heterozygous females (X^t/X^n) with variegated coat colour, cells in which the translocation carrying *X* (X^t) was condensed showed expression of the recessive coat colour genes (due to X^n). In contrast regions where the dominant coat colour was visible contained cells having a condensed X^n.

The results mentioned above also imply that the translocated autosomal segment is also inactivated along with *X*-chromosome (*position effect* variegation). The studies of Russel and Montgomery (1970) have confirmed that position effect variegation occurs in mice carrying Cattanach's translocation. The autosomal segment attached to *X*-chromosome (X^t) shows maximum inactivation in the region adjacent to *X*; the genes located far from *X* may remain active.

Inactivation of one *X* chromosome is said to occur in the early embryonic stages. In any cell, the choice as to which *X*-chromosome will become condensed is random. Inactivation is permanent, and all daughter cells of a given cell will show condensation of the same *X*-chromosome as the one in the parent cell. There is no condensation of the *X*-chromosome in the germ line cells of females.

X-chromosome inactivation is often interpreted in terms of a phenomenon called *dosage compensation*. With two *X*-chromosomes the female has a double dose of sex-linked genes as compared to the male with the single *X*. The inactivation of one *X*-chromosome in the homogametic sex appears to compensate for the double dose of sex-linked genes in females. Abnormal individuals with multiple *X*-chromosomes have only one active *X*, all the remaining *X*-chromosomes being condensed, giving further support to dosage compensation.

BALANCE THEORY OF SEX DETERMINATION

Drosophila melanogaster has four pairs of chromosomes in somatic tissues ($2n = 8$) of which the fourth pair represents sex chromosomes. The females have two *X* chromosomes and males

have *XY*. However, the genetic basis of sex determination of *Drosophila* is not directly related with the presence of *XX* or *XY*, but on the ratio of *X* chromosomes to autosomes as described below.

In 1910 T.H. Morgan discovered a white-eyed mutant in *Drosophila* which was recessive to the normal red-eyed condition of the wild type. After making several crosses between the red and white-eyed flies, Morgan concluded that the gene for white eyes was located on the *X* chromosome, and there was no allele for this gene on the *Y* chromosome.

Further evidence for the *Y* playing no role in sex determination came in 1916 from the work of Calvin B. Bridges. He found some exceptional females which appeared to show that the inheritance of the *X*-linked white-eyed trait was exclusively maternal. Bridges explained that this was due to the two *X* chromosomes of the mother remaining together during meiosis instead of segregating. This phenomenon is known as nondisjunction and results in an egg with two *X* chromosomes or an egg with no *X*; both types of eggs have the normal 3 autosomes. When an *XX* egg is fertilised by a *Y* bearing sperm, the result is an *XXY* fly which happens to be a normal fertile female. Evidently, the *Y* chromosome neither interferes with development of the female phenotype, nor contributes to expression of the male sex. If the second type of egg which is without an *X* is fertilised by a *Y* bearing sperm, it results in a *YO* embryo which does not develop. If an *X* bearing sperm fertilises on *XX* egg, the resulting *XXX* embryo is viable and grows into a sterile fly designated as *superfemale* or more appropriately *metafemale*. If an *X* sperm unites with an egg without an *X*, the zygote has *XO* sex chromosome constitution and grows into a phenotypically normal but sterile male.

The studies detailed above led Bridges to the following conclusions on sex determination in *Drosophila*: that the *Y* chromosome does not determine maleness; at least one *X* must be present to produce a viable embryo; the *Y* chromosome is necessary for male fertility; the *XXX* constitution is an abnormal female phenotype not always viable, and termed *superfemale* or *metafemale*; the *XXY* individuals were normal fertile females, and *XYY* were fertile males.

In 1922 Bridges found a triploid fly having the entire set of four chromosomes represented 3 times. Sometimes during oogenesis there can be nondisjunction in all the chromosomes to produce a diploid unreduced egg. If such an egg is fertilised by a haploid *X* sperm, it results in a triploid fly having 3*A*3*X* chromosomes where *A* denotes one set of autosomes. When a triploid fly is crossed to a diploid male 2*AXY,* it produces an interesting range of phenotypes and chromosomal constitutions with different properties of autosomes and sex chromosomes as described below.

Female parent 3*AXXX* produces 4 types of gametes: 2*AXX*, 2*AX*, *AXX*, *AX*. Male parent 2*AXY* produces 2 types of gametes: *AX*, *AY*.

These gametes are expected to produce the following *eight* types of fertilisations:

Chromosomal constitution	*Phenotype*	*Ratio of sex chromosomes to autosomes X/A*
3*AXXX*	Triploid female	1/1 = 1.00
2*AXY*	Diploid male	1/2 = 0.5
3*AXX*	Triploid intersex	2/3 = 0.67
3*AXY*	Metamale	1/3 = 0.33
2*AXY*	Diploid female	2/2 = 1.00
3*AXXY*	Triploid intersex	2/3 = 0.67
2*AXXX*	Metafemale	3/2 = 1.5
2*AXXY*	Diploid female	2/2 = 1.0

Bridges compared the phenotypes of the embryos with the corresponding chromosomal constitutions. He found that a correlation exists between the ratio of *X* chromosomes to complete sets of autosomes which determines the phenotypic sex. When this ratio is 1:1 (1.0) for instance, the phenotype of the fly is normal female; when the ratio is 1:2 (0.5) it is a normal male. But when the ratio is in between *i.e.* 2:3 (0.67), the phenotype is also a mixture of male and female characters designated as intersex. When the number of *X* chromosomes increases in proportion to the autosomes to give a ratio of 3:2 (1.5) the phenotype shows exaggeration of female characteristics. Such flies referred to as superfemales or metafemales are weak and sterile. Lastly, an increase in the number of complete autosomal sets in proportion to *Xs* in a ratio 1:3 (0.33) results in supermales or metamales which are also weak and sterile.

On the basis of above mentioned results Bridges proposed his *Balance Theory of Sex Determination*. He suggested that the *X* chromosome contains genes that appear to trigger the expression of all those genes in the cell which control femaleness; the autosomes trigger the expression of male determining genes. The ratio between the number of *X* chromosomes and the number of complete sets of autosomes determines the sexual phenotype. However, this quantitative relationship is found to apply only to *Drosophila* and not to any other animal species. In plants of *Melandrium*, the male plants have *XY* and females *XX* chromosomes. Studies have been done on triploid and tetraploid plants. Correlations have been found between the number of sex chromosomes against the complete autosome sets in determining phenotypic sex. Plants with following chromosomal constitutions were females: 2*X*2*A*; 2*X*3*A*; 3*X*3*A*; 4*X*4*A*. The male plants had following constitutions: *XY* 2*A*; *XY*3*A*; 2*XY*3*A;* 3*XY*3*A*; and 3*XY*4*A*. It is evident that in contrast to *Drosophila*, the *Y* chromosome in *Melandrium* determines male sex; the *X* chromosome determines femaleness. The role of autosomes is not clear; perhaps they determine femaleness.

L.V. Morgan (wife of T.H. Morgan) in 1921 discovered a strain of *Drosophila* flies in which 100 per cent nondisjunction occurred. The females of this strain always produced female offspring identical to the maternal parent and male offspring identical to the paternal parent. It appeared as if the daughters were receiving both *X* chromosomes only from the mother, and the sons were getting the single *X* from their father. Out of this strain stocks were developed in which all the females were white-eyed and all the males were red-eyed. The idea was to see whether each female would transmit its two white-bearing *X* chromosomes only to the daughters, and the male transmitted its normal red-bearing *X* chromosome only to the sons or not.

When L.V. Morgan mated the white-eyed females to the red-eyed males, indeed, the progeny consisted exclusively of white-eyed females and red-eyed males. To explain this Morgan suggested that the females produced two types of eggs—one type with nondisjunctional *XX* chromosomes the other with only one *X* chromosome. They explained the production *XX* eggs as resulting from a physical attachment between the maternal *X* chromosomes and called this strain *attached-X*. This was confirmed later when L.V. Morgan cytologically observed that the two *X* chromosomes were physically joined at the centromere.

SEX MOSAICS

Among insects sometimes one part of the body is male, the other part female producing sex mosaics known as *gynandromorphs* or *gynanders*. These have been studied most thoroughly in *Drosophila* which has no sex hormones so that tissues develop autonomously. The sexual phenotype is determined by the number of *X* chromosomes against sets of autosomes. If during

mitosis in early embryogenesis there is nondisjunction between the two *X* chromosomes, some cells will have *XX*, others *XO* constitution. The descendents of *XX* containing cells will result in the development of female tissues while descendents of *XO* cells will produce male tissues. The resulting individual will be a spectacular mosaic. Gynandromorphs are usually bilateral with one side of the body male, the other side female. There are irregular gynandromorphs also in which the proportions of male to female tissues are variable depending upon the time and stage of embryo development when *XX* nondisjunction took place.

Sex mosaics are sometimes mistaken for intersexes. In a gynandromorph the boundary between male and female regions is always sharp and distinct, whereas in intersexes, all parts of the body may have a mixture of male and female characteristics showing a condition intermediate between maleness and femaleness.

Sex mosaicism occurs in humans also. Chromosome preparations from peripheral blood may show *XO/XX* or *XO/XXY* cells frequently observed in Turner's and Klinefelter's syndromes (discussed in a later chapter). As in *Drosophila,* if there is nondisjunction of two *X* chromosomes during early stages of embryo development, it results in mosaicism.

GENOTYPIC MECHANISMS FOR SEX DETERMINATION

Sex is not always determined on the basis of chromosomes. Sometimes genes present on autosomes or sex chromosomes influence phenotypic expression of sex. Sturtevant discovered the influence on sex of a *transformer* gene present on an autosomal chromosome in *Drosophila melanogaster. Transformer* or *tra* is a recessive gene with no effect on sexual development of heterozygous individuals. But when present in the homozygous condition *(tra/tra),* it can transform an individual with two *X* chromosomes, that would otherwise be a female, into a phenotypic sterile male.

A gene which induces reversal of sex is also known in human beings. When it acts, the genetic sex of an individual is different from the phenotypic sex. In the condition known as *testicular feminisation* the individual has *XY* chromosomes but is phenotypically female. Internally they do not have a uterus, fallopian tubes or ovaries, instead they have undescended, degenerate testes and vas deferens. They have normal levels of the male hormone, androgen. This is due to a dominant *X*-linked gene which has no effect on females but in males it makes the target tissues insensitive to the androgenic hormone. It is said that cyclic *AMP* (adenosine monophosphate) which binds the hormone at the outer membrane of cells and transports it inside, is lacking in the cells. Since the target tissues are not able to respond to androgen, the phenotype is female. The gene responsible for androgen insensitivity can be traced by pedigree analysis of an affected family which has been having an excess of girls for many generations. A woman who is carrier of the gene would be normal, but one-half of her children having *XY* chromosomes will become androgen insensitive females.

In rats a dominant gene is known to induce male sterility. The gene has no effect in females, but in males it causes inhibition of mitosis in spermatogonial cells which fail to produce spermatocytes and sperms, and sterility results. An excess of cyclic *AMP* is produced which interferes with the normal stimulating function of gonadotropic hormones from the pituitary glands in inducing cell divisions in spermatogonial cells.

In bees and wasps belonging to the Hymenoptera, fertilised eggs produce diploid males, while unfertilised eggs divide parthenogenetically to produce haploid males. Meiosis takes

place during egg formation, but the male bees produce sperm by mitosis. The queen bee uses this sperm to produce females whereas her unfertilised eggs develop into haploid males. It first appeared therefore, that sex determination in Hymenoptera is based on haploidy and diploidy. Later on it was found out from studies on the wasp *Bracon hebetor* by Whiting that a gene called *X* determined sex. There are nine alleles of this gene designated as *Xa*, *Xb*, *Xc* and so on. A wasp that gets a combination of any two different alleles such as *Xa/Xb*, *Xa/Xc*, etc., becomes female. But when a single allele is present in the haploid condition such as *Xa*, *Xb*, *Xc*, etc., or in a homozygous diploid condition such as *Xa/Xa*, *Xb/Xb* etc., the wasp becomes a male. Because the number of different alleles is large, most matings take place between unrelated individuals and the homozygous condition is rarely obtained. The diploids therefore become females. In the wood lemming *Myopus schisticolor,* females could be either *XX* or *XY*. Both types of females are phenotypically identical, fertile and normal!

ENVIRONMENTAL CONTROL OF SEX

The sea worm *Bonellia* is a fantastic example. It is a sexually dimorphic organism in which the female is large, about 10 cm long, whereas the male is only 3 mm long and lives parasitically inside the cloaca of the female. The egg on fertilisation produces a free swimming larva. If there are no females around, the larvae settle down on the bottom of the sea and develop into females. But if mature females are available, the larva attaches itself to its proboscis and grows by drawing nourishment from the female, into a male! It then migrates down the proboscis to reside in the cloaca of the female.

Experiments have been done where the growing larvae were removed from the proboscis and made to live away from the females. Such larvae developed into *intersexes* which were neither male nor female. Even extracts of the proboscis induced the development of maleness. Although the true mechanism of sex determination in *Bonellia* is not known, it is tentatively explained that a sex hormone produced by the female might be stimulating the expression of some larval genes for maleness.

In the marine annelid *Dinophilus* also the female worm is much larger than the male. The eggs are of two different sizes. The large eggs on fertilisation develop into females, the small ones into males. As there are no sex chromosomes in *Dinophilus,* it appears that the quantity of egg cytoplasm determines sex. *Ophryotrocha* is also a marine annelid which is male if there are 15–20 segments in the body; it is female if the number of segments is more. If some segments which contain the genital regions are cut off from a female, it develops a testis and becomes a male.

In one species of frog sexual development is influenced by environment. There is *XY* chromosomal mechanism, and in waters at normal temperatures, male and female offspring are produced in the ratio 1:1. However, if the temperature of water is maintained at 90°F, only males are produced.

Plants offer many more examples for environmental influence on sex. If soil conditions favour good growth *Equisetum* (horsetail) will show female characteristics; on poor soils the plants show male characteristics. Plants of *Canabis sativa* (hemp) with *XY* mechanism normally

produce only male flowers. If seeds are sown in summer months of May to July the resulting plants show male and female phenotypes in the ratio 1:1. If seed sowing is delayed beyond July, the proportion of female plants increases. If seeds are sown in winter when days are short and nights long, only female plants are produced. In muskmelon *(Cucumis melo)* and cucumber *(C. sativus)* Buyers *et al.,* have found that ethylene regulates sex expression.

The gypsy moth *Lymantria dispar* is distributed throughout countries of Europe, North Africa and Asia. In 1911 Richard Goldschmidt made crosses between moths from different geographical locations and found the phenotypic sex of the progeny different from expectations.

When *XX* male was crossed to an *XY* female from the *same* locality the progeny consisted of males and females in the ratio 1:1. But when he crossed moths from geographically different places such as females from Korea and males from Tokyo, some intersexes were produced in the first or second generation. In the reciprocal cross only normal males and females were produced. According to Goldschmidt sex in the gypsy moth was controlled by male factors (*M*) carried on *X* chromosome, and female factors (*F*) on the *Y* chromosome. The "strength" of these factors were different in the geographical strains crossed, resulting in imbalance which produced intersexes. The *F* and *M* factors are of equal "strength" in moths of the same locality. Thus if a strong male (*MsMs*) from one locality is crossed to a weak female (*FwFw*), the progeny males (*MsMw*) are normal, but females (*FwMs*) are intersexes instead of normal.

HORMONES AND SEX DETERMINATION

Hormones are known to play a role in sex determination and are sometimes considered as environmental factors.

Poultry breeders have often noticed that a good egg-laying hen starts showing male characteristics in old age. Not only does the comb develop on the head, but it crows like a rooster and even starts mating with hens. The explanation came from post mortem anatomical investigations. In poultry females have a single functional ovary. In the centre of this ovary is a small mass of tissue similar to the testis. The female hormones produced by the ovary ensure the normal development and functioning of female characters, at the same time suppressing growth of the small testis-like tissue. If however, due to infection the ovarian tissues are destroyed, the small testicular tissue gets released from inhibition. It grows and produces male hormones due to which the hen starts showing a shift towards maleness.

It is possible to change sex in poultry under experimental conditions by administering the male hormone testosterone to new born chicks. The normal functioning of ovaries is suppressed and the bird becomes a male. The technique has potential for inducing sex reversal in human beings as well, but so far there is no record of its application. Some cases of natural sex reversal are known in human beings. Since they are due to abnormalities they will be dealt with under Human Genetics.

The freemartin cattle is one of the best examples for illustrating the role of hormones in sex determination. When dizygotic twins, consisting of one male and one female are born, it is found that in most cases the female twin shows male characteristics and is sterile. Such a phenotypic male which is genetically female is called a freemartin. The condition originates when there is fusion of the two foetal membranes resulting in joining of the blood vessels of the male and female twins. The male calf starts producing sex hormones earlier than the female calf. The male sex hormones cross the placenta and reach the female and induce the development of male characteristics. The same thing is known to happen in pigs, goat and sheep.

QUESTIONS

1. What type of sex determining mechanism exists in (*a*) spiders, (*b*) *Sphaerocarpos,* (*c*) fishes (*d*) man?
2. The *XO* sex chromosome constitution is normal when present in the insect *Fumea,* is normal but sterile in *Drosophila* and abnormal in the human. Explain.
3. The gene for barred feathers (*B*) in poultry is sex-linked and dominant, its recessive allele (*b*) producing solid black colour in homozygous fowls. Silkiness in feathers is due to an autosomal recessive gene (*s*), its dominant allele (*S*) producing nonsilkiness. What would be the result of mating a solid black cock heterozygous for silky feathers with a barred, silky hen?
4. How many Barr bodies would be expected in buccal smears of following individuals:
   (*a*) *XYY*, (*b*) *XXXX*, (*c*) *XXYY*, (*d*) *XO*, (*e*) *XX/XXX* mosaic?
   What are the characteristic features of *X* chromosome inactivation?
5. (*a*) What would be the phenotypic sex of fruit flies with following chromosomal constitutions:
   (*i*) 3*AXXY*, (*ii*) 2*AXY*, (*iii*) 2*AXXY*, (*iv*) 3*AXX*
   (*b*) How would you distinguish between a *Drosophila* intersex and superfemale?
6. Carnation eyes and short wings in *Drosophila* are two recessive mutants due to genes *c* and *s* respectively.
   A female with carnation eyes and long wings was crossed to a normal eyed, short winged male. All the males in the progeny had carnation eyes, but half had long wings, and half short wings. All the female progeny had normal red eyes, but half had normal wings and half had short wings. What were the genotypes of the parents and of the progeny males and females?
7. A female *Drosophila* homozygous for red eyes was crossed to a hemizygous white eyed male. The result was a bilateral 'gynandromorph' with a red eye in the female half and a white eye in the male half. Explain.
8. The plants of *Melandrium* are dioecious. What would be the phenotypic sex of plants with following chromosome constitution: (*i*) 3*AXXX*, (*ii*) 3*AXY*, (*iii*) 4*A*4*XY*.
9. What explanation can you give for a *Drosophila* gynandromorph which is (*a*) 75% female and 25% male (*b*) 50% male and 50% female.
10. White eyes in *Drosophila* is a sex-linked recessive trait (*w*) while bright scarlet eyes are due to an autosomal recessive gene(s). Normal eye colour in wild type flies is dull red. A white eyed male heterozygous for scarlet eyes was crossed to a scarlet eyed female heterozygous for white eyes. What phenotypes are expected in their progeny?
11. Name an organism in which phenotypic sex is controlled by following factors: (*a*) size of eggs, (*b*) water temperature, (*c*) ethylene, (*d*) nutritional status of the soil, (*e*) female hormones, (*f*) geographical location, (*g*) onset of old age.

SELECTED READINGS

Arcos-Teran, L. 1972. DNA Replikation und die natur der spät replizierenden orte im *X* chromosom von *D. melanogaster. Chromosoma* **37 :** 233.

Barr, M.L. and Bertram, E. G. 1949. A Morphological Distinction between Neurones of the Male and Female and the Behaviour of the Nucleolar Satellite during Accelerated Nucleoprotein Synthesis, *Nature* **163 :** 676.

Bridges, C.D. 1925. Sex in Relation to Chromosomes and Genes. Amer. *Nat.* 59: 127.

Brown, S. W. 1966. Heterochromatin. *Science* **151 :** 417.

Cattanach, B.M. 1975. Control of Chromosome Inactivation, *Annu. Rev. Genetics* **2 :** 1.

Chandra, H.S. and Brown, S.W. 1975. Chromosome Imprinting and the Mammalian *X* Chromosome. *Nature* **253 :** 165.

Doncaster, L. and Raynor, G.H. 1906. On Breeding Experiments with Lepidoptera. *Proc. Zool. Soc. London.* **1** : 125.

Goldschmidt, R.D. 1955. Theoretical Genetics. Univ. California Press. Berkeley.

Goodale, H.D. 1909. Sex and its Relation to the Barring Factor in Poultry. *Science* **29 :** 1004. Hannah-Alava, A. 1960, Genetic Mosaics. *Sc. Amer.* **202** : 118.

Lillie, F.R. 1917. The Freemartin: A Study of the Action of Sex Hormones in the Foetal Life of Cattle. *J. Exptl. Zool.* **23** : 371.

Morgan, T.H. 1910. Sex-limited Inheritance in *Drosophila. Science* **32 :** 120.

McKusick, V.A. 1960. Human Genetics. Prentice-Hall, N.J.

Sturtevant, A.H. 1945. A Gene in *Drosophila* that Transforms Females into Males. *Genetics* **30** : 297.

Whiting, P. W. 1939. Multiple Alleles in Sex Determination of *Habrobracon. J. Morph.* **66** : 323.

11

Variations in Chromosome Number

So far we have been dealing with genetic events on the assumption that genetic material is always constant in different species and during the life time of an organism. But this is not always so. Variations of many types occur causing harmful, useful or otherwise interesting alterations in the genome.

In a broad sense all changes in the genetic material which produce an altered phenotype are referred to as mutations. If the mutation involves cytologically visible changes in chromosomes we call them variations or aberrations in chromosomes. But if the change occurs in a gene, is cytologically invisible, it is conventionally referred to as a gene or point mutation. In this chapter only variations in chromosome number will be described.

Changes in chromosomes that are easiest to identify cytologically are those where an increase or decrease in chromosome number has occurred. Such numerical variations are known as polyploidy. When there is addition of one or more *entire* sets of chromosome (a set means all the chromosomes which constitute a haploid complement), the condition is known as *euploidy*. When one or more *single* chromosomes are added or deleted we call it *aneuploidy*. Sometimes variations involve *structural* changes in chromosomes and cytologically visible although sometimes difficult to identify. They are of four types: *deletions* or *deficiencies* involving loss of chromosomal fragments; *duplications* where segments of chromosomes are added; *inversion* involving intrachromosomal rearrangement of a portion so that the gene sequence becomes inverted; *translocations* in which an interchromosomal rearrangement occurs so that a segment of one chromosome becomes attached to another nonhomologous chromosome. These structural aberrations are dealt with in the next chapter.

EUPLOIDY

Most diploid sexually reproducing organisms have an alternation between a haploid and a diploid state in the life cycle. The haploid state, mostly confined to germ cells, is characterised by the presence of a single set of chromosomes (n). When two haploid germ cells produced by a male and a female parent unite during fertilisation, the zygote formed contains two haploid sets of chromosomes and it becomes diploid ($2n$). Mitotic divisions of the zygote produce a diploid adult organism of which each cell contains one haploid set of chromosomes from the

male parent, the other from the female parent. Since both sets of chromosomes are morphologically identical, pairs of identical chromosomes exist in a cell nucleus. The two members of a pair are said to be homologous to each other. When meiosis starts in the germ mother cells, homologous chromosomes attract each other and start pairing. The exactness of pairing depends upon the identical nature of the homologues.

In euploidy an organism acquires an additional set of chromosomes over and above the diploid complement. If one additional set is present the condition is known as triploid ($3n$); if two then tetraploid ($4n$); addition of three sets is called pentaploid ($5n$), and of four sets hexaploid ($6n$) and so on. This is known as a polyploid series and the individuals are said to show euploidy.

Depending upon the source of the additional chromosome set, euploids are classified into two types, *autopolyploids* and *allopolyploids* which are described below.

(*a*) Autopolyploids arise when the additional sets originate from the same species. For example, if the haploid set of a species is designated *A*, the diploid is *AA*, triploid *AAA*, tetraploid *AAAA*, and so on. Autotetraploids can arise through one of the following ways: (*i*) fertilisation of an egg by two or more sperms giving rise to a zygote with three or more sets of chromosomes; (*ii*) normal mitotic division in the diploid zygote in which chromosomes duplicate but cell division fails to occur so that four haploid sets of chromosomes produce a tetraploid nucleus; if the same mitotic error occurs during embryo development it results in some tetraploid tissues in an adult diploid individual; (*iii*) failure of meiotic division in germ mother cells so that unreduced diploid gametes are formed instead of haploids.

Although autopolypoids have homologous genomes, yet those having odd numbered sets of chromosomes such as $3n$, $5n$, $7n$, $9n$, and so on, show a high degree of sterility. This is because during meiosis, pairing between two homologous chromosomes only results in normal segregation of the haploid set into gametes. If three homologues are present they may or may not become paired to form a trivalent (Fig. 11.1). Since pairing in any region is restricted to only two homologues at a time, third homologue may fail to pair and remain a univalent, or may pair at some places to form a trivalent. During anaphasic segregation at meiosis *I*, two of these homologues may move into one daughter cell, and the third into the other daughter cell. Since all chromosomes of the haploid set have three homologues each, their random distribution or independent assortment will cause the resulting gametes to have varying numbers of different homologues. In this way a true haploid or a true diploid gamete would be rarely formed. Instead, unbalanced gametes with chromosome numbers ranging between n and $2n$ would be formed. Such gametes are not viable and triploids consequently are sterile.

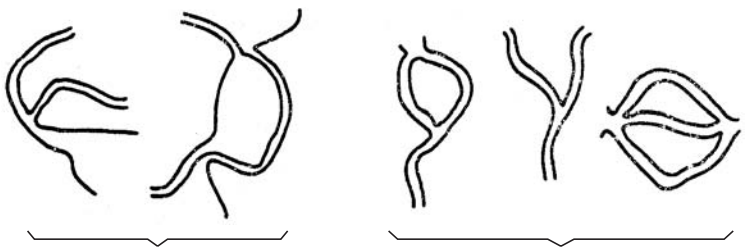


Fig. 11.1 Trivalent configurations in meiosis.

It is noteworthy that most popular varieties of seedless watermelons, bananas, Indian carpet grass, and European pears and apples are triploid. These triploid plants have resulted from fertilisation between diploid gametes from tetraploid plants and haploid gametes from

diploid plants. Once formed, the triploid plants are healthy and robust and are propagated through asexual cuttings.

Autotetraploids are either fertile or only partially sterile. As there are no univalents and trivalents formed, the four homologues can segregate to form viable diploid gametes, so that tetraploids are often fertile. Groundnut, potatoes and coffee are well known examples of autotetraploid species. Among cereals, autotetraploid rye is grown in Sweden and Germany. Some of the giant sized plants of *Oenothera lamarckiana* which DeVries first noticed in Holland and attributed to a mutation, had later turned out to be autotetraploids.

Polyploidy is more common in plants than in animals. More than 50 per cent of angiosperms are known to be polyploids. There are some explanations for this. Plants are mostly *hermaphrodites* or bisexual organisms in which sex chromosomes do not play a significant role in normal growth and development. An increase in the number of chromosomes is therefore desirable as it increases phenotypic variability and magnifies the expression of some favourable traits. In animals on the other hand, polyploidy leads to a disturbance in the balance between sex chromosomes and autosomes. An increase in the number of sex chromosomes markedly affects sexual development. Due to this, polyploidy in animals is restricted to those species which are hermaphrodites such as leaches and earthworms, or those which develop parthenogenetically (without fertilisation of egg), as in shrimps, aphids and some lizards. A second reason why polyploidy is more prevalent in plants is that the problem of sterility is easily overcome through asexual methods of reproduction in plants. Moreover, if one portion of a diploid plant becomes polyploid, for instance a branch bearing fruits, it is possible to propagate that branch through budding and grafting for raising new plants. Such techniques are obviously not applicable to animals except that individual polyploid cells can be excised and cultured in the laboratory. In mammals liver cells are often polyploid. Even germ cells such as primary spermatocytes are sometimes polyploid in mouse (Fig. 11.2) and in man. In plants cells of the tapetum which nourishes the male gametophyte, and endosperm cells which support the growing embryo, are also polyploid. The root nodules of leguminous plants frequently show polyploidy.

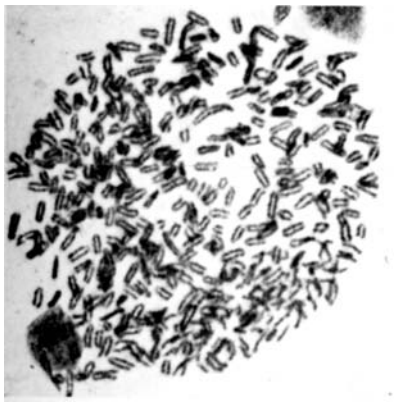


Fig. 11.2 Polyploid cell in spermatocyte (squash prepration)

From evolutionary point of view polyploidy has played a significant role in evolution of plant species. The origin of some important crop plants such as barley, potatoes, grass (*Dactylis glomerata*), lotus and many ornamental plants is due to polyploidy.

(*b*) Allopolyploidy: This is the second type of euploidy where the additional set of chromosomes comes from a different species. For example, suppose a diploid species with two

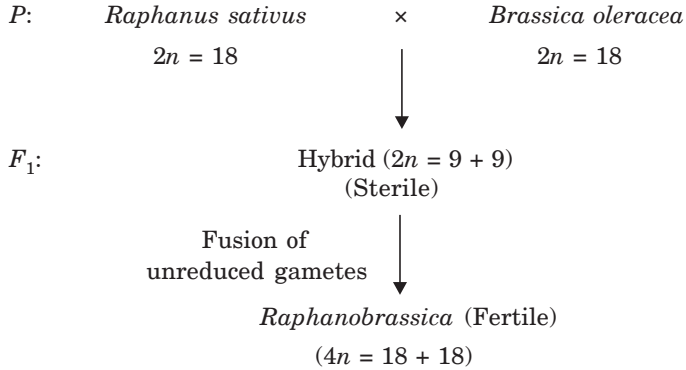
chromosome sets *AA* crosses naturally or artificially with another species *BB*. The offspring produced would be *AB* which is viable but sterile. This is because during meiosis the chromosomes belonging to the set *A* do not find homologous partners in chromosomes of *B*. Due to failure of pairing at anaphase *I*, the chromosomes move at random towards the two poles. Thus each gamete gets an unbalanced mixture of *A* and *B* chromosomes and sterility results.

There is one way of restoring fertility to a sterile hybrid (*AB*). If during mitotic division in the *AB* hybrid all the chromosomes are allowed to divide but cell division is inhibited, the result would be a tetraploid nucleus with two sets of *A* and two sets of *B* chromosomes (*AABB*). Therefore, when meiosis starts, all chromosomes belonging to one set of *A* will find homologous partners with the remaining *A* chromosomes and perfect pairing will result. Similarly, the two sets of *B* chromosomes will pair with each other and viable fertile gametes would be formed. Such an allopolyploid individual is called an *amphidiploid*.

It is possible to induce *amphidiploidy* artificially by treating young buds or seeds with the alkaloid colchicine, a mitotic poison which inhibits spindle formation, consequently cell division. This leads to all the duplicated chromosomes becoming included in a single tetraploid nucleus.

Raphanobrassica is an interesting example of a newly synthesised genus for illustrating allopolyploidy. In 1927 a Russian geneticist Karpechenko made a cross between *Raphanus sativus* (radish) and *Brassica oleracea* (cabbage) with the aim of producing a new plant that would have the roots of radish, and in the aerial portions would bear cabbage. The hybrid that was actually formed had the roots of cabbage and tops of radish plant.

The hybrid produced between radish and cabbage proved useless economically. But it proved very important genetically. Both radish and cabbage plants are diploid with 18 chromosomes. Thus gametes from each parent had 9 chromosomes, and their union produced the F_1 hybrid with 18 chromosomes. This hybrid was sterile because the 9 chromosomes of radish did not pair with the 9 chromosomes of cabbage. Sometimes however, viable pollen and ovules were produced having all 18 chromosomes. Fusion of such unreduced gametes produced tetraploid ($4n = 36$) plants with 18 chromosomes of radish and 18 of cabbage. Pairing took place *amongst* the radish chromosomes to form 9 pairs; similarly the cabbage chromosomes also formed 9 pairs. Normal segregation gave rise to viable gametes. The hybrid therefore became fertile and was given the name of a new genus *Raphanobrassica*. This is a beautiful demonstration of how a new genus can be artificially synthesised through allopolyploidy.



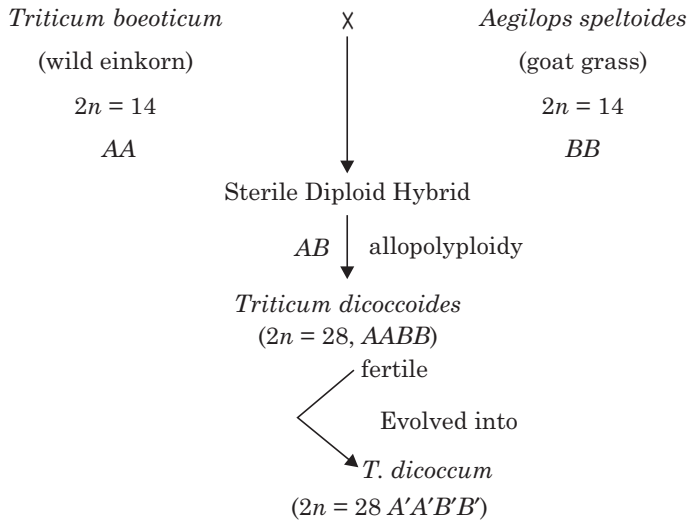
The genus *Triticale* demonstrates the efforts of man to create a new cereal by crossing wheat and rye. A hexaploid *Triticum* ($2n + 44$) is crossed to the diploid *Secale* ($2n = 14$). The tetraploid hybrid undergoes chromosome duplication to produce the octoploid *Triticale* which

combines the characters of wheat and rye. It is resistant to diseases affecting both wheat and rye, and the flour made from its grains has very high protein content. Therefore, efforts are being made to develop it for commercial use as a crop plant.

A number of cultivated plants are allopolyploids. One of the most important cereals, wheat, represents an allopolyploid series of diploid, tetraploid and hexaploid species. The series is represented by three groups designated *Einkorn* (single seeded), *Emmer* and *Vulgare*.

The *einkorn* group consists of two primitive diploid ($2n = 14$) species, namely *Triticum monococcum* and the wild *T. boeoticum*. Although not of much use for human consumption because the grain is tightly enclosed in the glumes, the einkorn species are useful as fodder. In some parts of Europe and the Middle East they are used for making dark breads.

The *Emmer* group consists of seven species of tetraploid wheats of which the most important are *Triticum dicoccum* (Persian emmer wheat) and *T. durum*. The origin of emmer wheats took place through hybridisation between an einkorn wheat and a wild species *Aegilops* (goat grass) as explained below.



Most of the emmer wheats are grown for animal feed, one *T. durum* has a high gluten content and is particularly useful for making chapatis in India and noodles in western countries.

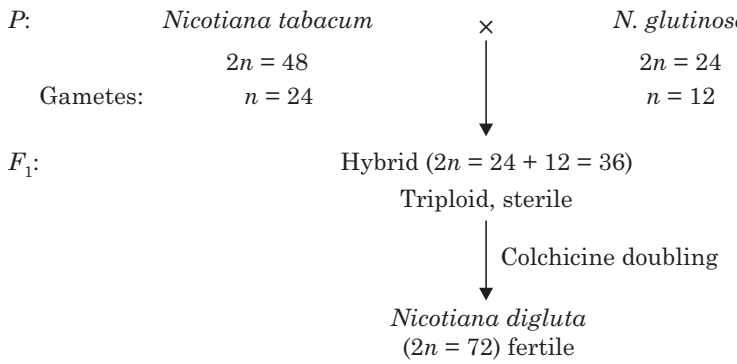
The *vulgare* group consists of five species of hexaploid wheats ($2n = 42$) including the economically important bread wheat *T. aestivum*. It is said to have originated through hybridisation between *T. dicoccum* (*A′A′B′B′*) and a different species of goat grass *Aegilops squarrosa* (*DD*) followed by chromosome doubling. The true bread wheat of today therefore, contains three genomes from three different wheats (*A′A′B′B′DD*; $2n = 42$).

Bread wheat has 21 pairs of chromosomes which show an interesting behaviour during meiosis. Normally, chromosomes of wheat species coming from different origins do not pair at meiosis. But the chromosomes belonging to *A*, *B* and *D* genomes that are present in hexaploid bread wheat pair with each other under one condition that chromosome No. 5 of *B* genome should be absent. Thus chromosome *I* of *A* pairs with chromosome *I* of *B*; chromosome *IA* can also pair with *ID*; and chromosome *ID* can pair with *IB*. Similar combinations of pairs exist for other chromosomes of *A*, *B* and *D* genomes. Such chromosomes which belong to different genomes, yet show pairing are called *homologous*.

Apparently there is a gene on the long arm of chromosome 5*B* of wheat which suppresses homologous pairing. Riley *et al.* in 1974 have given the name pairing homologous or *ph* to this gene. They have also found that at the beginning of meiosis the positions of the chromosomes on the nuclear membrane are determined by this gene, thereby affecting their pairing behaviour.

In cotton it has been possible to trace the origin of American cottons from hybridisation in the past between New World and Old World cottons. The American cultivated cottons have 52 chromosomes whereas the wild American cottons have only 26 chromosomes. The Indian cultivated cottons also have 26 chromosomes but these are morphologically different from the 26 chromosomes of the wild New World varieties. It appears that some time in the past the American wild cotton must have crossed with the Old World cultivated cotton to produce a hybrid with 13 New World and 13 Old World chromosomes. Chromosome duplication in this hybrid gave rise to the present day tetraploid ($2n = 52$) cultivated cottons in America.

Clausen and Goodspeed synthesised a new species of *Nicotiana* (tobacco) by induction of polyploidy as follows.



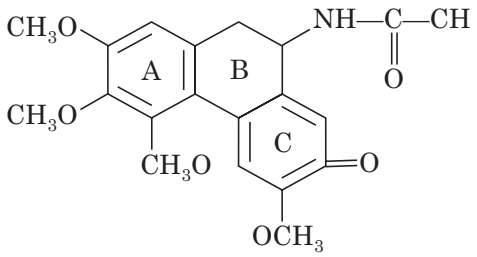
Like autopolyploids, allopolyploids are also of common occurrence in plants. Among animals they are extremely rare for some well defined reasons. There are usually no fertilisations of the interspecific type due to different behavioural patterns. Even if fertilisation is induced artificially, the hybrids formed show defects and do not grow normally. They cannot reproduce vegetatively and do not live long enough to allow chromosome doubling to take place in their germ mother cells to form diploid gametes. Nevertheless, allopolyploidy is found in some animals such as lizards (*Cnemidophorns*), fishes (*Poeciliopsis*) and salamanders (*Ambystoma*), all of which are parthenogenetic.

ARTIFICIAL INDUCTION OF POLYPLOIDY

In 1937 O.G. Eigsti first succeeded in inducing tetraploidy in cells of onion root tips by the chemical *colchicine*. This is an alkaloid obtained from the seed and corm of liliaceous plant *Colchicum autumnale*. In high concentrations the drug is highly toxic to cells. If used in concentrations as low as 0.01% to 0.5% it inhibits the formation of spindle fibers.

The molecular basis of colchicine action is now known. In the electron microscope spindle fibres of eukaryotes are found to be made up of numerous hollow, cylindrical structures known as microtubules, about 240 Å in diameter, and several microns in length. The hollow core of the microtubule is surrounded by a shell made up of globular protein called tubulin. Colchicine binds specifically to microtubules; one mole of colchicine binding to one mole of tubulin. The

binding leads to breakdown of microtubules (details in Chapter 6). Consequently spindle fibres do not form.



Colchicine

Due to absence of spindle fibres, chromosomes which have already duplicated remain in metaphase for a much longer time than they do in untreated cells, a condition described as the *c*-metaphase. Their centromeres divide, and as there is no anaphasic movement (for lack of spindle fibres), the nuclear membrane is formed around them and the cell enters interphase. Such a nucleus resulting from a *c*-mitosis contains double the number of chromosomes which would have otherwise been distributed to two daughter nuclei. In plants colchicine is applied to growing tips, meristematic cells, seeds or axillary buds in aqueous solution or mixed with lanolin.

ANEUPLOIDY

Aneuploids are individuals with an uneven number of individual chromosomes (Greek *aneu* meaning uneven). The cell has one or a few chromosomes more or less than the normal diploid number. In a diploid cell one complete set of chromosomes is present twice so that two homologues are present for each chromosome. If there is one additional chromosome it leads to a particular chromosome being represented by three homologues instead of usual two, and the condition is known as *trisomy* (symbolised as $2n + 1$). If there are two additional chromosomes in a cell so that one chromosome comes to be represented by four homologues, it is called *tetrasomy* ($2n + 2$). It can also happen that the two additional chromosomes are homologues belonging to two *different* chromosomes of the complement so that two chromosomes are represented each by three homologues. Such an individual is said to be trisomic for two different chromosomes ($2n + 1 + 1$).

The alternative situation also exists involving loss of a chromosome from a diploid complement. This results in the presence of one chromosome without a *homologous* partner called monosomy ($2n - 1$). If both homologues of a chromosome are missing, it is called nullisomy ($2n - 2$).

Aneuploidy results from an abnormal event of *nondisjunction* of single chromosomes during meiosis. The paired homologues might fail to separate at metaphase so that both homologues go to the same pole whereas the other pole receives neither homologue. Out of the four gametes formed from such a cell, two will have an extra chromosome each (trisomy) in addition to the haploid number ($n + 1$). The other two gametes will have one less (nullisomy) than the haploid number ($n - 1$). If fertilisation takes place with a normal haploid gamete (n) from the other parent, then these four gametes will produce two types of aneuploid individuals—trisomic ($2n + 1$) and monosomic ($2n - 1$). Nondisjunction could occur in autosomes or in sex

chromosomes. Unbalanced gametes can also arise due to failure of pairing (asynapsis or desynapsis) between meiotic chromosomes so that they remain as univalents. At metaphase univalents move at random to either pole leading to formation of unbalanced gametes. Trisomics can also originate from translocation heterozygotes (see Chapter 21), and from ionising radiation when it causes spindle disturbances. An autosomal trisomy has been reported by McClure *et al*. (1969) in the chimpanzee which resembled the human Down's syndrome in all respects. The Klinefelter's syndrome is known in male *XXY* sheep which showed testicular hypoplasia but not mental retardation. In plants trisomy has been extensively studied and complete series prepared, first in *Datura* by Blakeslee, and later in maize, barley, tomato and rye.

Trisomy in Datura

Datura stramonium, also known as Jimson weed or thorn apple proved to be most favourable for illustrating trisomy. In 1922 A. F. Blakeslee and J. Belling from an extensive study found a correlation between the morphology of the capsule and leaves and presence of trisomy for each chromosome of the complement. Normally *Datura* has 12 chromosomes in its haploid cells, therefore 12 different trisomics are possible. The investigations of Blakeslee and Belling started when they found a plant with 25 chromosomes, instead of 24 in a diploid cell. This plant further showed a globular capsule and broader leaves as compared to diploid plants. It was inferred that the globe phenotype was due to trisomy of one particular chromosome.

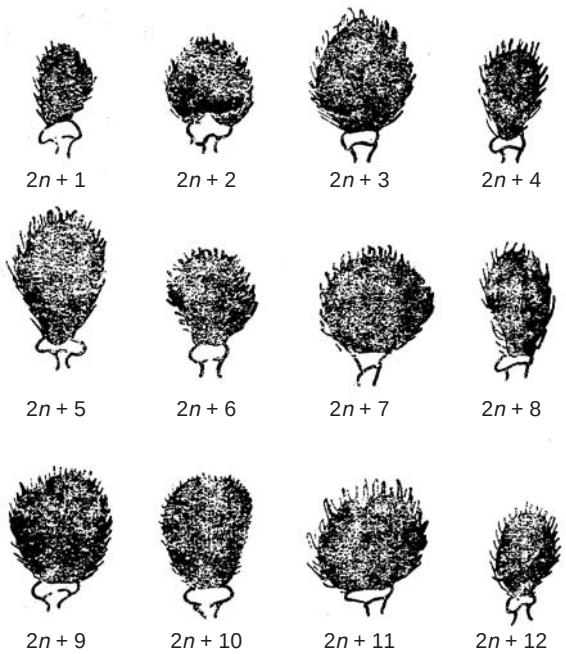


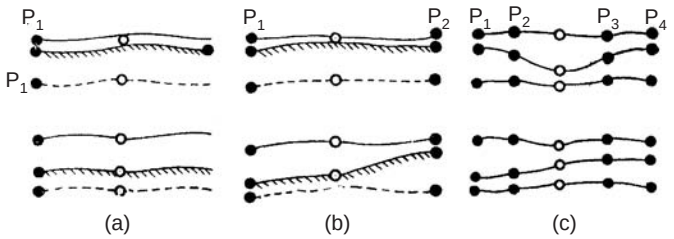
Fig. 11.3 The shape of capsule related with each of 12 different trisomies in *Datura.*

Crosses were then made between normal male plants and globe female plants. The globe phenotype appeared in about 50% of the progeny. Reciprocal crosses in which pollen came from the globe plants resulted in all normal progeny. Evidently, the extra chromosome passes to two of the four megaspores formed after meiosis in the developing female gametophyte. Pollen grains with the extra chromosome are not viable. Similarly, trisomics for all the 12 chromosomes were discovered, each producing a distinct capsular phenotype. Figure 11.3 shows

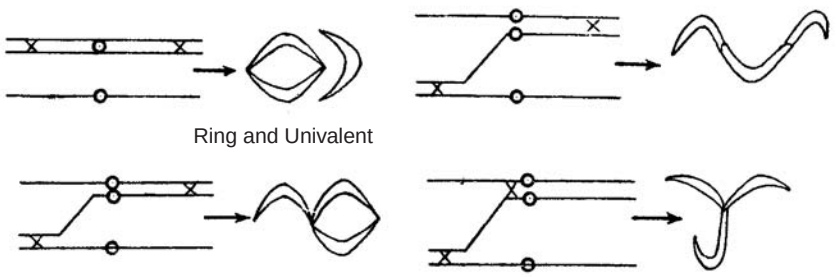
the different types of capsules resulting from trisomic states of the 12 chromosomes which have been arbitrarily designated *A–L*.

Chromosome Pairing in Trisomics

In trisomics three homologues of one chromosome are available for pairing. At any point only two chromosomes can pair and genetic exchange can take place. Different pairing configurations will form depending upon the number of points at which pairing is initiated: (*a*) when pairing is initiated at a single point P_1, there is no genetic exchange, a bivalent and a univalent are formed; pairing is influenced by specific affinity between chromosomes and particular regions of chromosomes. (*b*) Pairing is initiated at two points (P_1 and P_2), for instance at the ends of the chromosomes. Depending upon which ends pair (there are six ends), they can result in trivalent, bivalent and univalents in the gametes. (*c*) When the number of pairing points increases, there is more exchange resulting in a large number of trivalents and few bivalents and univalents. The following diagram explains the 3 situations.



When chiasmata form in all available segments, the metaphase configurations are indicative of the pairing relations. The bivalent and univalent condition becomes a ring and a univalent; the trivalent becomes a "frying pan". When chiasmata do not form in all segments, then rod bivalents, univalents, chain trivalents and *Y*-shaped configurations form at metaphase. These have been studied thoroughly in maize, rye and *Drosophila*.



ANEUPLOIDY IN DROSOPHILA

Nondisjunction of two *X* chromosomes leading to aneuploidy is discussed in Chapter 10. The strain of attached *X*-females further explains the consequences of aneuploidy. Besides sex chromosomes, autosomes in *Drosophila* can also have nondisjunction. The tiny dot-like chromosome 4 if present in the trisomic state produces viable but phenotypically deformed

offspring. But if one of the very large chromosomes 1 or 2 is present in the trisomic state, it results in lethality. Since there are many genes present on these two chromosomes, a gain or loss of one chromosome upsets the genetic balance to an extent enough to cause the lethal condition.

ANEUPLOIDY IN HUMANS

Nondisjunction can occur in both sex chromosomes and autosomes leading to a number of disorders in human beings, each phenotypically distinct from the other. Chromosome studies on foetuses obtained from spontaneous abortions have shown aneuploidy in about 20% of cases. It is perhaps a good thing that abnormal embryos die early before birth. But this is not always so. Unfortunately, some of the aneuploid zygotes survive to produce abnormal individuals which are born to suffer throughout their lives.

The most well known amongst the mentally retarded conditions is that of Mongolism or Down's syndrome resulting from trisomy of a *G* group chromosome (No. 21). The method of classifying chromosomes is described in Chapter 21. Individuals affected by this syndrome are characterised by a short stature, round face, short fingers and mental retardation of varying degrees. Their average life span ranges between 10 and 20 years, and *IQ* between 25 and 49. Their sexual development is incomplete/impaired and most of them, especially males are frequently sterile. Some are fertile and have produced children.

Besides Down's syndrome, mental retardation can also result from trisomy of chromosome 18 (Edward's syndrome) and chromosome 13 (Patau's syndrome). Nondisjunction of the sex chromosome leads to *XO* (Turner's syndrome) and *XXY* (Klinefelter's syndrome) conditions, all of which are described in detail elsewhere.

Sometimes nondisjunction occurs during mitotic divisions in the zygote. This leads to the formation of two or more cell lines with different chromosome numbers. Such an event results in a mosaic individual with phenotypic abnormalities related to the chromosomal anomalies in the cell lines. Thus Turner's females may be chromosomally *XO*/*XX*; Klinefelter's *XX*/*XXY* or *XXX*/*XXY*/*XXXY*; Down's syndrome $2n/2n + 1$, and so on.

HAPLOIDS

Diploid organisms have progeny when their unfertilised eggs develop parthenogenetically. Haploid animals have been produced experimentally in amphibians by chemical or mechanical treatment of unfertilised eggs. Haploids are more frequent in plants. By controlling pollination haploids have been produced in barley, maize and tobacco.

In potato interspecific crosses have been made using the natural autopolyploid *Solanum tuberosum* ($4n = 48$) and *S. phureja* to give dihaploids. The dihaploids are useful in potato breeding. In a strict sense the term dihaploid is used in the case of an allotetraploid, and trihaploid for an allopolyploid, and polyhaploid for an allopolyploid. It is possible to distinguish between unfertilised (haploid) seedlings from the fertilised (diploid) ones if dominant seedling characters are present in the male parent. The fertilised endosperm which determines the

phenotype of the seed inherits the dominant character from the male parent. In maize it is common to use seedling markers in the male parent to identify spontaneous haploids in cross breeding work. The importance of haploids lies in production of homozygous diploid lines by colchicine treatment and to use them in hybrid breeding. The doubled haploids are a good source of improved genetic material. However, on selfing the doubled haploids, their phenotypic homogeneity goes on decreasing in successive generations.

In *Datura innoxia* haploids were first developed from pollen grains by culturing the anthers (Guha and Maheshwari, 1964). When mature anthers were cultured, embryo-like structures called embryoids (later confirmed by the authors to have arisen from pollen grains) emerged from the inside of the anthers. The embryoids could produce haploid plantlets. Rice, tobacco, *Atropa*, tomato and a few more plants have yielded haploids by similar methods.

QUESTIONS

1. Why are triploid and pentaploid plants usually sterile? How then are the plants of triploid sugarbeets and watermelons maintained?
2. More than 50% of higher plants are natural polyploids, whereas animals are rarely so. Give reasons. If you had to hunt for more examples of polyploidy in animals, what type of species would you aim at?
3. A plant is trisomic for a certain chromosome so that 3 alleles P^1, P^2, P^3 are present in each diploid cell. What would be the genotypes of the gametes formed with respect to these alleles?
4. The cross between plants of radish and cabbage produced a useless hybrid having roots of cabbage and tops of radish. Yet the cross was significant in other ways and was repeated for other plants. Explain giving examples.
5. Fill in the blanks:
   (*a*) Both capsule shape in *Datura* and mongolism in human beings are due to ..... of chromosomes.
   (*b*) When there are $2n + 2$ chromosomes in a cell, the condition is known as .....
   (*c*) Nullisomy is represented as .....
   (*d*) The plant ..... is an example of an amphidiploid.
6. With diagrams only describe the meiotic behaviour of an autotriploid ($2n = 9$) showing (*a*) pairing at pachytene, (*b*) configurations at metaphase *I* and anaphase *I*, (*c*) the chromosome constitution of the four meiotic products.
7. A plant (*AA*) homozygous for a hypothetical gene *A* becomes trisomic so that the third chromosome carries the *a* allele. When this plant is crossed to a diploid plant of genotype *aa* (*a*) what proportion of progeny plants would be trisomic and *Aaa* in genotype? (*b*) what proportion would be diploid?

SELECTED READINGS

Astaurov, B.L. 1969. Polyploidy in Animals. *Annu. Rev. Genetics* **3 :** 99.

Blakeslee, A.F. and Belling, J. 1924. Chromosomal Mutations in the Jimson Weed, *Datura stramonium J. Heredity* **15** : 195.

Bungenberg de Jog, C.M. 1958. Polyploidy in Animals. *Bibliographica Genetics* **17** : 111.

Clausen, R.E. and Goodspeed, T.H. 1925. Interspecific Hybridisation in *Nicotina* II. A tetraploid *glutinosatabacum* hybrid. An Experiment Verification of Winge's Hypothesis. *Genetics* **10** : 278.

Ford, C.E. 1969. Mosaics and Chimaeras. *Br. Med. Bull.* **25** : 104.

Love, A. and Love, D. 1943. The Significance of Differences in the Distribution of Diploids and Polyploids. *Hereditas* **29** : 145.

McClure, H.M. *et al.,* 1962. Autosomal Trisomy in a Chimpanzee: Resemblance to Down's Syndrome. *Science* **165** : 1010.

Macgregor, H.C. and Uzzell, T.M. Jr. 1964. Gynogenesis in Salamanders Ralated to *Ambystoma jeffersonianum. Science* **143** : 1043.

McFadden, E.S. and Sears, E.R. 1946. The Origin of *Triticum spelta* and its Free-threshing Hexaploid Relatives. *J. Heredity* **37** : 107.

Stebbins, G.L. 1950. Variation and Evolution in Plants. Columbia Univ. Press, New York.

Walia, Karvita 1972. Effect of Asafoetida on Mouse Chromosomes. *Cytologia*, **37** : 719.

Winge, O. 1917. The Chromosomes—Their Numbers and General Importance. *Comp. Rend Trav. Lab.* Carlsberg **13** : 131.

12

Variations in Chromosome Structure

Structural changes within chromosomes do not change the amount of genetic material nor total number of chromosomes in a cell. Basically they are visible alterations in the shape of chromosomes and sequence of their segments which change the linear order of genes on a chromosome. Also known as chromosomal aberrations, they are transmitted through mitosis and meiosis and lead to changes in phenotypes and in the expected genetic ratios. Their study is useful for gaining knowledge about the nature of chromosomes and gene mapping.

Chromosomal arrangements occur spontaneously due to natural causes from time to time. In grasshoppers about one in a thousand individuals appears to carry a newly arisen cytologically visible chromosomal rearrangement in the germ line. In human beings also their incidence is high and would be deleterious, but luckily most of them are eliminated by natural selection.

Structural variations in chromosomes are of four distinct types: deletions, duplications, translocations and inversions. These and a special type of *B* chromosome are described.

DELETION OR DEFICIENCY

A deficiency means deletion of a small portion of a chromosome resulting in loss of one or more genes. A deficiency originates from breakage occurring at random in both chromatids of a chromosome (called chromosome break), or only in one chromatid (chromatid break). The breakage may be caused by various agents such as radiation, chemicals, drugs or viruses at any time during the cell cycle, either in somatic or in germ cells. Depending upon its location, a deletion may be *terminal* when a single break occurs near the end of the chromosome; or *interstitial* when two breaks occur in a middle portion of the chromosome.

Each break produces two raw ends which may behave in one of following three ways: (*a*) there might be reunion of the broken ends called *restitution* so that the original chromosome structure is restored; (*b*) the broken ends may not unite giving rise to a chromosomal segment without a centromere which is eventually lost during cell division; (*c*) if two single breaks occur in two different chromosomes in a cell, the deleted segment of one chromosome may unite with the raw broken end on the other chromosome; this is called *exchange union*.

Fate of a Deleted Fragment

If the fragment does not have a centromere (acentric), then at metaphase it will not be able to get attached to spindle fibres and move towards a pole with other centric chromosomes. It will remain at the centre of the cell and will not be included within any of the two daughter nuclei. It will be free in the cytoplasm and will eventually be lost (Fig. 12.1). In this way, the cell will lose one or more genes contained in the deleted fragment.

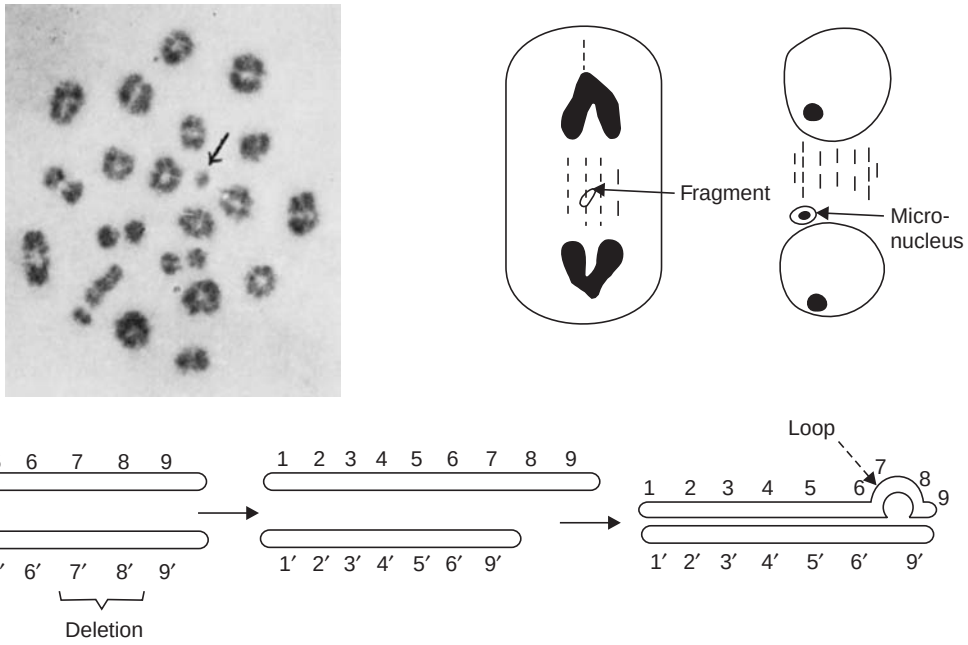


Fig. 12.1 Acentric chromosome fragment (arrow) at diakinesis.

A diploid cell has a homologue of the chromosome which has lost a segment. The corresponding segment of the intact homologue will have alleles of the genes that the cell has lost. Such a cell is said to be heterozygous for a deficiency. A very small deficiency in the heterozygous state is viable, but if homozygous it is lethal. When a deletion is large it is lethal even in the heterozygous state.

If a deletion occurs in cells of the germ line, then 50% of the gametes formed will have a deleted chromosome and 50% gametes would be normal. This would result in half the offspring with phenotypic abnormalities related to the genes carried on a small deleted fragment. If the deficiency occurs in a developing embryo, some cells would have normal chromosomes and other cells would have the deficiency. This could produce a mosaic individual with two different phenotypes.

Detection of Deficiency

The occurrence of a deficiency can sometimes be inferred from the results of a genetic cross when a rare recessive phenotype unexpectedly appears in the progeny. Consider a cross between two parents *DD* and *dd* where *D* controls the dominant expression of a trait, and *d* is the recessive allele. The *F*1 is expected to show the dominant trait and have the genotype *Dd*. If on the contrary, some *F*1 individuals show the recessive phenotype, one explanation could be sought in a deletion of the chromosomal segment bearing gene *D*. Since other interpretations

are also possible, it is best to confirm the occurrence of deficiency from a cytological study of the chromosomes as described below.

Deficiencies are best observed in preparations of homologously paired chromosomes at meiotic prophase either in large sized plant chromosomes or in polytene chromosomes. Normally during pachytene homologous chromosomes are intimately synapsed throughout their length. If one of the homologues is deficient over a small length, the corresponding portion of the second homologue has nothing to pair with. It therefore, forms a loop (Fig. 12.1), which is clearly visible in cytological preparations and is clear-cut proof that deficiency has occurred.

Suppose further that the deleted segment carries the dominant gene *D*. The recessive allele *d* is still present in the other homologue and expresses itself even in the single dose because the dominant allele *D* which normally suppresses it is absent. Such a phenomenon whereby a single recessive allele expresses itself in absence of the dominant allele is called *pseudodominance*. This also explains the results of the cross described above. *Pseudodominance* is similar to the hemizygous condition found exclusively in males where a recessive gene or a single *X* chromosome is expressed. Pseudodominance effect is observed in autosomal genes.

Due to the fact that deficiencies can produce unique phenotypic effects and the ease with which they can be identified by loop formation, they are important cytological tools for mapping genes. In *Drosophila* this method has been used to locate a number of genes on the banded polytene chromosome. In general it can be stated that each band represents a distinct gene. Since there are 5000 bands in *D. melanogaster* it is believed that there are about 5,000 genes in this species.

Deficiencies also produce phenotypic abnormalities in man. The cat-cry syndrome (*cri-du-chat*) where the baby cries like a cat is due to a deletion in the short arm of chromosome 5. A deletion in the long arm of chromosome 22 (*Ph*1 or Philadelphia chromosome) leads to chronic granulocytic leukemia. These examples are discussed in detail in Chapter 21.

In plants deficiencies are not easily transmitted to the progeny because their presence in developing pollen grains leads to pollen sterility. Nevertheless they have been observed in some

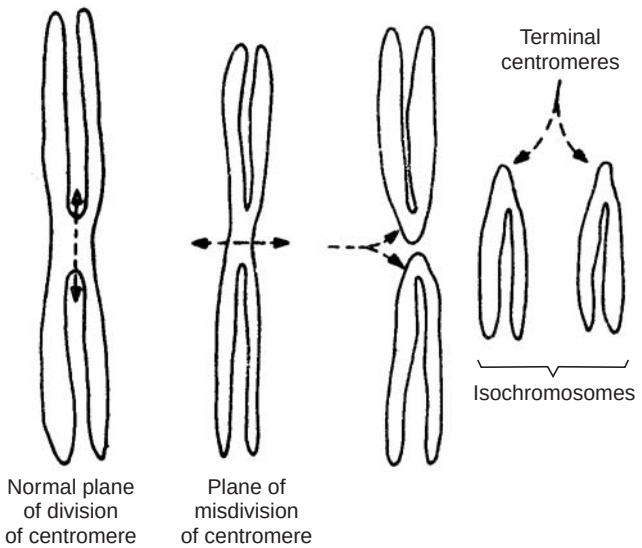


Fig. 12.2 Formation of telocentric chromosomes due to misdivision of centromere.

plants. In maize Creighton and McClintock have found that small deficiencies are viable even in the homozygous state. A special kind of single break occurs through the centromere of a metacentric chromosome giving rise to two isochromosomes with terminal telomeres. Such an event is also called misdivision of centromere (Fig. 12.2).

DUPLICATIONS

A duplication involves attachment of a chromosomal fragment resulting in addition of one or more genes to a chromosome. Whenever there is a duplication in a chromosome, there is a corresponding deletion in another chromosome. Following types of duplications are known (Fig. 12.3: the diagrams are self explanatory).

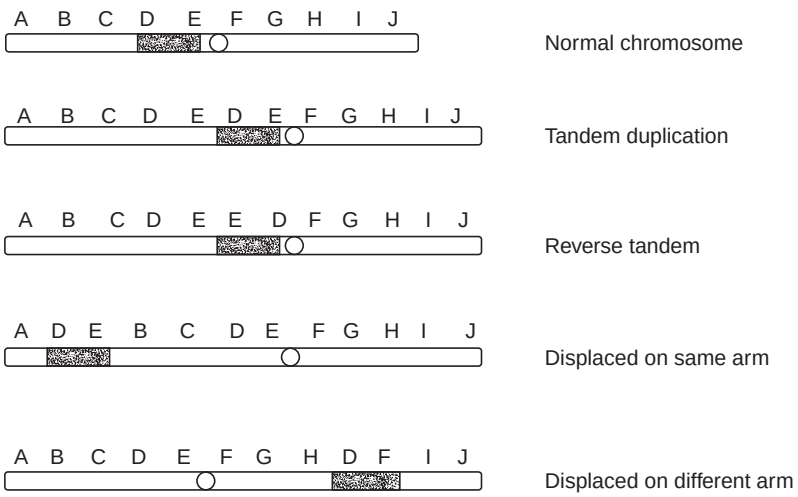


Fig. 12.3 Diagrams showing different types of duplications.

The phenotypic effect produced by a duplication is illustrated by the attached-*X* females in *Drosophila*. Consider such flies which are homozygous for some recessive sex-linked traits. It is found that when a fly receives a fragment of an *X* chromosome carrying the wild type allele from its male parent, then only the dominant phenotype is expressed. The recessive alleles of the same gene although present in the homozygous condition, are not able to express themselves. Evidently the presence of a single dominant allele in a duplication is enough to produce the wild type phenotype.

The origin of duplications can be traced to unequal crossing over during meiosis. Normally homologous chromosomes are paired in a perfect manner so that identical loci lie exactly opposite each other. The mechanism ensures that after crossing over between nonsister chromatids, equal exchange products are formed. If paired chromosomes are *misaligned*, it is not possible for exchange to take place between exactly opposite locations on two chromatids. Instead, exchange occurs between adjacent points on two chromatids so that one resulting chromatid will have a duplication, the other a deletion. Such an exchange is called unequal

crossing over. A gamete that receives a chromosome with a duplication will be diploid for some genes. When it fertilises a normal gamete, the zygote will have three sets of those genes that are present in the duplicated segment.

Bar eyes is a dominant *X*-linked trait in *Drosophila* females which provides a range of interesting phenotypes resulting from duplication. In a homozygous wild type female there is a large oval compound eye (nonbar) with about 779 facets. The Bar trait reduces the eye to a vertical bar with very few facets. Bridges analysed the salivary gland chromosomes of *Drosophila* and found that the Bar gene (*B*) was present on a region designated 16*A* of the *X* chromosome.

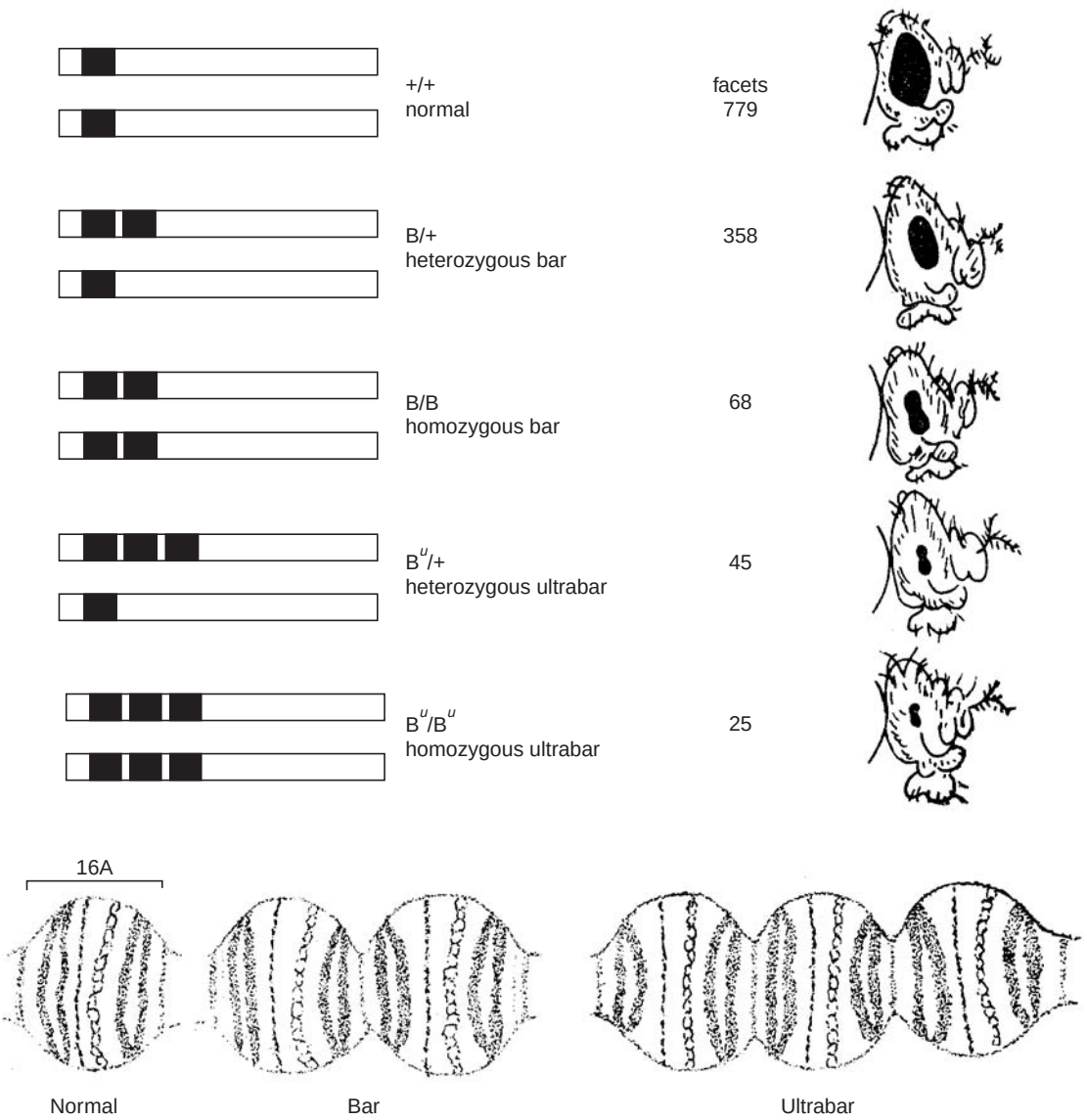


Fig. 12.4 Diagrams showing correlation between duplication of region 16A and inheritance of bar eye trait in *Drosophila*. Below is enlargement of 16*A* region of the chromosome.

When the band in the 16*A* region is present in duplicate in one *X* chromosome of the female (*i.e.* heterozygous for the duplication *B*/*X*), it results in an elongated Bar-shaped eye, smaller than the wild type (+/+) due to the presence of only 358 facets. When a female is homozygous for the duplication (*B*/*B*), the Bar-shaped eye is further reduced in size and has 68 facets. If there is unequal crossing over in a female homozygous for Bar (*B*/*B*), it results in one chromatid where the 16*A* region (Bar locus) is present in triplicate, and the second chromatid with only one Bar locus. Such a heterozygous triplicate condition produces a phenotype known as ultrabar (B^u) with only 45 facets. If the triplicate condition becomes homozygous (B^u/B^u), the result is a very small eye with only 25 facets (Fig. 12.4). Unequal crossing over is also responsible for a rare human haemoglobin known as haptoglobin, described later.

The Bar locus in *Drosophila* provides an explanation for *position effect*. According to this phenomenon the expression of a gene becomes altered when the position of the gene is physically changed.

Cytologically, a duplication is identified by the same method as deficiency, since in the heterozygous condition the extra fragment forms a loop in one of the two homologues.

TRANSLOCATIONS

Sometimes a segment of a chromosome becomes detached and unites with another nonhomologous chromosome. Such an interchromosomal rearrangement is called translocation. The rearrangements are of following types (Fig. 12.5):

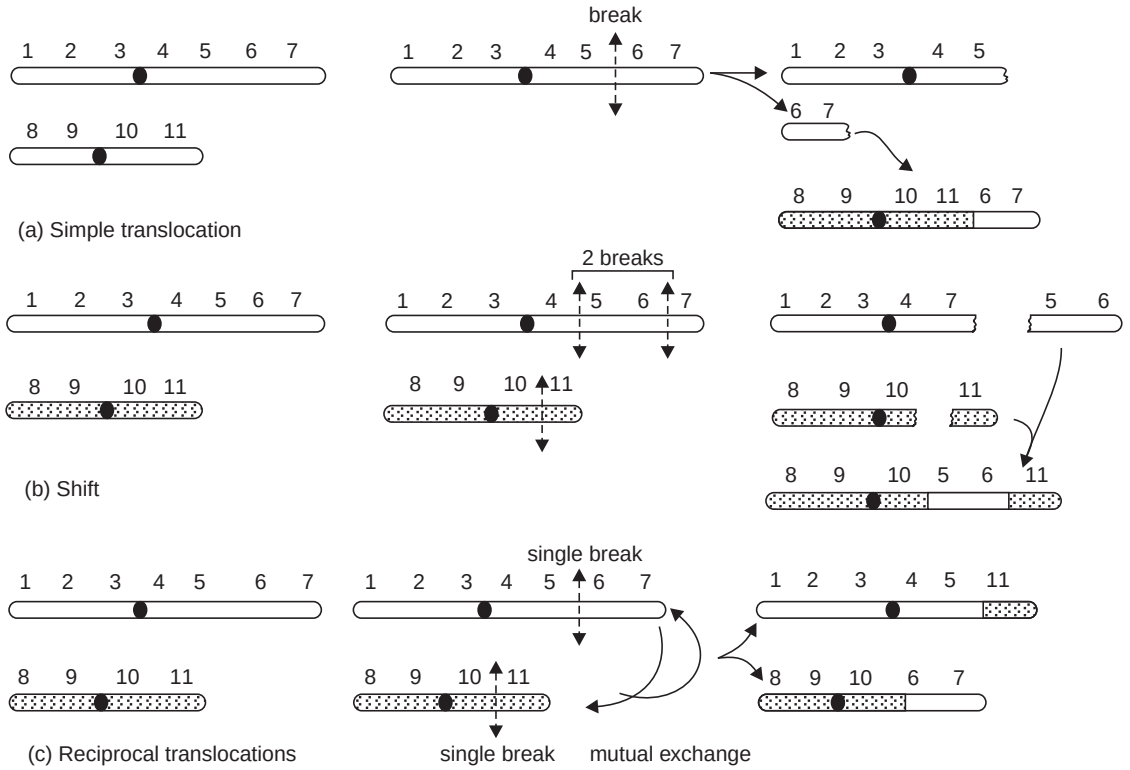


Fig. 12.5 The various types of translocations.

1. *Simple Translocation*: A single break occurs in a chromosome, and the broken fragment becomes attached to the end of another chromosome. However, due to the presence of "nonsticky" telomeres at the unbroken ends of a chromosome, such a terminal attachment of a segment does not take place.

2. *Shifts*: In this type three breaks are involved. Two breaks occur in a chromosome to produce an interstitial fragment. This fragment becomes inserted into one of the arms of another nonhomologous chromosome in which a single break has produced two "sticky" ends.

3. *Reciprocal Translocations*: These are the most frequent and extensively studied translocations. A single break occurs in each of the two nonhomologous chromosomes followed by a mutual exchange of the broken fragments. This results in two new chromosomes each having one segment of the other chromosome. Rarely, two breaks occur in each of the two chromosomes followed by exchange of intercalary segments. If the centromere containing segment of one chromosome is joined to the acentric piece of the other nonhomologue, the exchange is called *eucentric*. But if two centric pieces from two nonhomologues join to form a dicentric chromosome, it is called *aneucentric*. In the next division the dicentric chromosome will form a bridge and the acentric fragment will be lost. Therefore *aneucentric* exchange unions are usually lethal. The eucentric reciprocal translocations produce viable gametes if both pairs of nonhomologous chromosomes exchange segments.

Reciprocal exchange of segments involves no loss of genetic material. There is a qualitative change in the sequence of genes which is transmitted during mitosis and meiosis. Reciprocal translocations represent an important group of interchromosomal structural aberrations in chromosomes.

4. *Multiple Translocations*: Sometimes more than two pairs of nonhomologous chromosomes may be involved in a translocation as observed *Drosophila* and *Oenothera*. In 1930 Stern studied a multiple translocation system in *Drosophila* in which a segment of the *Y* chromosome became attached to the *X* chromosome. At the same time a reciprocal translocation occurred between the *X* and chromosome *IV*. This resulted in a female with 9 chromosomes instead of 8.

5. *Half Translocations*: When the nucleus containing two broken chromosomes is small, the broken ends are not widely separated in space and have better chance of undergoing reciprocal exchange. This is true for the small compact nucleus in the head of a sperm. In oocytes, on the contrary, due to the large nuclear volume the distance between the broken ends of non-homologous chromosomes may be so great that the chance for an exchange union is relatively small. In such a case only one exchange union may take place, leaving the other two broken ends free. This is called half translocation.

Cytology of Reciprocal Translocations

In translocation homozygotes meiosis is normal with regular bivalent formation at pachytene. At anaphase, movement to the poles is normal and viable gametes are produced. On the contrary in translocation heterozygotes pairing is complicated due to segments that have been exchanged between nonhomologous chromosomes. Instead of bivalents therefore, cross-shaped configurations of quadrivalents are formed at Metaphase I because of synapsis between homologous segments. Such interchange figures are more easily recognised in plant species with large chromosomes.

A translocation heterozygote has two normal and two interchange chromosomes. Following the rules of pairing these four chromosomes will form a cross-shaped interchange configuration shown in Fig. 12.6. The fate of chromosomes and the type of gametes that will be

produced depend upon the frequency and distribution of chiasmata. The chance that one or more chiasmata are formed in a certain segment depends on three factors: the length of the segment, the amount of exchange in the given organism, and characteristic properties of the segment that relate to chiasma formation. The formation of chiasma in the separate segments determines the frequencies of the different metaphase configurations. Conversely from the latter frequencies it is possible to estimate the chance that each segment has one or more chiasmata, or the intensity of crossing over in the segment.

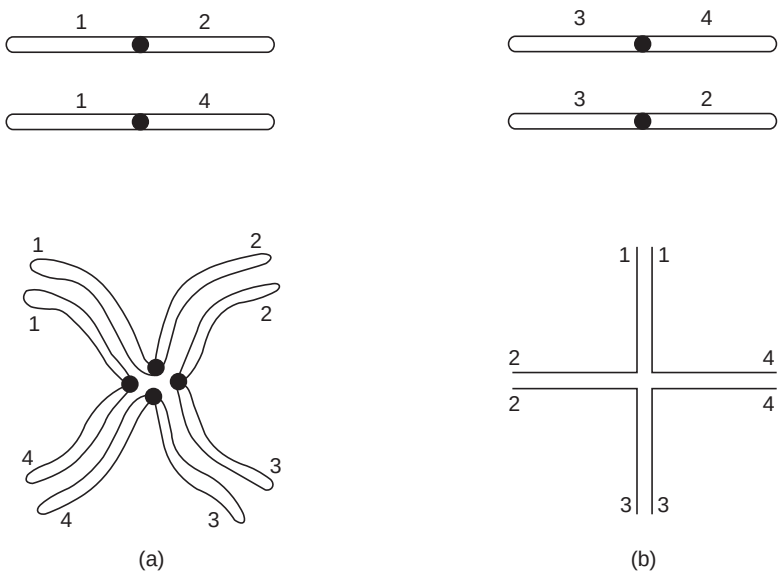


Fig. 12.6 The cross-shaped interchange figure formed in a translocation heterozygote.

When two opposite segments in the cross have a chiasma each, it results in two rod bivalents. If chaismata occur in two adjacent terminal segments, a trivalent and a univalent are formed. However, the ring and chain quadrivalents are the most common translocations. Sometimes two heterozygous reciprocal translocations may occur in the same cell. When there is one chromosome common to both, a multivalent of chromosomes is formed (hexavalent). When two interchanges occur in the same two chromosomes a quadrivalent can result.

It is possible to determine whether or not two translocations share a chromosome. This is done by making a double heterozygote and by observing whether two quadrivalents, a hexavalent, or a single quadrivalent are formed. In plants like barley, *Datura*, maize, rye and some others, a series of interchanges are known, involving all chromosomes at least once. The chromosomes involved in an interchange can be determined when the interchange is hybridised successively with all interchanges of the series, called a *tester set*. In one case a quadrivalent will be observed at meiosis; the known interchange and the one to be analysed have both chromosomes in common. In two cases a hexavalent is formed (the known and unknown interchange share one chromosome); and in the remaining cases two quadrivalents are observed. This is an efficient method of determining which chromosomes are involved in an unknown interchange.

In each of the four arms of the cross-shaped quadrivalent one chiasma is usually formed. At diakinesis two events take place: repulsion between homologues causing their separation, and terminalisation (movement) of chiasmata towards the distal end of the arms. At metaphase

therefore, the interchange figure becomes oriented to form an open ring or a twisted, zig-zag configuration (Fig. 12.7). In case chiasma does not form in one of the four arms, the cross-shaped complex opens up to form a chain. Anaphasic movement of chromosomes towards the poles takes place in one of the three different ways described below (Fig. 12.8).

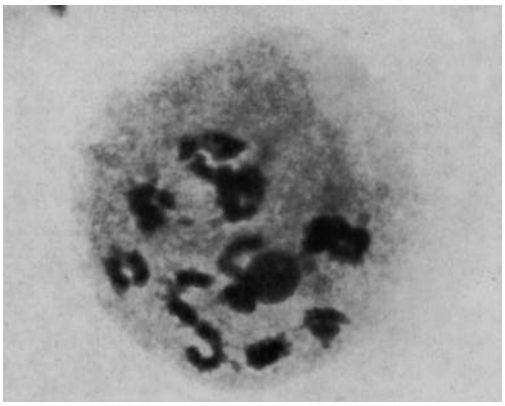


Fig. 12.7 A zig-zag configuration formed at pachytene due to absence of chiasma in one of the arms.

1. *Alternate Segregation*: The twisted orientation ensures perfect disjunction so that both translocated chromosomes 1′ and 2′ go to one pole, and both untranslocated chromosomes 1 and 2 go to the other pole. Thus all the remaining gametes will receive a full complement of genes and would give rise to viable individuals.

2. *Adjacent-1 Segregation*: This will take place in the open ring configuration whereby one translocated and one untranslocated chromosome will go to the same pole, in this way chromosome 1 and 2′ will go to one pole whereas 1′ and 2 will go to the other pole.

3. *Adjacent-2 Segregation*: Again in the open ring configuration, two homologous chromosomes 1 and 1′ will go to one pole, the other two homologues 2 and 2′ go to the other pole.

It is evident that both the adjacent types of segregation will give rise to gametes with duplications and deficiencies which would cause semisterility. The proportion of inviable gametes produced would be determined by the frequency of germ line cells having ring configuration. In animals gametes with duplication deficiency genomes are viable, but the zygote does not survive. Homozygous translocations can give rise to viable individuals if the paired homologues have normal crossing over and segregation at meiosis.

The site in the cross-shaped figure where crossing over occurs is important in estimating sterility. With respect to crossing over, each arm has a distinct *interstitial segment* which lies between the centromere and the break point of the translocation; the second is called *pairing segment* which represents portions of the arms of the cross beyond interstitial segments. Crossing over in the pairing segments has no effect on the segregation pattern as only homologous are exchanged. The ring and zig-zag arrangements obtained are in fact due to crossovers in the pairing segments. If crossing over occurs in the interstitial segment then nonhomologous portions are exchanged leading to the production of unbalanced gametes.

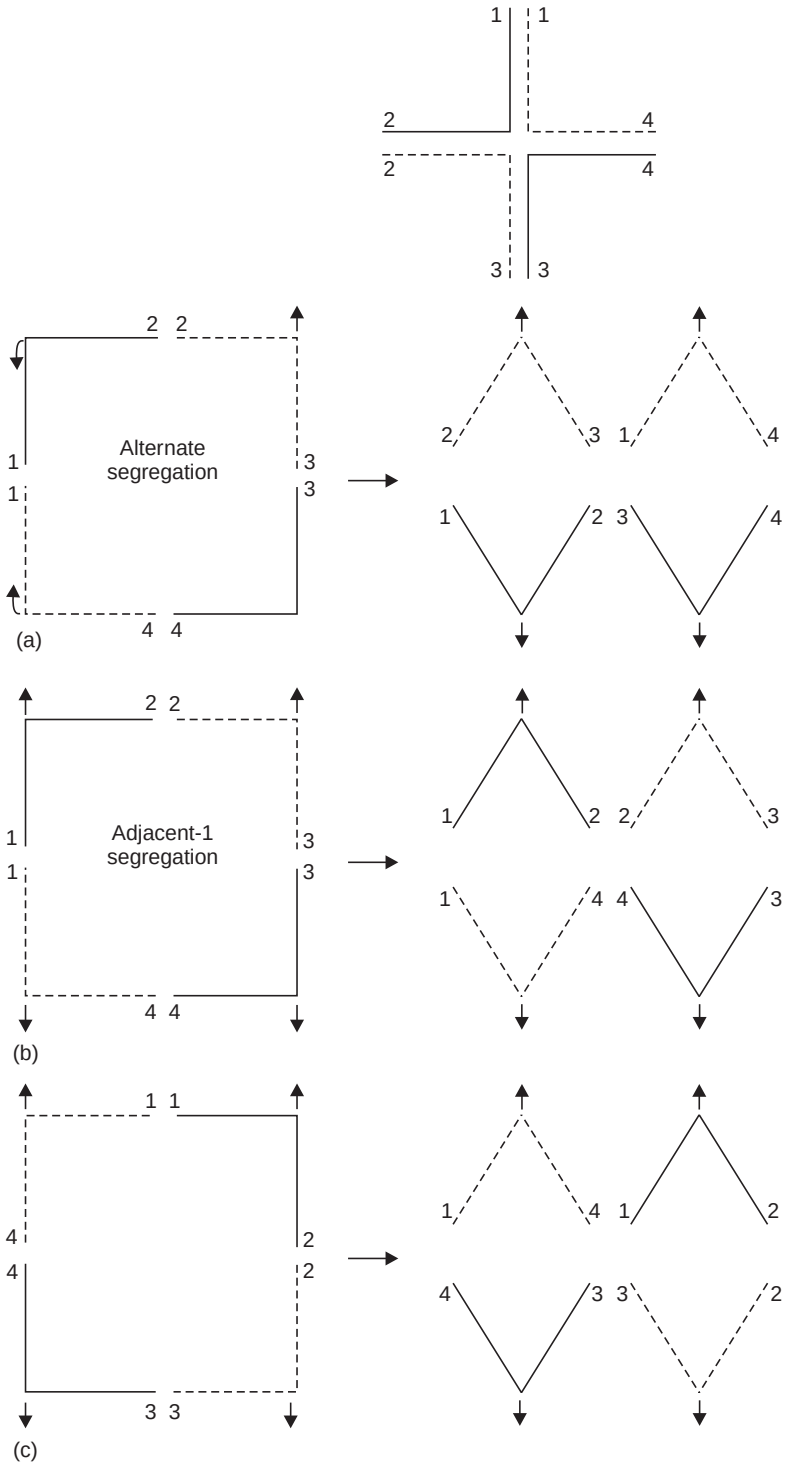


Fig. 12.8 Schematic diagram showing orientation and ways of movement of chromosomes to the poles in a translocation heterozygote.

Genetic Methods of Detecting Translocations

Translocations can be detected by performing genetic crosses and observing gene segregation. When translocation heterozygotes are selfed or crossed with each other, the progeny is of three types: normal homozygotes, interchange heterozygotes, and interchange homozygotes in the ratio 1 : 2 : 1 (Fig. 12.9 *A*, *B*).

If there are two reciprocal translocations in a cell that do not share any chromosome, then there is independent segregation. But if two translocations share a common chromosome there can be two consequences: (*a*) the same homologue of the common chromosome is involved in both translocations. The resulting balanced gametes are of two types, one with both translocations, the other with none. Alternate segregation is necessary to produce them. Thus when heterozygotes are selfed the progeny is in the ratio: 1 homozygous for both, 2 heterozygous for both, and 1 homozygous normal. (*b*) there is one chromosome shared by two translocations, but here one homologue is involved in one, the other homologue in the other translocation as indicated below.

This happens normally when two independently formed translocations are combined in one individual by hybridisation. Here also only two types of gametes are formed: one type having one translocation and the third chromosome normal; the other having the other translocation and also one chromosome of the three normal. When heterozygotes are now selfed they produce one homozygote for one translocation and also for one normal chromosome; two double heterozygotes and the normal types are formed due to crossing over in the different segment.

Translocations affect linkage relationships between genes in two ways: (*a*) in the homozygote linkage is changed; the genes in the translocated segment are not linked with the genes in the chromosome where they originally belonged; they are now linked to other genes. Study of this change in linkage can be used to detect a translocation and identify the involved chromosome. (*b*) In the translocation heterozygote all the genes on all the involved chromosomes are linked. This is because usually only balanced gametes take part in fertilisation or only balanced zygotes survive. Balanced gametes arise when either all interchange chromosomes or all normal chromosomes are present in one gamete.

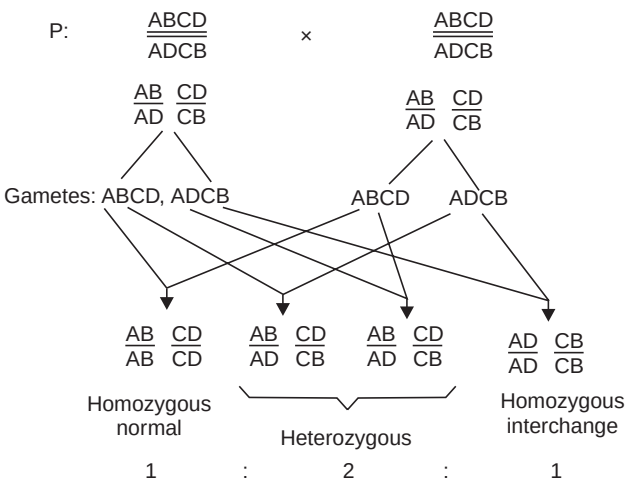


Fig. 12.9 A Types of progeny obtained on selfing a translocation heterozygote.

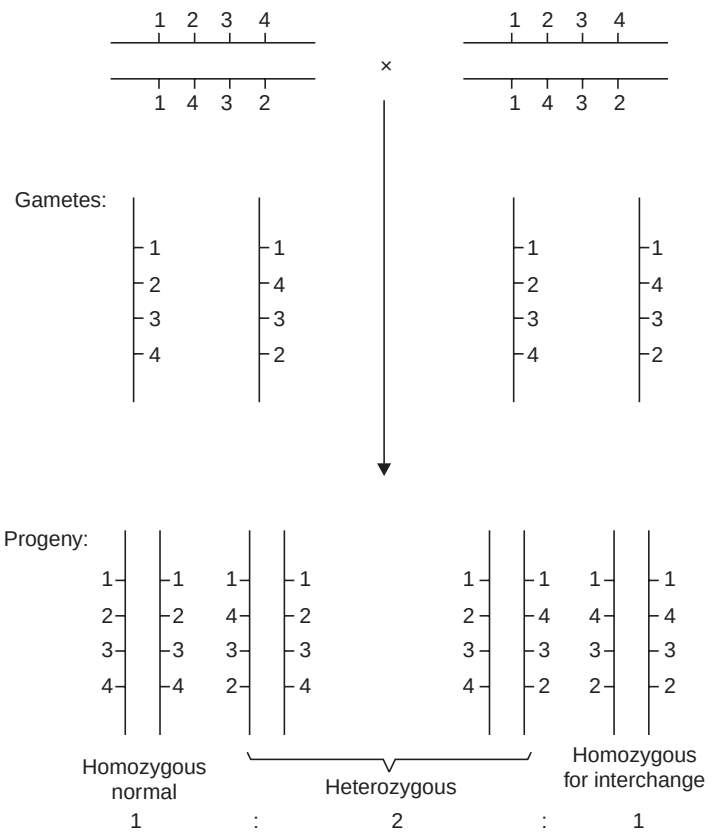


Fig. 12.9 B Types of progeny obtained on crossing two translocation heterozygotes. Fig. A shown diagrammatically.

Recombination between genes on different chromosomes, that is, between the gene and translocation takes place between the interchange breakpoint and the locus. The percentage crossing over between a locus and the break point can be calculated. The translocation can be detected from mitotic chromosomes. The heterozygotes can be located from multivalents at meiosis. The two homozygous types, the normal and interchange are not distinguishable from each other as both produce bivalents. An easier method of identifying heterozygotes is through their semisterility. The simplest analysis between the gene and translocation using heterozygotes can be done by a test cross which yields a 1 : 1 segregation for the translocation and also for the gene.

Translocation in Oenothera

The various species of the plant *Oenothera* (Onagraceae) are heterozygous for multiple translocations and show rings of chromosomes at meiosis. There are 14 chromosomes in the diploid cell of which some or all may be involved in translocations. On this basis the species of *Oenothera* form a graded series. *O. hookeri* is distinct in having 7 pairs of chromosomes and no translocations. The other species form rings of 6, 8, 10, 12 or 14 chromosomes at meiosis (Fig. 12.10). *O. lamarckiana* has a ring of 12 chromosomes and one bivalent pair. In *O. muricata* all the 14 chromosomes are united to form a giant ring. *O. biennis* shows one ring of 8 and another of 6 chromosomes.

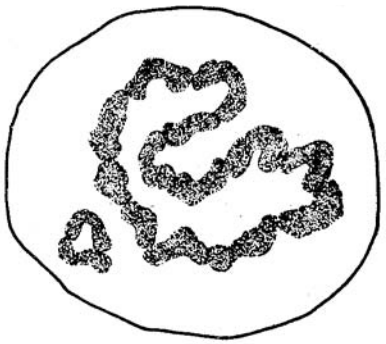


Fig. 12.10 A ring of 12 chromosomes formed due to translocation in *Oenothera*.

Similar instances of interchange heterozygosity are also known in some other plants such as *Rhoeo discolor* (Commelinaceae), *Isotonia* (Lobeliaceae), *Hypericum* (Hypericaceae) and 6 more genera of the family Onagraceae besides *Oenothera*. It is rare in animals. A few genera of scorpions like *Isometrus*, *Buthus* and *Tityus* show translocation heterozygosity and ring of chromosomes at meiosis.

There are certain genetic mechanisms which enforce permanent translocation heterozygosity in *Oenothera*. The cytogenetics of *Oenothera* has been worked out extensively but cannot be discussed here in detail.

INVERSIONS

Inversions result when there are two breaks in a chromosome and the detached segment becomes reinserted in the reversed order. They are classified into two types depending upon the inclusion or absence of the centromere within the inverted segment. Thus when both breaks occur in one arm of the chromosome it leads to a paracentric inversion; when a break occurs in each of the two arms, the centromere is included in the detached segment and leads to a pericentric inversion.

C H R O . M O S O M E	
C H R O . S O M O M E	Paracentric inversion
C H O M . O R O S O M E	Pericentric inversion

Meiosis is normal in inversion homozygotes. In heterozygotes pairing between homologous chromosomes is affected in the region of the inverted segment. Consequently, there is a suppression of recombination and fertility is impaired.

Paracentric inversions: This type of inversion is identified in the heterozygote by formation of a pairing loop at pachytene. If the size of the loop is large enough, chiasma formation will take place within it. When a single chiasma forms between an inverted and a normal segment, the two chromatids involved will produce one dicentric chromatid and one acentric fragment after exchange (Fig. 12.11). The other two chromosomes will be normal.

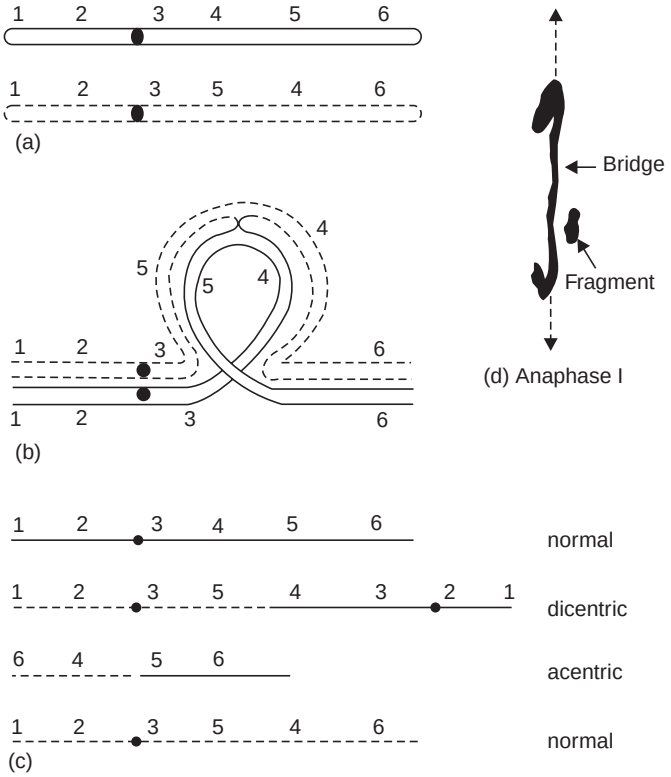


Fig. 12.11 Result of crossing over within the loop of a paracentric inversion. A chromosome bridge due to a dicentric chromosome and an acentric fragment are formed.

At anaphase I the dicentric chromosome will be pulled towards both poles, it will form a bridge that will ultimately break. The acentric fragment due to its inability to move would be eventually lost. Consequently, of the resulting four gametes, two would be normal and two deficient in chromosome segments. In plants deficient gametes are not viable (pollen grains that are deficient usually abort and are nonfunctional). In animals such gametes take part in fertilisation but either the zygote or the embryo aborts.

In an individual heterozygous for a paracentric inversion therefore, viable offspring are produced only by two of the four chromatids which did not have chiasma formation between them in the region of the loop. In each chromatid the gene sequence in the inversion segment will be of the nonrecombinant, parental type. Consequently, none of the offspring would be recombinants for genes present within the inverted segment. In this way a paracentric inversion suppresses recombination throughout its length.

In some insects and in *Drosophila*, individuals heterozygous for an inversion do not show reduction in fertility. In fact paracentric inversions occur frequently in natural populations of *Drosophila*. There are two explanations for this. One is absence of crossing over in male meiosis. The second is occurrence of four products of female meiosis in linear order of which the middle two egg nuclei have the deficiency; the peripheral two nuclei are functional and fertilised. They produce viable offspring of the parental type.

Pericentric inversions: In an individual heterozygous for a pericentric inversion, the centromere is present within the loop. When chiasma formation takes place within the inverted segment the chromatids resulting after exchange do not form a dicentric and acentric fragment as in a paracentric inversion heterozygote. Instead, they have one centromere each, but are deficient for some segments whereas other segments are duplicated (Fig. 12.12). The exchange segments produce inviable gametes and offspring. As in the case of pericentric inversions, the two chromatids not involved in crossing over only produce viable offspring with parental combination of genes present in the inverted segment.

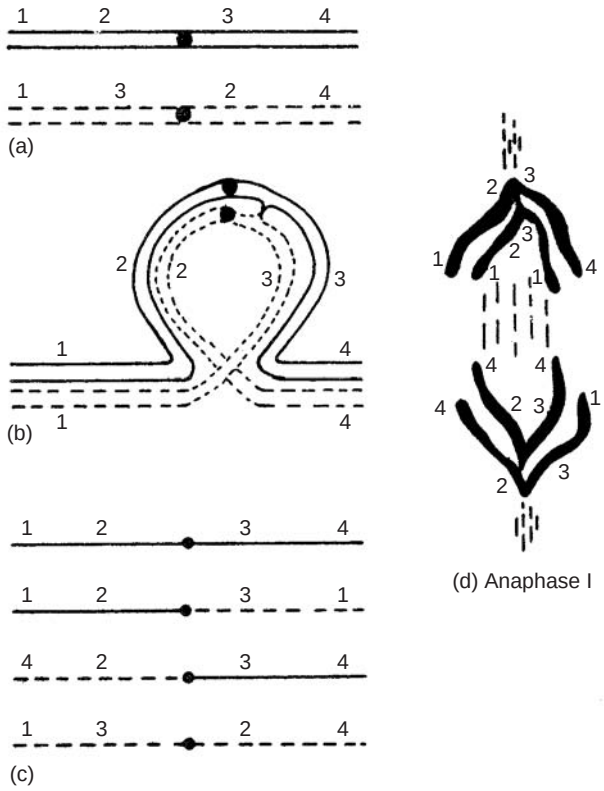


Fig. 12.12 Result of crossing over in the loop of a pericentric inversion; formation of duplication and deficient segments.

Due to the suppression of recombination the genes present in the inverted segment segregate as a single unit called *supergene* within a population. Inversions are easy to identify in the banded polytene chromosomes of *Drosophila* larvae and have been extensively studied.

B-CHROMOSOMES

In addition to the normal chromosome complement, a number of plant and animal species have extra chromosomes called *B*-chromosomes (normal complement in such cases is designated *A*). They are smaller than the *A* chromosomes, they do not pair with any *A* chromosome during

meiosis, and apparently do not serve any vital function in the organism. However, they persist in the population without conferring any obvious advantages. Their number is variable within species and among individuals within a population, in others they may be lacking. As they are not required for normal growth and reproduction, they have been considered to be genetically inert and dispensable. However, the recent work on corn and rye has shown that they have a few active genes and they perform certain functions.

B-chromosomes have been extensively studied in plants of *Zea mays* (maize) and *Secale cereale* (rye). They behave abnormally during mitotic division in the uninucleate pollen grains which gives rise to a small generative cell and a large vegetative cell. At anaphase of mitosis, the daughter chromatids of *B*-chromosomes fail to separate even though the centromeres have divided. Due to nondisjunction both chromatids move together toward the pole which forms the generative nucleus. Later on when the generative nucleus divides to form two male gametes, *B*-chromosomes segregate normally. There is preferential fertilisation of eggs by male gametes which carry *B*-chromosomes.

QUESTIONS

1. Suppose in a cross between two plants of genotypes *AA* and *aa* some of the F_1 progeny show the phenotype of *aa*. What would you suspect? How would you proceed cytologically to confirm your suspicion?
2. The gene *A* has an allele *a*. Is it possible for a pair of homologous chromosomes in a diploid cell to carry alleles *Aaa*? If so, explain the genetic mechanism involved.
3. A plant heterozygous for an inversion *ABCDE/ABDCE* is crossed to another plant homozygous for an inversion *ABCDE/ABDCE*. What percentage of gametes produced by both parents would have a complete haploid set of genes under the following conditions; (*a*) occurrence of one cross-over in the inverted segment *DC*, (*b*) absence of crossover in the inverted segment in all spore mother cells.
4. A plant of *Oenothera* is heterozygous for a reciprocal translocation between two nonhomologous chromosomes *L*. *MNO* and *PQ*. *RS* where period indicates location of centromers. Draw figures to show (*a*) their pairing behaviour at pachytene, (*b*) the kinds of gametes that would form if a single crossover occurred between *P* and *Q*, (*c*) the kinds of gametes resulting from a cross over between *M* and *N*.
5. Comment on (*a*) aneucentric chromosomes, (*b*) Bar locus in *Drosophila*, (*c*) translocation heterozygotes in *Oenothera*.

SELECTED READINGS

Baker, W.K. 1968. Position Effect Variegation. *Adv. Genetics* **14** : 133.

Bridges, C.B. 1925. Sex in Relation to Chromosomes and Genes. *Amer. Nat.* **59** : 127.

Burnham, C.R. 1956. Chromosomal Interchanges in Plants. Bot. Rev. **22** : 419.

Carlson, P. 1978. The *B* Chromosome of Corn. *Annu. Rev. Genetics* **15** : 5.

Cleland, R.E. 1962. The Cytogenetics of *Oenothera. Adv. Genetics* **11** : 147.

Ford, D.E. *et al.,* 1959. A Sex Chromosome Anomaly in a Case of Gonadal Dysgenesis (Turner's syndrome). *Lancet* **1** : 711.

Garber, E.D. 1972. Cytogenetics, An Introduction. McGraw-Hill, New York.

Grell, R.F. 1967. Pairing at the Chromosomal Level. *J. Cell Physiol.* **70** Suppl. 1 : 119.

Khush, G.S. and Rick, C.M. 1968. Cytogenetic Analysis of the Tomato Genome by Means of Induced Deficiencies. *Chromosoma* **23** : 452.

Kihara, H.Q. and Shimotsuma, M. 1967. The Use of Chromosomal Interchange to Test for Crossing Over and Chromosome Segregation. *Seiken Zoho* **19** : 1.

McClintock, B. 1942. The Fusion of Broken Ends of Chromosomes Following Nuclear Fusion. *Proc. Natl. Acad. Sci.* U.S. **21** : 458.

Muller, H.J. 1956. On the Relation between Chromosome Changes and Gene Mutations. *Brookhaven Symp. Biol.* **8** : 126.

Rhoades, M. M. 1933. A Cytological Study of a Reciprocal Translocation in *Zea. Proc. Natl. Acad. Sci. U.S.* **19** : 1022.

13

Nature of the Genetic Material

By the beginning of the twentieth century Mendelism had gained enough ground as the physical basis of inheritance. Biochemists then focussed their attention towards identifying the chemical nature of the hereditary material. Friedrich Miescher, a Swiss biochemist working in South Germany had already in 1869 identified a chemical compound of pus cells and salmon sperm. He presumed that this chemical was contained in the large nuclei of these cells and gave it the name *nuclein*. Later on it was found to be acidic and called *nucleic acid*. During the first quarter of the present century, all plant and animal materials were found to contain nucleic acid. In fact nucleic acids were found to be of two types designated DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). In 1924 Feulgen introduced a specific stain for DNA by which this type of nucleic acid was found to be localised in nuclei and chromosomes in contrast to RNA which occurs primarily outside the nucleus in the cytoplasm.

Simultaneously, proteins were being investigated extensively. As the physical nature of DNA was not well understood at that time, it failed to illuminate any problem in genetics. Many workers therefore, assigned the role of hereditary material to the proteins. It was only from the experiments of Griffith in 1928 and the correct interpretation of his results in 1944 by Avery, Macleod and McCarty that transmission of DNA in heredity was correctly understood.

DNA IS THE GENETIC MATERIAL IN BACTERIA

Griffith was working with the micro-organism *Diplococcus pneumonia* also called *Pneumococcus*, a bacterium causing pneumonia. Some strains of this bacterium have a smooth polysaccharide capsule which causes disease and are designated as virulent *S* type strains. In a mutant strain the bacteria have no capsule, and when grown on the surface of an agar plate, they form small rough colonies and represent avirulent *R* strain.

The pathogenicity of the bacterium (ability to produce pneumonia) lies in the capsule. Colonies which appear rough due to lack of capsule are nonpathogenic or avirulent. The presence or absence of the capsule and its chemical composition are heritable features of the bacterium. Capsules of different compositions are classified as type II and type III. In the same decade, mutations were being studied as stable heritable changes in genes. It was found that one in a million IIS bacterium could mutate to an avirulent, nonencapsulated strain of RII type. Similarly,

and with the same frequency, a IIIS strain could mutate to a type designated IIIR. Very rarely, a IIR bacterium could mutate back to IIS, and a IIIR to IIIS. It is noteworthy that a IIR could mutate back only to IIS, never a IIIS; a IIIR to IIIS, but never IIS.

Griffith found that when living *S* type bacteria are injected into mice they show symptoms of pneumonia and die; if living *R* type is injected, the mice continue to live. It was also found that if *S* type of bacteria are killed by heating to 65°C they become avirulent. Griffith mixed heat-killed IIIS cells with living IIR and injected them into a mouse. Contrary to expectations, the animals did get pneumonia. When blood of affected mice was sampled, living IIIS *Pneumococcus* was found in it (Fig. 13.1). These IIIS cells could multiply and maintain their characteristics for many generations. How did this happen?

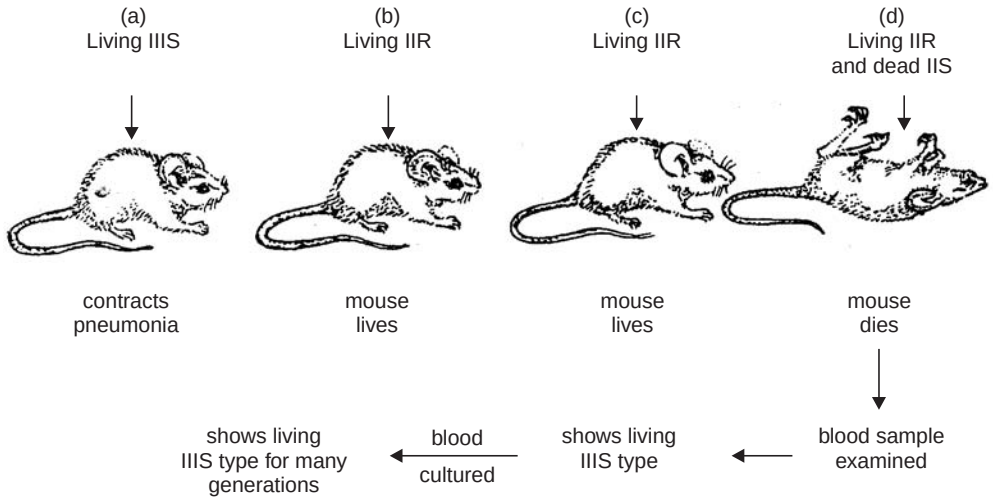


Fig. 13.1 Experiment of Griffith illustrating transformation of IIR cells of *Pneumococcus* into IIIS type.

If mutation had occurred, living IIR could mutate to living IIS, not IIIS type. The temperature at which IIIS cells were heat-killed turned out to be important. When very high or very low temperatures were used for killing IIIS, the avirulent strain failed to show features of the virulent strain. Evidently some heat-stable component present in dead bacteria of type IIIS could confer characteristics of this strain on living IIR cells. The component was hypothesised by Griffith to be a transforming principle which was released by heat-killed cells of type IIIS and taken up by avirulent IIR cells, thus transforming their hereditary properties into those of type IIIS.

By further experiments an *in vitro* method was devised for detecting transformed cells of *Pneumococcus*. When serum containing antibodies specific for IIR cells is added to a tube, the IIR cells agglutinate and clump at the bottom of the tube. The transformed cells are not agglutinated; instead they grow and produce a cloudy suspension.

Stated briefly, transformation occurs· when (*a*) bacteria of one genotype (for example *R* strain of *Pneumococcus*) are recipients of DNA extracted from bacteria of a different genotype (example *S* type of *Pneumococcus*); (*b*) transformed cells undergo stable alteration in genotype and new phenotype is expressed through many generations. Although Griffith did not identify the transforming principle, yet his experiments set the stage for the discovery of the hereditary role of DNA.

In 1944 three scientists, O. T. Avery, C.M. Macleod and M. McCarty isolated and identified the chemical nature of Griffith's transforming principle. They could show experimentally that

highly purified DNA from heat-killed type IIIS cells could transform IIR cells into IIIS type in the *in vitro* system. They could also demonstrate that enzymes that degrade DNA (the deoxyribonucleases) could destroy the transforming factor. Addition of the enzyme ribonuclease (which degrades RNA) or proteolytic (protein digesting) enzymes had no effect on transforming ability. The DNA induced transformation was a permanent and heritable characteristic. This was the first conclusive evidence that DNA indeed was the transforming principle, and consequently also the genetic material involved in hereditary mechanism in bacteria.

Later on other bacteria like *Hemophilus influenzae, Bacillus subtilis* and *Escherichia coli* were also shown to undergo transformation. Furthermore, bacterial strains which are sensitive to antibiotics like penicillin and streptomycin can acquire permanent resistance to these antibiotics by transformation with DNA from resistant strains.

Mechanism of Transformation

Transformation is the process of adding a foreign DNA fragment from a donor genome into genome of a recipient cell. The donor fragment passes through the cell membrane of the recipient cell (which may or may not belong to the same species) and becomes incorporated into the latter's genome through recombination. Transformation is detected by the presence of new cell phenotypes in the recipient cell's progeny.

The actual process of transformation is achieved in at least 3 steps as follows:

(*a*) Interaction between a transforming DNA fragment and the recipient bacterium. For successful transformation of a *Pneumococcus* cell the DNA fragment must have a minimal size of 800 nucleotide pairs; for *Bacillus* 16000 nucleotide pairs are necessary. The number of DNA fragments which can interact with a recipient cell is about 10, as these many receptor sites are present on the cell membrane. The recipient cell must be physiologically *competent* to receive a donor DNA fragment. This state of competence is acquired during the brief period when the recipient cells are dividing at a maximal rate.

(*b*) The next step involves movement of DNA fragment in the recipient cell cytoplasm by association with an enzyme system. At the time of its initial contact with the recipient cell, the donor DNA fragment is double stranded; but inside the recipient cell it becomes single stranded.

(*c*) Finally, the single stranded donor DNA gets covalently incorporated into the double stranded recipient genome and is replicated along with the host cell chromosome.

EVIDENCE THAT DNA IS THE GENETIC MATERIAL IN VIRUSES

The conclusion reached by Avery, Macleod and McCarty was not immediately accepted. Some investigators suggested that the DNA preparations contained a mutagen which caused mutation to the *S* form. Others pointed out that transformation was due to traces of some specific protein remaining in the DNA samples used. One apparent reason why the significance of this work was only gradually recognised was the absence of knowledge about bacterial genetics at that time. The existence of a chromosome capable of exchanging genes was not fully known in bacteria.

Moreover doubts continued as to whether DNA or protein was functioning as the genetic material. DNA was mostly ruled out because its only variable components were the four nitrogenous bases (the sugar and phosphate being identical in all nucleotides). On the other hand the molecular complexity of a protein, due to twenty different amino acids in various sequences could better

account for the diversity required in a genetic material for performing the varied functions in an organism.

Looking back it seems that it was perhaps due to lack of knowledge that the great achievement of Avery, Macleod and McCarty was not rewarded with a Nobel Prize. Their discovery however, did have the impact of initiating studies in molecular genetics with the use of micro-organisms.

One of the finest demonstrations that DNA is the genetic material came in 1950's through investigations on bacterial viruses by A. D. Hershey and Martha C. Chase. They were studying the mechanism by which the virus infects *E. coli.*

Bacterial viruses known as bacteriophages are obligate parasites which must infect a host cell in order to reproduce. The bacterial viruses most studied are those that infect *E. coli* bacteria particularly those classified as *T*2. It was known at that time that *T*2 consists exclusively of DNA and protein. In EM the *T*2 virus has a hexagonal body and a protruding tail. When a suspension of these viruses is mixed with a suspension of *E. coli* and left at 37°C, the viruses infect the host cells by becoming attached by their tails to the bacterial cell wall. After about 30 minutes of this attachment, the bacteria burst (lyse) releasing several hundred newly synthesised viruses.

The noteworthy point in the above experiment is that the original infecting viruses can still be seen attached to the bacterial membrane. Hershey and Chase became interested to find out what molecular event occurred between the time of *T*2 attachment to *E. coli* and the release of new virus progeny. Obviously, some material contained in the infecting viruses must have passed into the host bacterium where it caused the formation of new viruses. This must be the genetic material of the virus.

Hershey and Chase now prepared bacterial viruses in which either the phosphorus or sulphur were radioactively labelled. By doing so they could distinguish between viral DNA and viral protein and trace them during the process of infection. They did this by first growing *E. coli* cells in a nutrient medium containing radioactive phosphate (^{32}P) and sulphur (^{35}S). The labelled *E. coli* cells were used as hosts for unlabelled viruses. The resulting virus progeny labelled with ^{32}P and ^{35}S was used to infect unlabelled *E. coli* cells in a series of experiments popularly known as the waring blender experiments (Fig. 13.2).

Since sulphur is present only in protein and not in DNA and phosphorus only in DNA but not in protein, it was clear that in the labelled viruses, the protein component was selectively labelled with ^{35}S and the DNA with ^{32}P. When a suspension of labelled viruses was mixed with unlabelled *E. coli* cells and left for a few minutes, the viruses were observed in *EM* attached to the bacteria. By stirring the mixture in a waring blender, the attached viruses were broken away from the host cell. The suspension was divided into two parts. One portion was incubated for some more minutes and the bacteria burst to liberate the newly formed progeny virus. The remaining suspension was cooled and centrifuged to separate the bacteria from the remains of attached viruses. The amounts of radioactive sulfur and phosphorus were assayed in each. They found that 85% of ^{32}P of the viruses had passed into the bacteria whereas 80 % of ^{35}S had remained in the viruses. It was apparent therefore that most of the ^{32}P containing DNA had entered the host bacterium whereas bulk of the proteins did not.

Experiments performed in a more refined way showed that almost all the virus DNA enters the infected *E. coli* cell and only about 3 per cent of the protein. Evidently the genetic material of the virus is contained in the DNA rather than in the trace of protein. With conclusions drawn from Hershey and Chase experiments, DNA was indisputably recognised as the hereditary material.

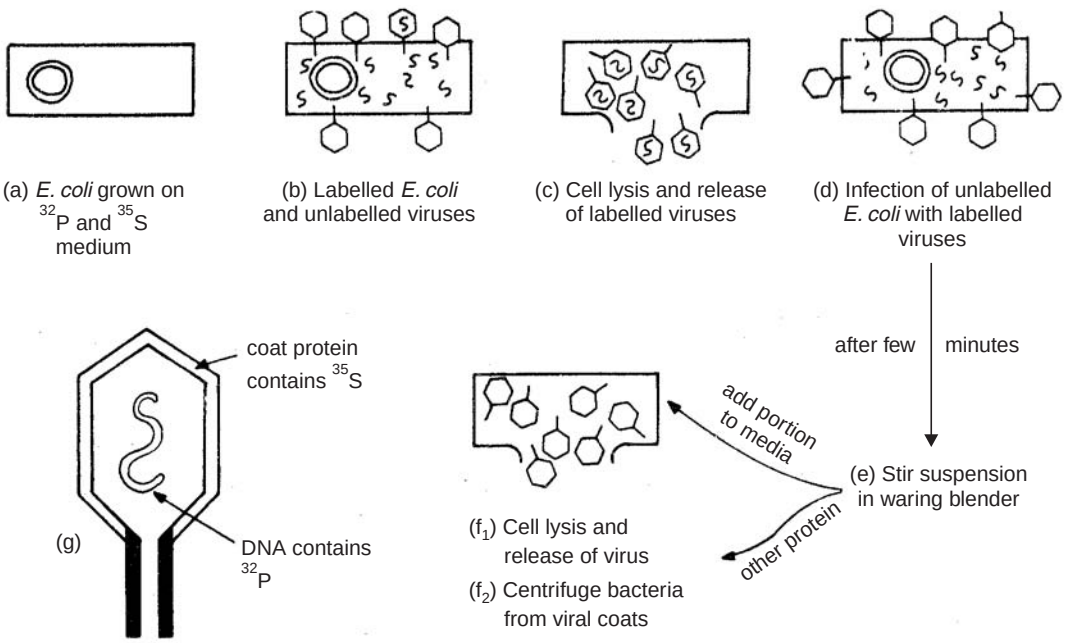


Fig. 13.2 Diagram showing the steps in the Hershey-Chase experiments.

GENES OF SOME VIRUSES ARE MADE OF RNA

Some investigators of 1950s focussed their attention on viruses containing protein and RNA, but no DNA. The poliomyelitis virus, influenza virus and the virus that causes mosaic disease in tobacco plants (TMV) were thus shown to have genes made of RNA.

Earlier in 1935 W. M. Stanley had reported the isolation of a crystalline protein from infected tobacco plants, which possessed the properties of tobacco mosaic virus. Continuing this work in 1950's, H. Fraenkel-Conrat and his colleagues at Berkeley, California found that Stanley's crystals in fact consisted of ribonucleoprotein—a compound of protein and RNA. They designed experiments to find out whether the genes of this virus were made of RNA or protein or both.

They separated the protein and RNA of the virus and inoculated them into healtly tobacco plants. However, to find out if the genes of TMV are made of RNA or protein, Fraenkel-Conrat and Singer carried out further experiments. The virus has a number of genetically different strains which can be distinguished from the type of symptoms they produce in infected leaves, and also from their differing contents of amino acids in proteins. For example, strain HR contains the amino acids histidine and methionine, both of which are absent from the standard TMV strain. They took these two strains and separated their protein and RNA. Then they reformed hybrid viruses by combining the protein of one strain with the RNA of the other strain. The hybrid viruses were made to infect healthy tobacco plants where they replicated to produce new viruses. The types of symptoms produced were also analysed. The new viruses had protein coats identical to the original virus from which RNA had been isolated. In each case the symptoms were identical to those of the strain from which RNA had been isolated to reconstitute the hybrid virus (Fig. 13.3). The same results were obtained with various combinations of nucleic acid from 4 different strains

and of protein from 3 strains. The conclusion was obvious that genes of RNA viruses are made of RNA and not protein.

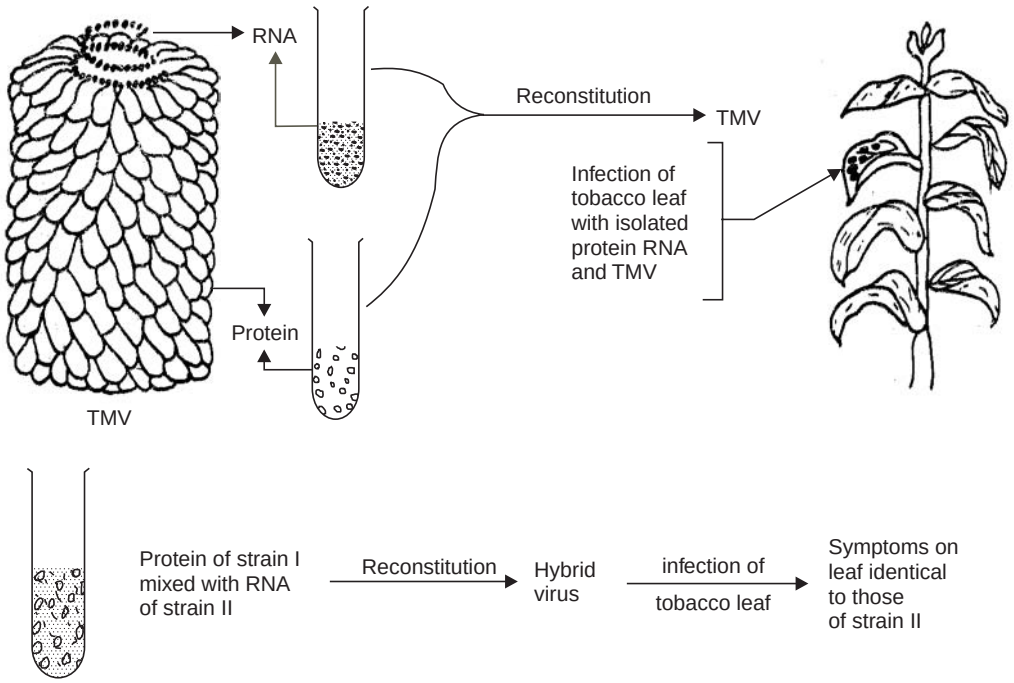


Fig. 13.3 Fraenkel-Conrat's experiment with reconstituted viruses demonstrating RNA as the genetic material.

BIOCHEMICAL EVIDENCE THAT GENES ARE MADE OF DNA

Quantitative estimations of DNA carried out on various organisms revealed the following constant features, providing additional support for DNA as the genetic material.

1. The amount of DNA per cell in an organism or a given species is remarkably constant and cannot be altered by changes in external or internal environment of the cell.

2. The amount of DNA per cell is in proportion to the complexity of the cell and to the amount of genetic information it contains. Thus mammals were found to have approximately 6pg per cell, reptiles 5, birds 2, crustaceans 3, sponges 0.1, higher plants 2.5, fungi 0.02–0.17, bacteria 0.002–0.06 and bacteriophage lambda 0.00008. This shows that higher the organism in the evolutionary scale, greater the content of DNA per cell.

3. Germ cells of higher plants and animals are haploid with only one set of chromosomes, and half the amount of DNA present in somatic cells of the some species.

4. The amount of DNA per diploid cell in a species is constant from one cell type to another.

In 1940s definite clues relating genes to DNA came after the discovery that ultraviolet radiation induces gene mutations. Working with maize plants Stadler found that the wavelength of UV light absorbed maximally by proteins was 280 nm, whereas DNA would absorb maximally at 260 nm. Since the wavelength most suitable for inducing mutations is also 260 nm, it was

inferred that DNA was the target for gene mutations and not proteins. Another confirmation came from studies on the chemical composition of chromosomes. DNA was found to be present exclusively in the nucleus, and absent from locations where chromosomes were not detectable.

Biochemical analysis of the base composition of DNA and its relation with the species members provided unequivocal evidence in favour of DNA as the hereditary material. In the years 1949 to 1953 E. Chargaff and his associates applied quantitative chromatographic methods for the separation and analysis of the four bases of DNA isolated from different species. They could draw the following conclusions: (*a*) The base composition of DNA varies from one species to another; (*b*) different tissues of the same organism have the same base composition in their DNA; (*c*) base composition of DNA in a given species does not alter with age or environment; (*d*) the number of adenine residues is equal to the number of thymine residues ($A = 1$) and the number of guanine residues is equal to the number of cytosine residues ($G = C$). It follows therefore, that the sum of purine residues equals the sum of pyrimidine residues ($A + G = C + T$); (*e*) the DNA specimens from closely related species have similar base composition, and of unrelated species have widely different base composition. Base composition of DNA can be used to classify organisms. An infinite number of different sequences of base pairs are possible in a DNA molecule, thus DNA is capable of encoding an enormous amount of information, and is more likely to be the genetic material of most organisms.

QUESTIONS

1. Why did Hershey and Chase use labelled sulphur and phosphorus in their experiments in preference to any other radioactive element?
2. (*a*) State briefly how the transforming principle in Griffith's experiment was identified, (*b*) Can human cells also undergo transformation? If so, give an example.
3. Describe an experiment which proves that the genetic material in some viruses is not DNA.
4. Give evidence that the genes of higher organisms are made of DNA.

SELECTED READINGS

Avery, O. T. *et al.*, 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. *J. Exp. Med.* **79**:137.

Chargaff, E. 1950. Chemical Specificity of Nucleic Acids and the Mechanism of their Enzymatic Degradation. *Experientia* **6**:201.

Fraenkel-Conrat, H. and Williams, R.C. 1955. Reconstitution of Tobacco Mosaic Virus from its Inactive Protein and Nucleic Acid Components. *Proc. Natl. A cad. Sci. U.S.* **41**:690.

Fraenkel-Conrat, H. and Singer, B. 1957. Virus Reconstitution. II. Combination of Protein and Nucleic Acid from Different Strains. *Biochem. Biophys. Acta* **24**:540.

Griffith. F. 1928. The Significance of Pneumococcal Types. *J. Hygiene* **27**:113.

Hershey, A.D. and Chase. Martha 1952. Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage. *J. Gen. Psysiol.* **36** :39.

Miescher, F. 1871. On the Chemical Composition of Pus Cells. *Hoppe-Seyler's Med. Chem. Untersuch.* **4**:441.

14

The Structure and Replication of DNA

Studies initiated in the 1940's indicated that DNA was the universal genetic material of all living organisms except certain viruses which had RNA as the genetic material. Soon after this was established, enthusiasm shifted toward understanding the structure of DNA. Many investigators tried various approaches, but the major breakthrough came in 1953 when James D. Watson and Francis Crick presented a model for DNA. Before going into the details of this model it is necessary to make a brief mention of some other discoveries which proved useful in creating this model.

Between the years 1949 and 1953 E. Chargaff and his colleagues at Columbia analysed the composition of DNA from various sources by refined quantitative chromatographic methods. They demonstrated base equivalence in DNA—the number of adenine bases was equal to thymine ($A = T$) and guanine was equal to cytosine ($G = C$). It follows that $A + G = C + T$. They also found out that $A + T/G + C$ ratio is specific for a given species. The significance of this finding became clear when the structure of DNA was finally elucidated by Watson and Crick.

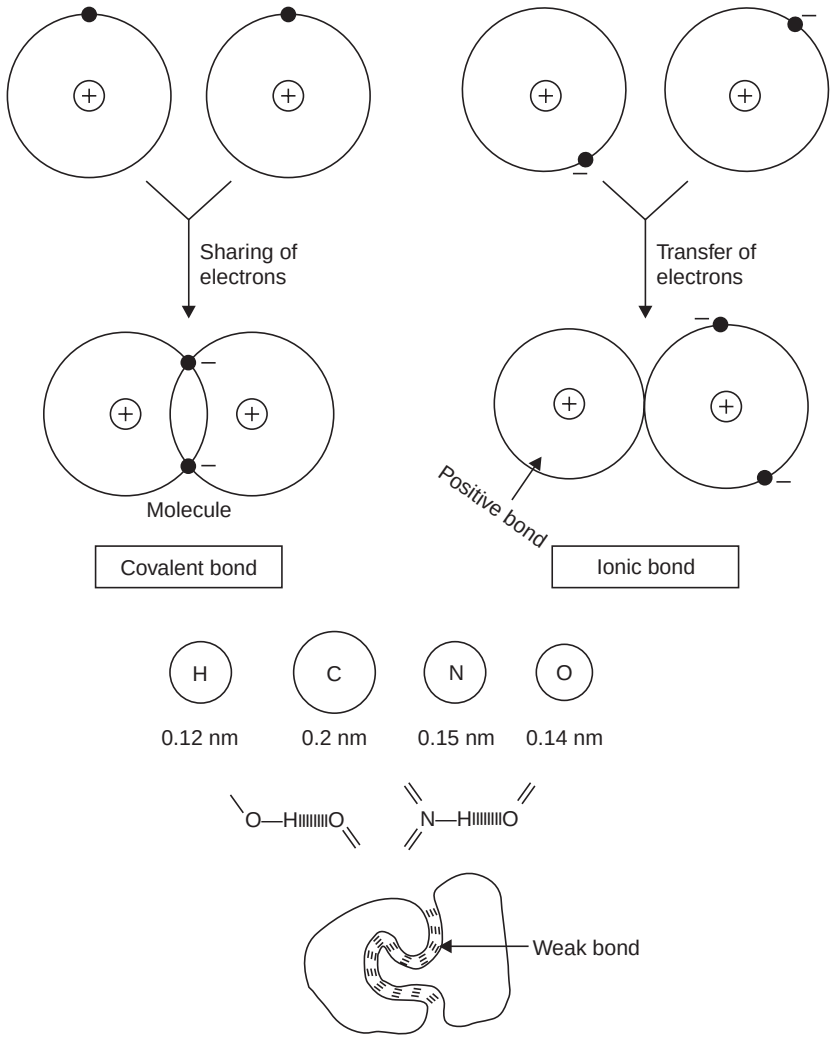
In 1951 Linus Pauling and Corey discovered the α-helical structure of fibrous proteins by X-ray diffraction technique. The α-helix is a spiral configuration maintained by *intramolecular* hydrogen bonds. In contrast β-configuration is an elongated configuration in some proteins stabilised by *intermolecular* bonds involving different protein chains. Later on, Pauling and Corey extended these ideas to the spatial geometry of the DNA molecule. They suggested three helical nucleotide chains for DNA with phosphate groups oriented inside the structure and the purine and pyrimidine groups projecting outwards. However, contemporary biochemists engaged in research on DNA structure held the view that hydrogen bonding between the bases probably played a role in providing stability to the molecular architecture of DNA. If this was so, then the bases in the two helical chains should be pointing inwards, towards each other, in contrast to the suggestion of Pauling and Corey.

In 1938 W. Astbury and F. O. Belly probed DNA structure by subjecting DNA fibres to X-ray diffraction analysis. But their samples were probably impure. During 1950 to 1953 more refined X-ray data were obtained on fibres of highly purified DNA by R. Franklin and M. H. F. Wilkins. Depending upon the degree of hydration, DNA exists in two forms, *A* and *B*. The *B* form is biologically important and shows purine and pyrimidine bases placed regularly along the molecule at a distance of 3.4 Å. They also found that another regularity in structure occurred at a repeating distance of 34 Å, that is after every 10 nucleotides. They concluded that the DNA molecule was not linear but twisted into a helix, one complete turn occurring every 34 Å.

TYPES OF BONDS

Covalent and Ionic Bonds

Atoms can attain a more stable arrangement of electrons in their outermost shell by interacting with one another. An ionic bond results from transfer of electrons from one atom to the other. A covalent bond is formed when electrons are shared between atoms. Often covalent bonds form with a partial transfer that leads to unequal sharing of electrons, resulting in a polar covalent bond.



Weak non-covalent bonds.

Van der Waals Forces: At very short distances any two atoms show a weak bonding interaction due to their fluctuating electrical charges. If the two atoms are too close together, however, they repel each other very strongly.

Each atom has a characteristic Van der Waals radius. The contact distance between any two atoms is the sum of their Van der Waals radii.

Two atoms will be attracted to each other by Van der Waals forces until the distance between them equals the sum of their Van der Waals radii. Although they are individually very weak, Van der Waals attractions can become important when two macromolecular surfaces fit very close together.

Hydrogen Bonds: The hydrogen bonds form when a hydrogen atom is "sandwiched" between two electron-attracting atoms (usually oxygen or nitrogen). Hydrogen bonds are strongest when the three atoms are in a straight line.

O --- H ||||||||O N ----- H ||||||||O

Examples of hydrogen bonds are seen in DNA and RNA. There are three hydrogen bonds between *G* and *C* pairs, two hydrogen bonds between *A* and *T* pairs. Amino acids in polypeptide chains are also hydrogen-bonded together. Molecules that can form hydrogen bonds with each other can also form hydrogen bonds with water molecules. Because of this competition with water molecules, the hydrogen bonds formed between two molecules dissolved in water are relatively weak.

Weak Chemical Bonds: Organic molecules can interact with other molecules through short-range noncovalent forces. Weak chemical bonds have less than half the strength of a strong covalent bond. However, they are strong enough to provide tight binding only when many of them are formed simultaneously.

THE DOUBLE HELIX

On the basis of the evidences available and a knowledge of interatomic distances and bond angles, Watson and Crick proceeded to construct molecular models of DNA. In 1953 they finally postulated a precise, three-dimensional, double helical model which could account for many of the observations on the chemical and physical properties of DNA (Fig. 14.2). The model also suggested a mechanism by which the genetic material could be accurately replicated. In 1962 Watson, Crick and Wilkins were awarded Nobel Prize for this work. In genetic discussion, one strand of the double helix is conveniently referred to as Watson (W), the other strand as Crick (C).

The basic *building blocks* of DNA are shown in Fig. 14.2. The combination of a base and a sugar is called a *nucleoside*. The first carbon atom of the sugar is linked to the nitrogen in 9th position in a purine base, and to the first nitrogen in a pyrimidine. The phosphate ester of a nucleoside is called *nucleotide*. The phosphoric acid (H_3PO_4) forms an ester linkage with one of the free hydroxyl groups in the pentose sugar which is a deoxyribose. The polynucleotide chain is made up of nucleotide units held together by phosphate diester linkages. One molecule of a phosphate joins two nucleoside units by forming

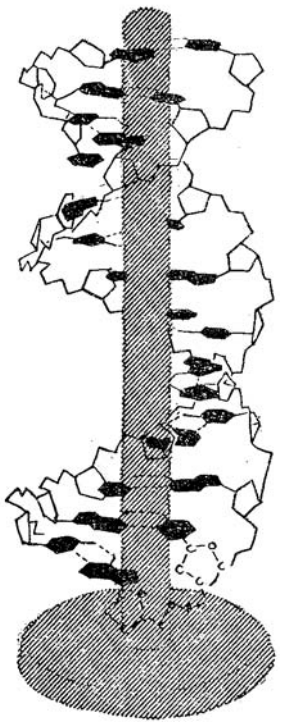


Fig. 14.1 A view of the double helix.

an ester linkage with the hydroxyl on the 3′ carbon atom of one sugar molecule and another ester linkage with the 5′ hydroxyl of the sugar of the adjacent nucleotide. All this information was available in the 1940's. However, the molecular arrangements of the component and the three-dimensional spatial geometry of DNA was not known.

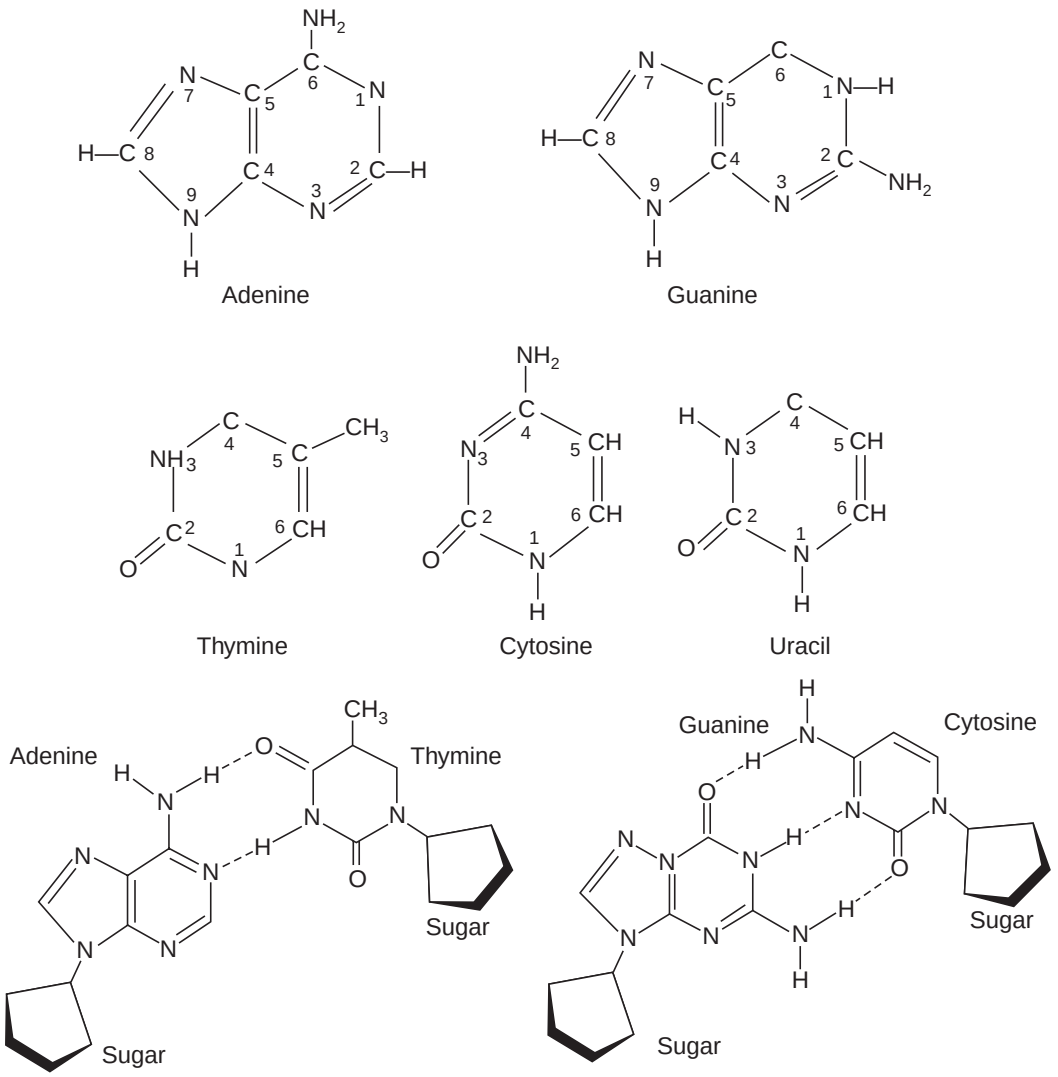


Fig. 14.2 Base pairing in DNA.

Watson and Crick's model consists of a double, right-handed helix in which polynucleotide chains are helically coiled about the same axis. The coiling of the two chains is such that they cannot be separated except by unwinding the coils; this is called *plectonemic* coiling. The sugar units on adjacent nucleotides are linked by phosphate groups to form an outer sugar-phosphate backbone. The only OH groups available for ester linkages on the pentose sugar are those on the 3′ and 5′ carbon atoms. Thus each phosphate links the 3′-carbon atom on one sugar to the 5′-carbon on the sugar of the next nucleotide (Fig. 14.3).

The purine (adenine and guanine) and pyrimidine (thymine and cytosine) bases are turned inwards and linked by hydrogen bonds, each base on one chain being paired with a base on the other chain. Base pairing is specific so that adenine pairs with thymine and guanine with cytosine. Thus the two polynucleotide chains have complementary base sequences. This means that if one chain has base sequence reading *ACCGATC* ... , then the *corresponding* region of the complementary chain will have *TGGCTAG* ..... Two hydrogen bonds form between adenine and thymine and three between guanine and cytosine. The bases are relatively hydrophobic molecules. Since DNA is normally in water solution, the long and flat base pairs remain turned on the inside of the double strand, leaving the sugar-phosphate backbone on the outside.

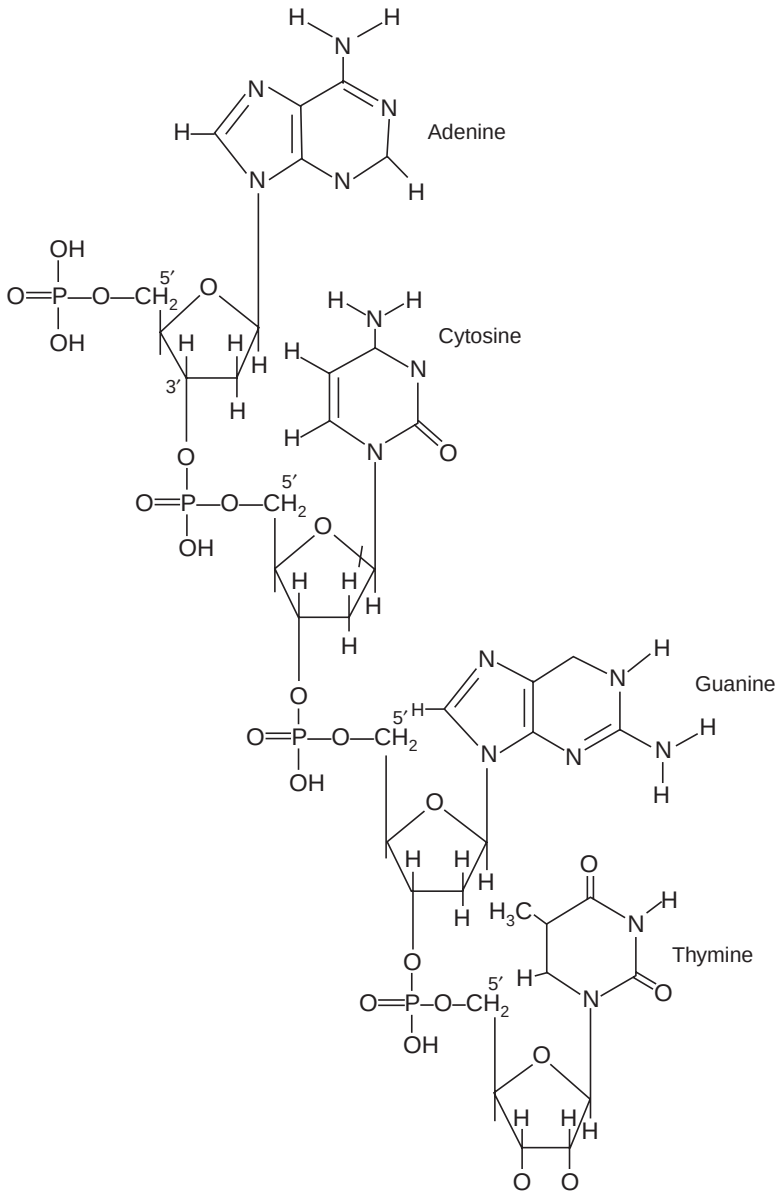


Fig. 14.3 Linkage of nucleotide units to form a single polynucleotide chain.

Base pairing also means that the phosphate sugar linkages run in opposite directions on the two chains. The phosphate sugar links go from a 3′-carbon to a 5′ carbon in one chain. In the complementary chain the phosphate-sugar linkages must go from a 5′-carbon to a 3′-carbon.

We say that one chain runs in 3′ to 5′ direction, the other chain in 5′ to 3′ direction. That is why the sugar molecules in one chain have the *O* position facing upwards, in the other chain ring *O* faces downwards. That is, the two chains are antiparallel.

Because of the opposite arrangement of the sugars in the two chains, and because the sugar binds to a position away from the centre of the base, the whole DNA molecule is constrained to coil or twist, forming a double helix. Each successive base pair in the stack turns 36° in the clockwise direction. The double helix thus makes a complete turn of 360° after 10 base pairs. The diameter of the helix is 20 Å. The chain makes a complete turn every 34 Å. Since there are 10 nucleotides in each chain in every turn of the helix, the distance of a single nucleotide base is 3.4 Å.

The base pairs do not uniformly fill up the space between the backbone of one chain and the backbone of the other. The space below a base pair is filled more than the space around the top. Consequently an empty *major groove* is formed and an empty *minor groove*. Because of the presence of numerous phosphate groups, the double helix is strongly acidic and carries a high density of negative charges. DNA is associated with cations and also with basic proteins which are considered to be present in the major and minor grooves.

THE A, B AND C FORMS OF DNA

The more refined X-ray diffraction studies of Wilkins and his colleagues of the 1950's, as well as circular dichroism and some other techniques of the 1970's have shown that DNA fibres can have 3 possible structures *A*, *B* and *C*. The observed structure depends upon conditions like base composition, salt content and relative humidity of the sample. The essential differences lie in the number of bases per helical turn, the inclination of bases to the axis of the helix and the pitch of the helix.

The *B* structure is found in DNA fibres of 92% relative humidity. They have a pitch of 3.4 nm and 10 bases per turn. The *B* form is closest to Watson and Crick's model. The *A* structure is shown by DNA fibres of 75% relative humidity in a solution of sodium, potassium or caesium. It differs from *B* form of DNA in that the bases are not flat but tilted, the pitch of the helix is 2.8 nm, and there are 11 bases per turn. The backbone structure is also slightly different. The *A* form of DNA is of interest as it is believed to resemble the structure of double stranded RNA and of DNA-RNA hybrids in solution. The *C* form of DNA is found in fibres of lithium DNA at 66% relative humidity. It has a pitch of 3.1 nm and 9.3 bases per turn.

The stability of the helix was earlier believed to be provided by the hydrogen bonds only. This idea is now modified and it is considered that the base stacking forces or 'apolar' bonds maintain stability of the double helix. The hydrogen bonds confer additional stability and specificity.

THE LEFT HANDED HELIX

The Watson-Crick model of DNA has been the basis for further work on DNA replication, mutation and recombination. Much evidence exists to establish the validity of this model. Nevertheless, the

model has been scrutinised by other workers, and from time to time other models have been suggested by Wu(1970), Arnott *et al.* (1968), Wang (1979) and a few others. Wang's model is based on artificially synthesised small length of DNA of known sequence and is described here.

Wang *et al.* (1979) crystallised DNA fragments of fixed sequence and analysed their structure by X-ray diffraction. They used a double helical fragment of DNA containing 6 base pairs in alternating sequence *CGCGCG*. The double helix was revealed to be left handed due to alternating conformations of the *C* and *G* residues. It has some new features not present in the Watson-Crick right handed helix of the *B* form of DNA. The fragment has antiparallel strands with Watson-Crick base pairing *A*-*T* and *G*-*C*. But the two strands have a left-handed helical sense. The sugar phosphate backbone has a zig-zag course. The sugar in molecules attached to the guanine bases have the ring *O* pointing upwards, while those attached to cytosine have ring *O* pointing downwards. Because of this alternating conformation of the sugar molecules, the repeat unit in the hexamer fragment is a *dinucleotide*, not a mononucleotide as in the Watson-Crick model. Furthermore, the helix contains 12 base pairs per turn. Each pair in the complementary chains is related to the next by a rotation of 60°. The bases themselves are tilted 7° from the helix axis.

There are two general positions which a base can occupy relative to the sugar residue called *anti* and *syn*. In the left handed helix the cytosine residues are in the *anti* conformation as in *B* DNA, but the guanines are in the *syn* conformation. The zig-zag course followed by the sugar phosphate backbone is due to the alternating conformations of the guanine and cytosine residues. Wang *et al*., have called this *Z*-DNA.

The bases are not stacked directly above one another, but are displaced laterally by 7Å. The guanines moreover are stacked upon the oxygen atom of the ribose sugar. The guanine residues in *Z*-DNA are located away from the centre; in *B* DNA they are closer to the centre.

Z-DNA looks like a cylinder and has less of a grooved appearance than *B* DNA. The minor groove is deep in *Z*-DNA and forms below the base pairs, in *B* DNA it is above the base pairs. There is no prominent major groove in *Z*-DNA as in the Watson-Crick model. Instead there is a convex surface above the base pairs.

Z-DNA has a smaller diameter of 18 Å and one complete turn of the helix has 12 base pairs or 6 dinucleotide pairs. A length of 44. 6° is covered by these 12 pairs; in *B* DNA the same length is occupied by 13 base pairs.

BENDS IN DNA

There is evidence from X-ray diffraction studies and from pattern of nuclease digestion (cuts produced by DNAase I and II and micrococcal nuclease) that a structure in DNA repeats itself after 10 or 20 base pairs. Crick and Klug (1975) suggested that at these points the double helix is bent or 'kinked'. Between the kinks the DNA is straight. One base pair at the kink is unstacked from the adjacent one while pairing of the strands is undisturbed. This produces a bend in the double helix towards the side of the minor groove. According to Selsing *et al*. (1979) there might be alternate stretches of *A* and *B* forms of DNA between the bends.

SUPERCOILED DNA

In Chapter 19 the supercoiling of DNA complexed with protein in eukaryotic chromatin is described (the solenoidal superhelix of chromosomes). Here, a different type of supercoiling in some forms of prokaryotic and organelle DNA is discussed. In the closed circular molecules of DNA, such as in mouse polyoma virus, human papilloma virus and mitochondrial DNA, the double helix itself becomes coiled into a new helical form called supercoiled DNA. It is also observed that covalently closed circular molecules of double helical DNA are usually underwound. They are said to be negatively supercoiled and in a relaxed state. The forces which confer stability on the double helix probably lead to the formation of supercoils in underwound DNA (Bauer, Crick and White, 1980).

A recently described enzyme DNA gyrase is said to maintain the negatively supercoiled conformation of DNA (Kano *et al.,* 1981). *In vitro* the enzyme is shown in convert underwound, closed circular DNA into a supercoiled form in the presence of ATP. It has also been suggested that DNA gyrase is involved in replication, transcription and recombinational activities of DNA.

The enzyme DNA gyrase consists of two subunits, the nal *A* protein and the cou protein; the first (nal *A*) subunit performs a nicking-closing function resulting in relaxation of supercoiled DNA. This activity is inhibited by nalidixic acid and oxolinic acid. The activity of the cou subunit though not well understood is inhibited by novobiocin.

Besides DNA gyrase, a number of enzymes which perform similar activities have been recently detected. Since they all produce interconversions in the topological isomers of DNA, Wang and Liu (1978) have called them DNA topoisomerases.

REPLICATION OF DNA

While presenting the double helical model of DNA, Watson and Crick had remarked, "It has not escaped our notice that the specific pairing (of bases) we have postulated suggests a possible copying mechanism for the genetic material." The complementary nature of the strands suggests that the process of replication must begin with the separation of strands. This idea was proved correct from experiments in which the distribution of labelled and unlabelled DNA strands in the progeny were studied after successive rounds of DNA replication.

Each strand of DNA sythesises a new strand, so that the resulting DNA has one parental and one new strand. This is called *semi-conservative* replication because only half of the parental molecule is conserved during replication. Two other modes of DNA replication were suggested by Delbruck (Fig. 14.4). *Conservative* method in which each of the two strands of the parent DNA is replicated resulting in two DNA molecules—one double helix with *both* parental strands, the other with *both* new strands. In *dispersive* replication the double helix breaks at several points forming many pieces. Each piece replicates and the pieces become reconnected at random. Thus two new helices of hybrid molecules are formed in which the parental DNA is distributed all over the four strands.

The final proof that DNA replication indeed takes place semiconservatively came from experiments performed by Meselson and Stahl. They used a novel technique for determining the density of purified DNA by ultracentrifugation. By this method they could separate *E. coli* DNA of different densities. DNA containing the heavy isotope ^{15}N is denser than DNA containing normal nitrogen ^{14}N. *E. coli* cells were grown in ^{15}N-labelled medium for several generations, so

that their chromosomes contained all ^{15}N-labelled DNA. A sample of fully labelled *E. coli* cells was drawn out and made to grow on a medium containing normal nitrogen. After a single replication (one cell generation time), a cell sample was taken out, its DNA was purified and analysed as described below. This was repeated for two successive generations.

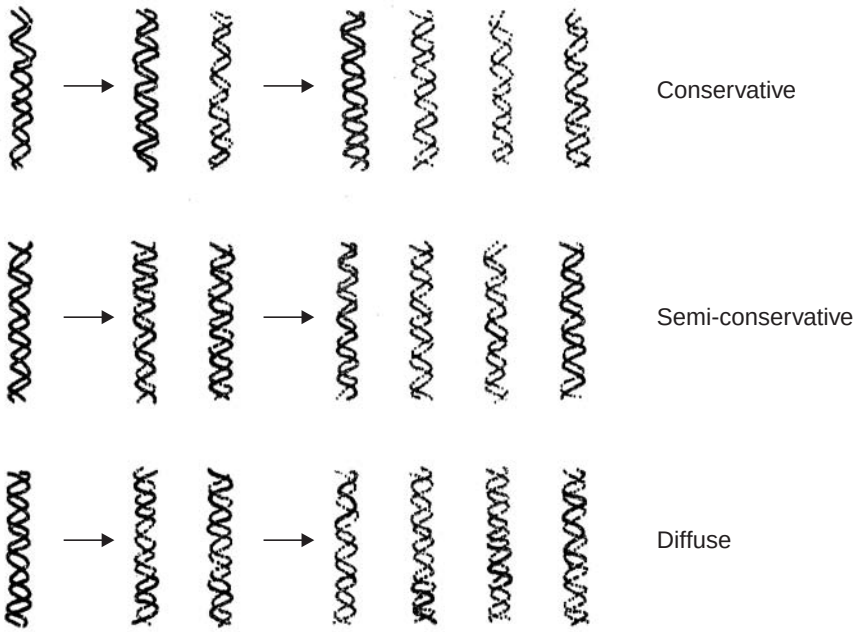


Fig. 14.4 Diagrams illustrating the three modes of DNA replication.

The method of separating heavy DNA from light DNA is called *density gradient equilibrium centrifugation.* Purified DNA is dissolved in a concentrated solution of caesium chloride (8.8 M CsCl in water) and centrifuged in a transparent quartz tube for 5–8 hours at very high speed. When centrifugal forces are applied, CsCl being denser than water tends to sediment towards the outside of the tube. This tendency to sediment is opposed by diffusion, and after several hours a state of equilibrium is reached. The concentration distribution of CsCl in the tube is stabilised with high density at the bottom of the tube and lowest at the top, and all grades of density in between. The DNA molecules in the tube come to rest in that region of the CsCl gradient that corresponds to their own density, and form a sharp band. The exact position of the DNA band in the density gradient is determined by using ultraviolet light to photograph the tube, since DNA absorbs ultraviolet light and CsCl does not. From the position of the DNA band the density of the DNA molecules present in the band can be calculated.

It was found that heavy DNA fully labelled with ^{15}N formed a band at the bottom of the CsCl solution. The DNA isolated from cells grown in low density (^{14}N) medium for one generation formed a band in an intermediate position. Whereas light DNA from low density medium after two cell generation's time formed a band at the top (Fig. 14.5). This shows that after one complete round of DNA synthesis, all the new DNA molecules are hybrids, consisting of an old heavy (^{15}N) DNA strand and one new light (^{14}N) DNA strand. The data provided evidence that DNA had replicated semi-conservatively.

Experiments similar to Meselson and Stahl's were performed on eukaryotic cells in which bromodeoxyuridine (BudR), a specific density label for DNA was used by Filner (1965) in *Nicotiana tabacum* and by Sueoka (1961) in *Chlamydomonas*. BudR being an analogue of thymidine becomes incorporated into DNA by replacing thymidine. The results showed that after one cycle of DNA replication in BudR, all the resulting DNA is of hybrid density. Thus eukaryotic DNA replication is also semi-conservative.

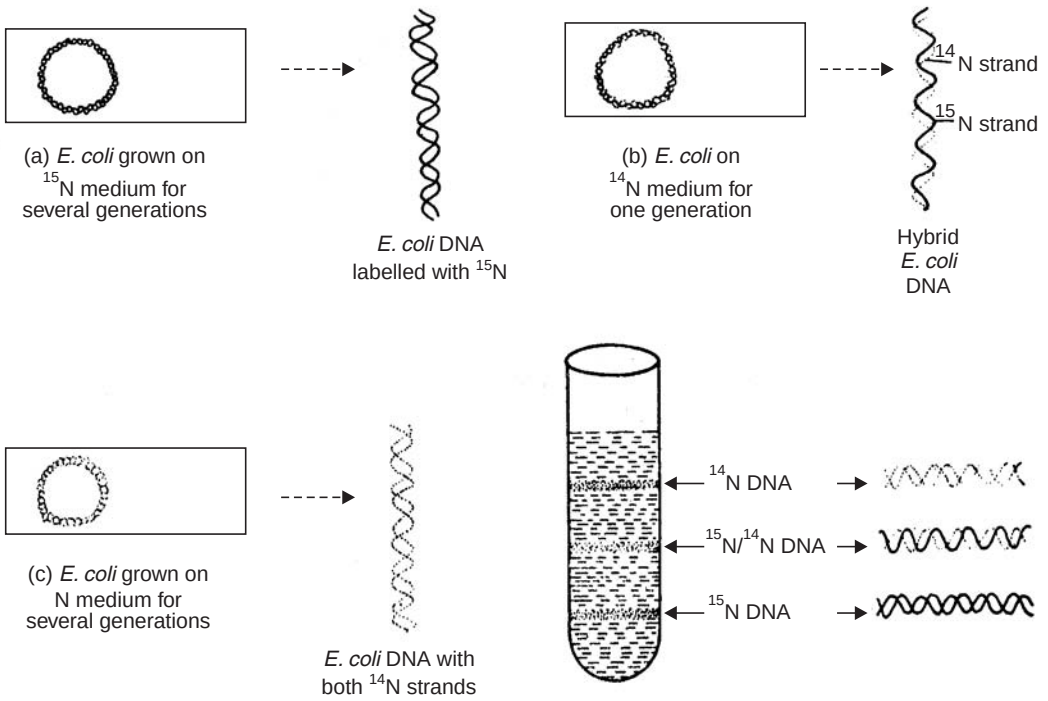


Fig. 14.5 Diagrammatic representation of the experiment of Meselson and Stahl.

Using the technique of autoradiography, J.H. Taylor (1957) studied the pattern of segregation of whole chromatids in chromosomes at successive mitoses after labelling. The results have been interpreted to demonstrate semi-conservative replication in eukaryotic chromosomes.

Root tips of the broad been (*Vicia faba*) were grown in a solution containing tritiated thymidine (*i.e.* thymidine labelled with ^{3}H called tritium, a radioactive isotope of hydrogen). The actively dividing cells in the root tip incorporated radioactive thymidine into the DNA in chromosome. The cells were allowed to divide for many generations in this medium, after which they were transferred to a medium containing normal thymidine. Here the cells were allowed to remain just long enough to undergo *one* cell division. The presence of radioactive isotope in chromatids was detected by autoradiography as described below.

Chromosome spreads are prepared on a slide which is then covered with a liquid emulsion of silver bromide. The slides are then stored in the dark for 6–8 weeks. During this time the decay of the radioactive isotope will reduce the silver bromide grains lying directly above them. The slides are then developed much like a photograph negative. The emulsion is washed out except for dark spots at places where reduced silver grains are present.

The results were clear and conclusive. Cells harvested after several rounds of replication in solution containing radioactive thymidine had both chromatids labelled (Fig. 14.6). But in cells which had divided once in normal thymidine, only one chromatid of each chromosome was labelled, the other chromatid was unlabelled. At each successive generation, the number of labelled chromatids in daughter cells was halved. The segregation of chromatids is therefore semi-conservative.

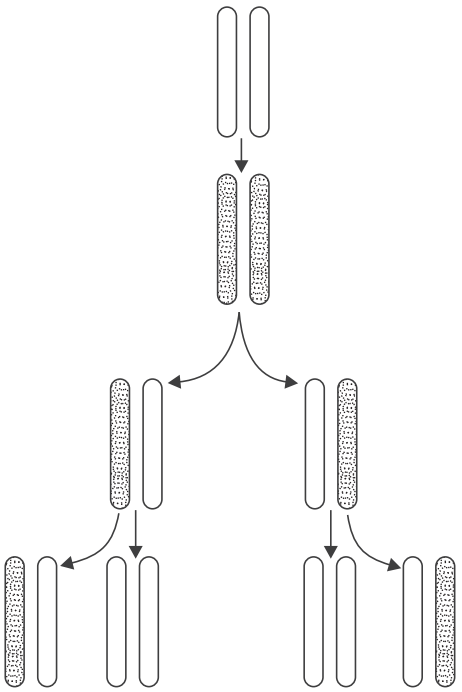


Fig. 14.6 Taylor's experiment illustrating semi-conservative segregation of chromatids.

THE NUCLEIC ACID ENZYMES — A General Account

Before discussing the molecular mechanism underlying replication, it is necessary to have some background knowledge about enzymes associated with nucleic acids. Except for the polymerases, most of the enzymes mentioned below were discovered during the 70s and 80s. Some of them play a key role in replication.

The Nucleases: Enzymes which catalyse the breakdown of particular bonds leading to fragmentation of nucleic acids are called *nucleases.* The enzyme DNase causes hydrolysis (addition of water) of the phosphodiester bonds that link nucleotides in a DNA molecule. Similarly RNase acts on the polynucleotide RNA.

The exonucleases are enzymes that attack a nucleic acid at its terminal nucleotide only. In the 3′ or the 5′ phosphodiester bonds an exonuclease will act on either the 3′ or the 5′ end of the linkage.

The endonucleases react only with those bonds which occur in the interior of a chain. They can cut a single polynucleotide chain into pieces. In the case of double stranded DNA, they can

produce nicks (gaps) in a single strand. Due to this the helix remains intact and the free ends at the gaps serve as substrates for exonucleases. There are several types of nucleases known which can degrade a nucleic acid in different specific ways.

The Ligases: These are enzymes which can join broken ends of two DNA chains by catalysing the synthesis of a phosphodiester bond between a 3′-hydroxyl group at the end of one chain and a 5′-phosphate group at the end of the other chain.

Restriction Enzymes: In the 1970s a new class of endonucleases has been isolated from micro-organisms whose action is limited only to specified nucleotide sequences. They are known as *restriction* enzymes and will produce breaks only within sequences which have two identical bases in adjacent positions such as *CCT*, *TTA*, *GGC*, *AAG*. Since at these positions the complementary strand also has identical bases (*GGA*, *AAT*, *CCG*, *TTC*), restriction enzymes act on both strands and produce a break. A number of different restriction enzymes have been isolated (such as Hind II, Hind III, Eco R 1 and others) which specifically break particular nucleotides. They are thus useful for breaking a DNA molecule into specific known type of smaller fragments.

The Polymerases: Enzymes such as DNA polymerase and RNA polymerase are involved in the synthesis of nucleic acids by addition of bases to a growing nucleotide chain. DNA polymerase can also perform nuclease activity by removing bases from a DNA strand, but in the opposite direction (3′ 5′) to that in which it adds bases (5′ 3′). The details of polymerase function are described later.

The Swivelases: The two helical strands of DNA are wound around each other as if two parallel strands had been coiled around a central axis. This is called plectonemic coiling and is revealed by X-ray diffraction data. This winding does not allow rotation or movement of helices about each other. DNA replication requires rotation at specific sites called swivels within the helical strands. Swivels allow free rotation of one part of the molecule relative to the adjacent part. In this way only the DNA between the growth point and the nearest swivel would have to rotate during replication. Enzymes that produce transient swivels are known as *swivelases.* They have endonuclease action on one of the strands thus allowing free rotation within the DNA molecule. The site is then immediately scaled leaving an intact strand for replication.

Unwinding Enzymes and Proteins: The unwinding of DNA during replication requires an endonuclease enzyme that makes a single break in a strand. This allows DNA to unwind in that region instead of the whole DNA molecule having to twist. The further opening up of local regions is aided by specific *unwinding proteins* which bind preferentially to single-stranded DNA, resulting in separation of strands. The binding of one unwinding protein to a DNA strand promotes the binding of a second, so that the helix continues to unwind. Another class of proteins called strand separation proteins (SSP) have also been isolated from both prokaryotes and eukaryotes.

MOLECULAR MECHANISM OF DNA REPLICATION

By late 1950s it became generally accepted that DNA replicates semiconservatively. Studies on molecular mechanisms underlying replication were initiated only after Arthur Kornberg in 1957 discovered the enzyme which is involved in DNA replication. This enzyme is today known as DNA polymerase I or pol I. Kornberg was awarded Nobel Prize in 1959. Later on two other

enzymes DNA polymerase II and DNA polymerase III (pol II and pol III) having similar properties as pol I were isolated. It appears that mainly DNA polymerase III is involved in DNA replication. DNA polymerase I fills gaps in DNA and also functions in repair of DNA. The function of DNA polymerase II is still not well understood.

DNA Polymerase I: Kornberg isolated this enzyme from living *E. coli* cells and used it for DNA replication *in vitro*. Since *E. coli* cells divide rapidly, their DNA duplicating once in every 20 minutes, it is a suitable material for isolating the enzyme in high concentration. Kornberg achieved *in vitro* DNA synthesis using DNA polymerase I and the following components: (1) all the four deoxynucleotides in the triphosphorylated form *i.e. dATP*, *dGTP*, *dCTP*, *dTTP*; (2) Mg^{++}; (3) a primer strand of DNA (or RNA) with a free 3′ OH end (primer DNA is a partially denatured double stranded DNA to which new bases are added and also serves as template for the sequence of bases added). The DNA polymerase I catalyses the addition of single nucleotide units to the free 3′-hydroxyl end of the primer DNA strand. The chain is synthesised in the 5′–3′ direction. The primer DNA is itself a short strand of DNA or a double strand with a nick in one strand exposing a free 3′-hydroxyl group and a 5′-phosphate group. The primer strand is base-paired with a long DNA strand that serves as the *template*. The name *template-primer* is now given to this structure.

DNA polymerase I isolated from *E. coli* has a single polypeptide chain of about 4000 amino acids and a molecular weight of 109,000. The enzyme molecule is roughly spherical in shape with a diameter of about 6.5 nm. In the purified state one molecule of the enzyme, at 37°C can add about 1,000 nucleotides per minute to a growing DNA chain. To show that the DNA synthesised *in vitro* by the purified DNA polymerase I has the same base sequence as the DNA used in the primer and template strand, Kornberg devised a method known as *nearest-neighbour base frequency analysis.* This method determines the frequency with which any two bases occur as adjacent or nearest neighbours in the DNA strand.

The *in vitro* system containing all the four triphosphorylated nucleotides has one of the nucleotides labelled with ^{32}P. Out of the three phosphates in the triphosphate linked to 5′ carbon of the sugar, the innermost or alpha phosphate group has the ^{32}P label. This phosphate is not eliminated as pyrophosphate in the reaction. But when DNA synthesis occurs, it forms second linkage with the carbon at position 3′ of the neighbouring nucleotide in the newly synthesised DNA chain. The DNA is then degraded by the endonuclease enzyme spleen phosphodiesterase which breaks DNA into single nucleotides at the 5′ phosphate-sugar linkages only. The ^{32}P containing alpha phosphoric acid remains attached to the 3′ carbon of the nearest neighbour nucleotide to the one on which it entered the DNA strand. The ^{32}P contents of all the four nucleotides are measured and found proportional to the frequencies with which they occurred next to the nucleotide that was labelled. The experiment is repeated three more times each time labelling the triphosphate of a different nucleotide.

Kornberg's experiments established beyond doubt that DNA polymerase I lead to *in vitro* DNA synthesis. However, in 1967 J. Cairns provided evidence that this may not always be so. He isolated a mutant strain of *E. coli* known as pol *AI* which was lacking in DNA polymerase, but still was able to grow and replicate its DNA normally. When these cells were exposed to UV and X-radiation, it was found that they could not repair damage to their DNA. From this it follows that DNA polymerase I is a repair enzyme and may not be responsible for replication.

Later on DNA polymerase I was shown to have exonuclease activity in both 3′ to 5′ and 5′ to 3′ directions (Fig. 14.7 A). It performs exonuclease activity mainly for removing a wrongly

placed nucleotide, as will be explained later. Further experiments proved that the DNA polymerase I molecule actually consists of two subunits: a larger polypeptide chain with molecular weight of about 76,000 daltons; a smaller chain with molecular weight of 36,000 daltons. The two subunits dissociate on treatment with proteolytic enzymes. The larger subunit possesses the repair and 3′ to 5′ exonuclease activities; the smaller subunit has 5′ to 3′ exonuclease activity.

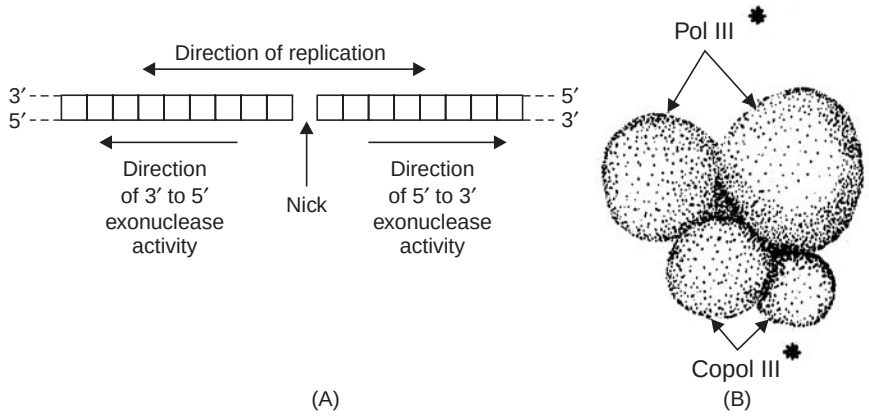


Fig. 14.7A. The exonuclease activities of DNA polymerase I.
B. DNA polymerase III.

DNA Polymerase II and III

From the pol AI mutants of *E. coli* two other enzymes were isloated namely DNA polymerase II and DNA polymerase III (pol II and pol III). The wild type *E. coli* cells contain all the three polymerase enzymes *i.e.* pol I, II and III which are coded for by structural genes pol *A*, *B*, and *C*. The exact function of DNA polymerase II in replication is not known with certainty. It has exonuclease activity in 3′ to 5′ direction. Mutants defective in pol II (pol *B*) have been isolated but they are phenotypically indistinguishable from the wild type. It is, therefore, likely that pol II is not necessary for replication.

DNA polymerase III was discovered in 1972 by T. Kornberg (son of Arthur Kornberg) and Gefter. This is the enzyme most actively involved in DNA replication. It also performs exonuclease activity in both 3′ to 5′ and 5′ to 3′ directions. Later on Kornberg showed that this enzyme is actually a dimer of pol III and pol III* and is known as DNA polymerase III* (pol three star). It requires a protein, copolymerase III* (copol III*) to function (Fig. 14.7B). Like pol I, DNA polymerase III* replicates a template-primer strand of DNA. It appears therefore that pol III*-copol III* complex along with ATP, a DNA template and an RNA primer are necessary to initiate replication.

DNA REPLICATION IN BACTERIA

The work of J. Cairns demonstrated DNA replication in bacteria in an elegant way. He used high resolution autoradiography for observing the distribution of radioactive label in the "chromosomes" of dividing *E. coli* cells.

The bacterial cells were grown on a medium containing tritiated thymidine. The labelled DNA was extracted from the cells and spread on EM grids. A very thin layer of the photographic emulsion was laid on the grids which were stored in the dark. Whatever radioactive thymidine

was present in the DNA, it reduced the grains of silver halide located directly above it. When the grids were developed and examined in the electron microscope, a track of reduce silver grains (black dots) was observed. Cairns study revealed that the chromosome of *E. coli* was a closed circle of duplex DNA; that replication began at a single point of origin and proceeded in a unidirectional manner round the whole circle, until it reached the point of origin.

However, later studies showed that after initiation at the point of origin, replication was bidirectional and proceeds stepwise as follows. First of all an endonuclease enzyme introduces a break or nick in one strand of the duplex at the point of origin thus allowing the two strands to untwist. The separation of strands is aided by an *unwinding protein* which binds to a small portion of the nicked strand. The successive binding of additional molecules of this protein leads to opening up of the duplex just before the replication fork. An *E. coli* cell contains about 10,000 copies of the unwinding protein. The break is perhaps joined by a ligase enzyme.

Replication begins and proceeds in both directions; it is visible in the electron microscope in the form of a replicating "bubble" or "eye" at the point of origin. Both strands of parental DNA are replicated simultaneously (Fig. 14.8), by movement of two replicating forks away from each other. A little later the circular replicating DNA appears like the greek letter theta (θ). The process of unwinding of the circular double helical DNA creates torque which results in supercoiling. An untwisting protein, also called *swivelase* appears to relieve tension by nicking one strand and allowing some unwinding to take place.

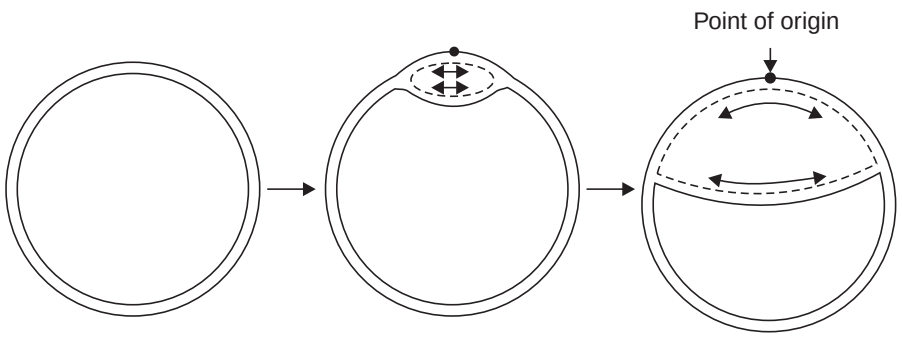


Fig. 14.8 Bidirectional replication of *E. coli* chromosome.

Both strands of the double helix are replicated in short fragments. The DNA polymerase enzymes described earlier are not able to polymerise native double stranded DNA. The enzyme DNA *dependent* RNA polymerase which takes part in transcription of DNA to RNA (described in the next chapter) is able to recognise specific initiation points or chromosomes. Due to the activity of this enzyme DNA first forms a number of short strands of complementary RNA called *primer* RNA, each about 100 nucleotides long in the 5′ to 3′ direction. The 3′ end of the primer RNA serves as a starting place for DNA polymerase to add nucleotide units of DNA. The DNA chain is replicated in the 5′ to 3′ direction and is complementary to a single strand of the parent· DNA duplex. After the DNA chain has attained a length of about 1000 to 2000 nucleotides, an exonuclease enzyme cuts off primer RNA from the 5′ end of DNA. In this way a number of DNA fragments, known as *Okazaki fragments* are generated during replication instead of one continuous linear strand (Fig. 14.9).

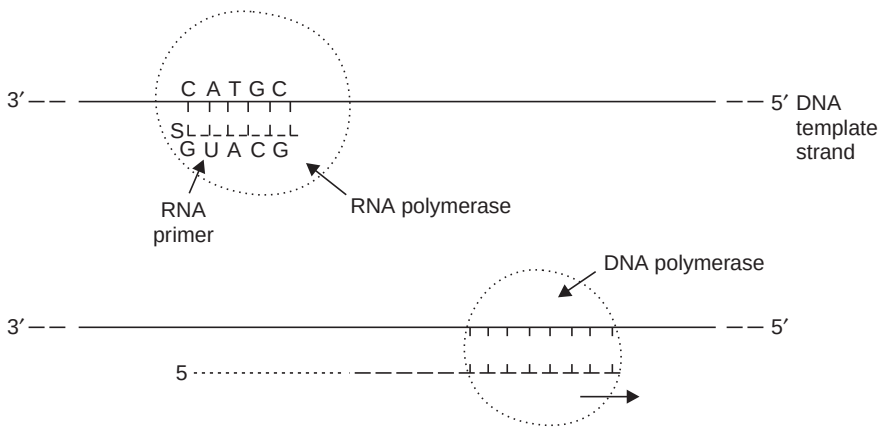
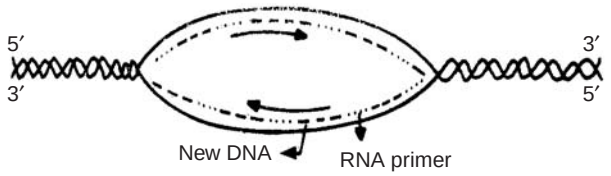
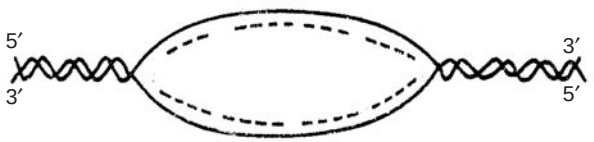


Fig. 14.9 Mechanism of formation of RNA primer in DNA synthesis.

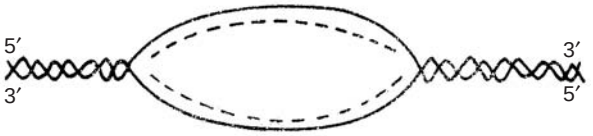


(a) Formation of Okazaki fragment



(b) Removal of primer by exonuclease activity of DNA polymerase I

(c) Gaps filled by DNA polymerase I



(d) Ends at gaps sealed by ligase

Fig. 14.10 Stages in replication of DNA.

The gaps between the DNA fragments are filled by DNA polymerase I which adds nucleotides complementary to the parental strand in the 5′ to 3′ direction. After the gaps are filled, the remaining nicks are removed by the enzyme ligase, which joins adjacent nucleotides by 3′ 5′ phosphodiester linkages.

In addition to its polymerising function, DNA polymerase enzyme also performs exonuclease activity by removing bases in 3′ to 5′ direction. If during the synthesis of a DNA chain an incorrect base is added, the DNA polymerase turns back, removes the wrong base in 3′ to 5′ direction, and again starts adding bases in 5′ to 3′ direction.

The idea that DNA replication takes place by first forming Okazaki fragments received strong support from studies on *E. coli* mutants deficient in DNA ligase enzyme. In such cells the fragments were seen to accumulate in very large numbers.

Cell division in bacteria is accomplished by replication of the single circular DNA chromosome into two identical daughter duplexes. The replication fork in *E. coli* moves at the rate of about 20–30 μm per minute. Replication of the entire genome is completed in about 40 minutes. This is followed by partitioning of cell cytoplasm into two daughter cells. The nuclear membrane is absent; there is no mitosis or spindle formation as in eukaryotes. It seems likely that there is some association between the replicating DNA duplex and the plasma membrane. Electron micrographs of *Bacillus subtilis* have shown that the plasma membrane invaginates to form a mesosome, and the daughter chromosomes become attached to either side of this invagination. There is some evidence from *in vivo* DNA transformation experiments that both the origin and terminal end of replicating DNA are preferentially bound to the membrane. However, as yet, the association between replicating DNA and membrane remains unclear. Further growth of the mesosome partitions the parent cell into two daughter cells.

Modification of Replicated DNA

After synthesis the bacterial DNA becomes modified to a conformation characteristic for that bacterium. There is a class of modification enzymes known as DNA *methylases* which add a methyl group to a newly synthesised adenine (methylate *A*), or sometimes to a cytosine. There is a specific DNA methylase for each species.

DNA REPLICATION IN EUKARYOTES

In eukaryotes there are only two different types of DNA polymerases in contrast with DNA polymerase I, II and III of prokaryotes. Furthermore the DNA of eukaryotes is a long linear molecule with several replication units. A diploid mammalian cell contains on an average about 6 pg of DNA in the *G* phase. This much DNA is equivalent to a length of 2 metres of a linear DNA molecule. If a single replication unit were to move along this length of DNA, it could complete replication within the 8 hour *S* phase only if its rate of movement is about 4 mm/min. This is obviously a very fast rate. The replicating fork actually moves at a slower speed (0.5 to 2.0 micron/min.) in eukaryotes adding about 2,600 bases per minute. In *E. coli* it moves faster adding about 6,000 bases per minute. It is, therefore, necessary that in eukaryotes replication be initiated at several points of origin.

Autoradiographic studies on labelling patterns of individual metaphase chromosomes have shown that multiple adjacent units initiate replication simultaneously. The most convincing

demonstration however, came from similar observations in giant polytene chromosomes. Here tritiated thymidine is incorporated *simultaneously* into a large number of different bands. By the same technique the egg in *Drosophila* is shown to have 6,000 replication forks and all the DNA synthesis is completed within 3 minutes.

The unit of replication is the *replicon*. The size of the replicon is estimated from the distance between adjacent initiation points (centre-to-centre distance). By autoradiography it has been found that units within the same cell are not uniform in size but fall within the range of 15–60 micron. Replicons in rapidly growing cells with short *S* phases are smaller than those in cells growing more slowly with longer *S* phases. Blumenthal (1973) has estimated that in *Drosophila melanogaster* replicons in embryonic cells are as short as 3–4 micron, whereas in a cell line of the same species they were about 13 micron long.

Experimental studies on cultured mammalian (Chinese hamster) cells have shown that the rate of DNA synthesis is not constant throughout the *S* phase, Kleveroz *et al.* (1975) found that synthesis is slow at the beginning of *S* phase, thereafter it increases. About 50% of replication occurs during the last hour of the 5.5 hour long *S* phase. The occurrence of multiple adjacent units has led to the concept that replication units exist in *clusters*. All units in a cluster do not replicate simultaneously, some being late replicating. In mammalian cells there are about 100 replicating units in a cluster.

The essential features of DNA replication are similar in eukaryotes and prokaryotes. After replication begins at a central point of origin in each unit, it proceeds in both directions away from the initiation site. Chain growth occurs by means of fork-like growing points. Electron micrographs therefore show a number of 'eyes' or 'bubbles', each formed between two replicating forks along the linear molecule. It appears that there are no specific term in DNA for stopping replication. The forks travel towards each other and the newly synthesised chains meet and fuse with chains synthesised on adjacent units (Fig. 14.11). In this way long DNA duplexes characteristic of eukaryotic chromosomes are produced.

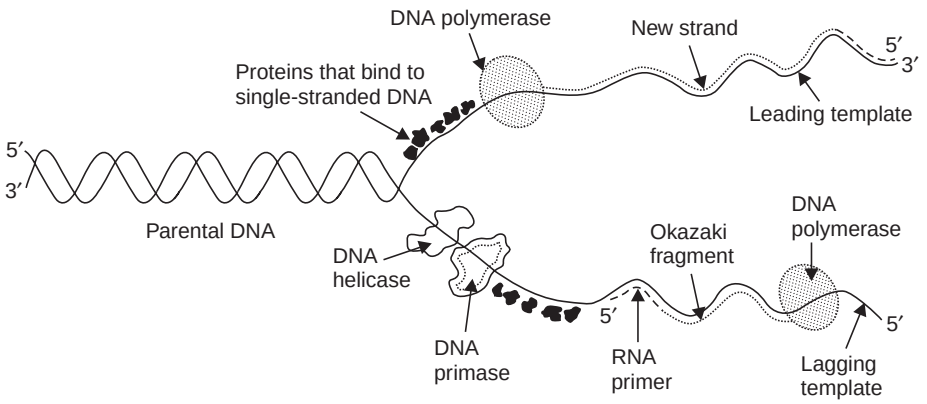


Fig. 14.11 Diagram giving an overview of DNA replication, binding of RNA polymerase, formation of Okazaki fragment, and involvement of a variety of proteins.

As in prokaryotes, the first step in DNA synthesis in eukaryotes is the formation of a primer strand of RNA about 10 nucleotides in length—catalysed by the enzyme RNA polymerase. After that DNA polymerase takes over and adds deoxyribonucleotides to the 3′ end of the primer

RNA. The Okazaki fragments thus formed are shorter in eukaryotes (about 100–150 nucleotides long) than in prokaryotes (1,000 to 2,000 nucleotides). The gaps between the fragments are filled up against the parent DNA template and their ends are joined by DNA ligase enzyme. The RNA primer is digested, starting from its 5′ end by the exonuclease activity of DNA polymerase.

Significance of the RNA Primer in DNA Synthesis

Why should DNA replication be initiated by the enzyme RNA polymerase and formation of RNA strand take place? Detailed analysis of DNA polymerase enzymes have revealed the fact that each polymerase enzyme can add nucleotides only to an already existing polynucleotide chain. These enzymes are not able to initiate new DNA chains. The point of origin in a DNA duplex is perhaps recognised by RNA polymerase, the enzyme which catalyses the synthesis of RNA on a DNA template. In other words, RNA polymerase is required for both RNA and DNA synthesis.

Synthesis of RNA primer on the DNA template continues until a stop signal is reached. The enzyme is then released and the RNA chain serves as a *primer* for addition of DNA nucleotides by DNA polymerase enzyme. However, the molecular mechanism which initiates DNA replication is not fully known.

REPLICATION IN KINETIC CLASSES OF DNA

On the basis of renaturation rates, Britten and Kohne in 1968 found DNA to consist of three types of sequences; highly repetitious, intermediate repetitious and unique. These have been termed *kinetic classes* of DNA. It has been found that in general both highly repetitious and intermediate repetitious types replicate throughout the *S* phase. In mouse and rat Kangaroo highly repetitious sequences may be late replicating.

When purified DNA is centrifuged in a caesium chloride density gradient, a portion of the DNA forms a band at a density different from the main DNA component. This is called *satellite* DNA and by renaturation studies has been shown to contain highly repetitious sequences. Satellite DNA is found to consist of replicating units with the same general characteristics as bulk DNA.

THE ROLLING CIRCLE METHOD OF REPLICATION

All circular genomes do not follow the same pattern of replication described in *E. coli.* In some bacteriophages (phi, lambda and phi *X* 174), in mitochondrial chromosomes and during bacterial mating an alternative method known as the rolling circle has been demonstrated. This method as seen in the small bacteriophage ϕX 174 will be described here.

ϕX 174 has a single stranded DNA molecule only about 6,000 nucleotides in length. Its newly formed DNA is linear, not circular. The process takes place in following way. Soon after ϕX174 enters the host *E. coli* cell its single stranded parent DNA (designated + strand) synthesises a complementary minus strand (with the help of host cell enzymes) to form duplex DNA (+ –). This is the *replicative form* (RF) of ϕXI74. An endonuclease enzyme produces a nick in the + strand, exposing a free 3′ end and a free 5′ end (Fig. 14.12). DNA synthesis begins by addition of deoxyribonucleotides by the enzyme DNA polymerase to the free 3′ end of the + strand using the – strand as template. The minus strand revolves while serving as template hence the name rolling circle. The addition of nucleotides displaces the free 5′ end outwards in the form of a free tail. The

rolling circle revolves a number of times thus increasing the length of the + strand by a corresponding number of + complements.

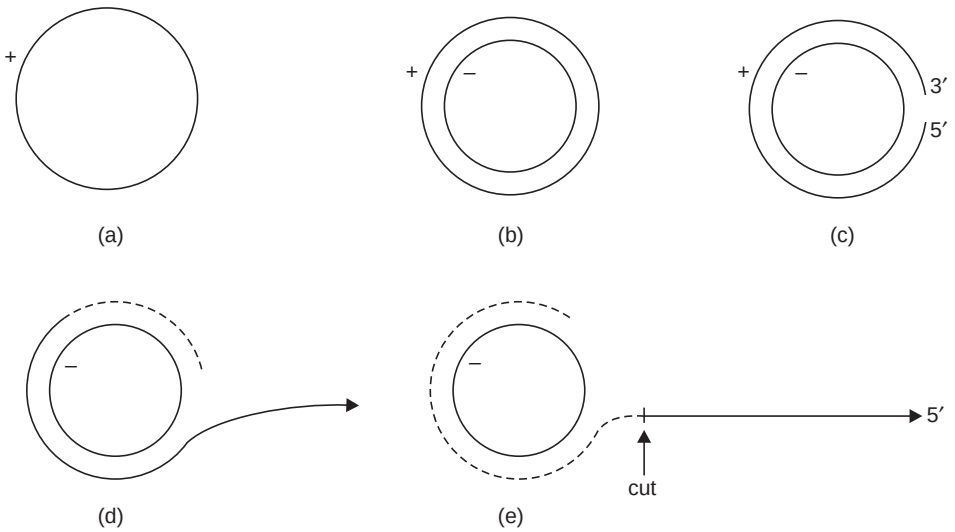


Fig. 14.12 The rolling circle method of DNA replication in $\phi X174$.

It appears that the tail is finally cut into the correct genome lengths by an endonuclease to provide as many copies of free, linear DNA molecules. The free ends joined by the newly formed linear DNA are joined by the enzyme DNA ligase to form closed, circular progeny chromosomes.

REPLICATION IN DNA VIRUSES

Replication in two DNA animal viruses SV40 (Simian virus 40) and polyoma with very small genomes (3×10^6 daltons) has been analysed in detail. In each case replication is initiated at a unique site in the viral genome. The viral genome controls only initiation of replication, and all subsequent processes of replication are dependent on the mammalian host cell. Chain elongation follows the discontinuous method by formation of Okazaki fragments. The primer RNA is about 10 nucleotides long with either adenine or guanine at the 5′ end. At the 3′ end any of the four bases can be present. The RNA primers are subsequently removed, and the fragments sealed by ligase.

QUESTIONS

1. Mention three important works of contemporaries that were used by Watson and Crick in formulating their model for DNA structure.
2. State one reason why (*a*) the bases should be turned inwards in the DNA double helix; (*b*) the DNA should be acidic in solution.
3. Mention three essential differences between Watson and Crick's DNA and left-handed model of the double helix.

4. By which biochemical technique could Meselson and Stahl establish the semiconservative mode of replication?
5. Fill in the gaps: (*a*) That chromosomes in higher plants replicate by the ..... method was first found out by ..... (*b*) Enzymes that produce nicks in the interior of a single strand are called ..... (*c*) Enzymes which break both strands of the double helix where two consecutive nucleotides are identical are called ..... (*d*) ..... is a polymerase enzyme with both nucleolytic and polymerising functions. (*e*) ..... was the discoverer of the mechanism by which the circular chromosome of *E.coli* replicates.
6. DNA synthesis is initiated by the enzyme RNA polymerase, not by DNA polymerase. Explain.
7. If the sequence of nucleotides in one chain of the double helix is *CCGTGCAT*, what would be the sequence in the *corresponding* region of the complementary chain?

SELECTED READINGS

Adegoke, J. A. and Taylor, J.H. 1977. Sequence Programming of DNA Replication Over the S phase of Chinese Hamster Cells. *Exp. Cell Res*. **104**: 47.

Altenburger, W. *et al.,* 1977. Comparative Analysis of Three Guinea Pig Satellite DNAs by Restriction Nucleases. *Eur. J., Biochem.* **73:** 393.

Axel, R. 1975. Cleavage of DNA in Nuclei and Chromatin with Staphylococcal Nuclease. *Biochemistry* **14**: 2921.

Baker, T.A. and Bell, S.P. 1998. Polymerases and the Replisome: Machines within Machines. *Cell* **92**: 296–305.

Bauer, W. R. *et al*., 1980. Supercoiled DNA. *Sc.Amer*. **243**: 100.

Berezney, R. *et al.,* 1981. Dynamic Association of Replicating DNA Fragments with Nuclear Matrix of Regenerating Liver. *Exp. Cell Res*. **132**: 1.

Botchan, P. 1976. An Electron Microscopic Comparison of Transcription of Linear and Supercoiled DNA. *J.Mol. Biol.* **105**: 161.

Botchan, P. *et al.,* 1971. Isolation and Chromosomal Localisation on Highly Repeated DNA Sequences in *D. melanogaster*. *Proc., Natl. Acad. Sci. U.S.* **68**: 1125.

Britten, R.J. and Davidson, E.H. 1971. Repetitive and Non-repetitive DNA Sequences and a Speculation on the Origins of Evolutionary Novelty. *Quart. Rev. Biol.* **46**: 111.

Britten, R. J. and Davidson, E. H. 1976. Studies on Nucleic Acid Reassociation Kinetics; Empirical Equations Describing DNA Reassociation. *Proc. Natl. Acad. Sci. U.S.* **73**: 415.

Britten, R.J., Davidson, E.H. and Kohne, D.E. 1968. Repeated Sequences in DNA. *Science* **161**: 529. Brown, P. O. and Cozzarelli, N. R. 1979. A Sign Inversion Mechanism for Enzymatic Supercoiling of DNA. *Science* **206**: 1081.

Cairns, J. 1963. The Chromosome of *E. coli. Cold Spring Harb. Symp. Quant. Biol.* **28**: 43.

Crick, F. H. C. and Klug, A. 1975. Kinky Helix. *Nature* **255**: 530.

Doolittle, W.F. and Sapienza, C. 1980. Selfish Genes, the Phenotypic Paradigm and Genome Evolution. *Nature* **284**: 601.

DuPraw, E.J. 1970. DNA and Chromosomes. Holt, Reinhart and Winston, New York.

Finch. J. T. and Klug, A. 1976. Solenoidal Model for Superstructure in Chromatin *Proc. Natl. Acad. Sci. U.S.* **73**:1897.

Hand, R. 1979. Eukaryotic Chromosome Replication and its Regulation. *In* Cell Biology, A Comprehensive Treatise. Vol. 2 Eds. Prescott, D. M. and Goldstein, P. Academic Press, New York.

Jain, H.K. 1981. Incidental DNA. *Nature* **288**: 647.

Johnston, *et al.,* 1979. Molecular Principles and the Enzymatic Machinery of DNA Replication. *In* Cell Biology, A Comprehensive Treatise. Vol. 2. Eds. Prescott, D.M. & Goldstein, P. Academic Press, New York.

Kolata, J. 1981. *Z*-DNA: From the Crystal to the Fly. *Science* **214**: 1108.

Kunkel, D.A. 1992. DNA Replication Fidelity. *J. Biol. Chem*. **267** : 18251–18254.

Liu, L.F. and Wang, J. C. 1978. DNA-DNA Gyrase Complex: The Wrapping of the DNA Complex Outside the Enzyme. *Cell* **15**: 979.

Meselson, M. and Stahl, F. W. 1958. The Replication of DNA in *E. coli. Proc. Natl. Acad. Sdi. U.S.* **44**: 671.

Orgel, L.E. and Crick, F.H.C. 1980. Selfish DNA, the Ultimate Parasite. *Nature* **284**: 604.

Orgel, L.E. *et al.,* 1981. Selfish DNA. *Nature* **288**: 645.

Stubblefield, E. 1976. DNA Replication in Chromosomes. *In* The Cell Nucleus. Ed. Busch.H. Vol. 2 p. 155.

Petes, T.D. and Fangman, W. L. 1972. Sedimentation Properties of Yeast Chromosomal DNA. *Proc. Natl. Acad. Sci. U.S.* **69**: 1188.

Taylor, J. H. 1974. Units of DNA Replication in Chromosomes of Eukaryotes. *Int. Rev. Cytol.* **37**: 1.

Wang, J. C. 1979. Molecular Structure of a Left-handed Double Helical DNA Fragment at Atomic Resolution. *Nature* **282**: 680.

Watson, J. D. 1977. Molecular Biology of the Gene. Benjamin Press, New York.

Watson, J. D. and Crick, F. H. C. 1953. The Structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* Vol. XVIII.

Watson, J. D. and Crick, F. H. C. 1953. Genetical Implications of the Structure of Deoxyribose Nucleic Acid. *Nature* **171**: 964.

Wickner, S. H. 1978. DNA Replication Proteins of *E. coli. Annu. Rev. Biochem.* **47**: 1163.

15

Gene Transcription and Translation into Proteins

Besides self duplication, DNA performs the important function of controlling the cell's metabolic processes by providing information for the synthesis of cellular enzymes and proteins. There is a relationship between nucleotide sequences in DNA and amino acids in proteins. The sequence of nucleotides determines exactly the sequence of specific amino acids (primary structure) in the protein as directed by the *genetic code*.

THE CENTRAL DOGMA

However, DNA itself does not directly order amino acid sequences. In 1958 Francis Crick suggested an RNA intermediate and proposed that there is a one-way sequential flow of information from DNA to RNA to protein. This relationship of information transfer between DNA and protein became the **Central Dogma** (meaning a set of beliefs) in molecular biology. In the years that followed, the Central Dogma was modified when it was found that in special cases RNA can direct the synthesis of DNA; under artificial conditions single-stranded DNA can specify protein (Fig. 15.1 A, B; Fig. 15.2).

Fig. 15.1A The Central Dogma.

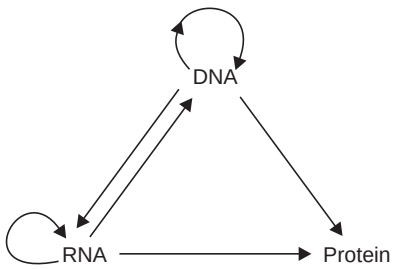


Fig. 15.1B Modification of the Central Dogma.

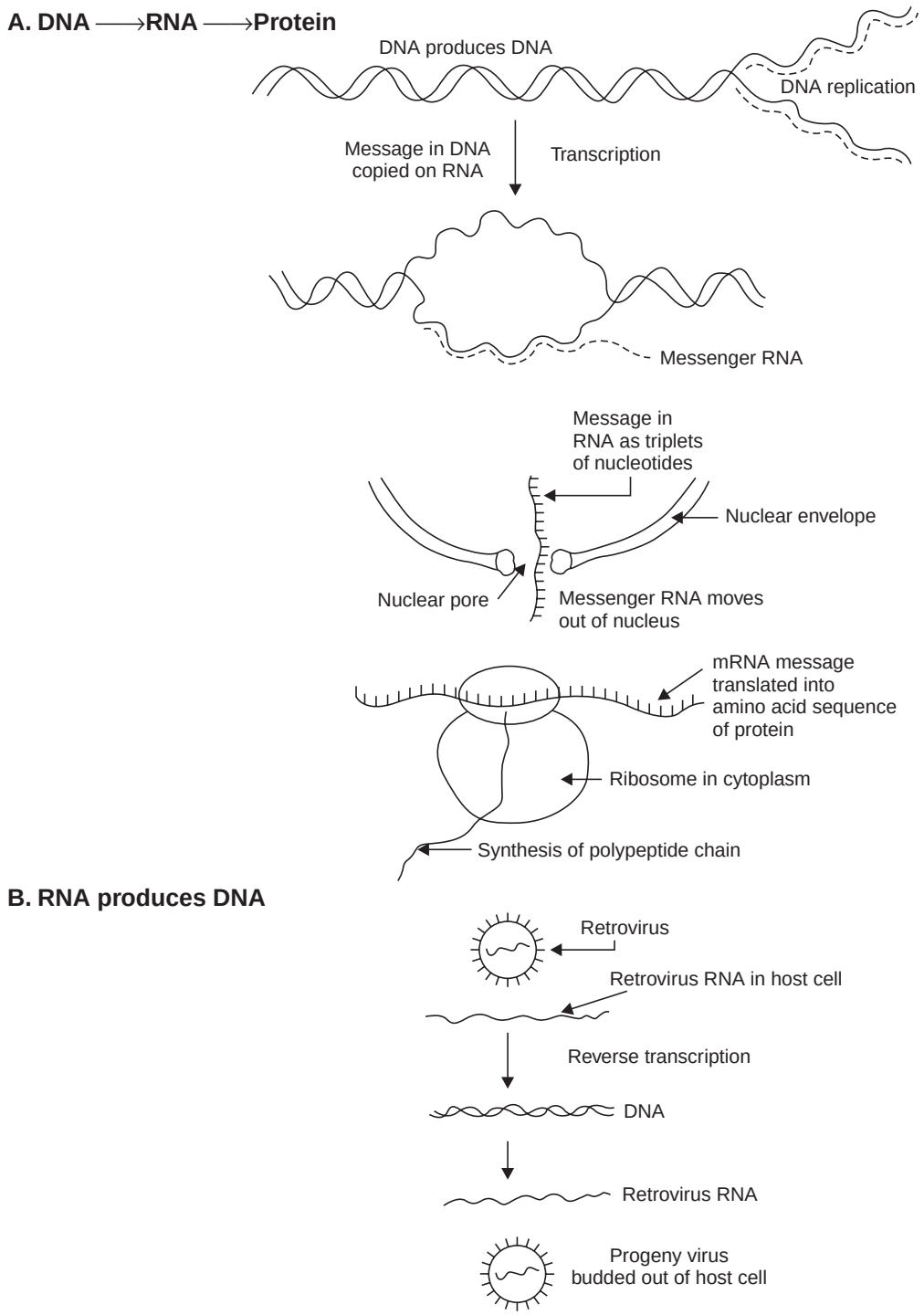


Fig. 15.2 A. Diagram representing the main features of the Central Dogma.
B. The discovery of retroviruses in which RNA gives rise to DNA, led to modification of the Central Dogma.

The One Gene-One Enzyme Hypothesis

As far back as 1915, the work of Archibald Garrod on inborn errors of metabolism in man had suggested a link between a gene and its product protein. But experimental proof for Garrod's hypothesis came almost three decades later from Beadle and Tatum's (1941) studies on the bread mold *Neurospora* (awarded Nobel Prize in 1958). These authors analysed the biochemical effects of genes by studying wild type prototrophs (strains that grow on minimal medium) and auxotrophs (nutritional mutants that grow only on supplemented medium) of *Neurospora* (see chapter 20). They first prepared mutants by UV irradiation of the normal strain. One such strain *thia* failed to grow on minimal medium unless thiamine was added; mutant *pdx* required pyridoxine for its growth.

Experiments of Beadle and Tatum confirmed the heritable nature of the defects in the mutants. When *thia* is crossed to the wild type, the resulting asci contain 4 ascospores from the wild type strain and 4 from *thia* strain. When each of the 8 ascospores is isolated and grown on minimal medium only 4 ascospores are able to grow; the remaining 4 grow after addition of thiamine to the medium. Beadle and Tatum could demonstrate that the UV-induced nutritional mutants in *Neurospora* had single gene mutations that were transmitted by Mendelian inheritance. Since each gene mutation produced a defect in a single enzyme, Beadle and Tatum proposed that *one gene produces one enzyme*.

COLINEARITY OF GENE AND ITS PRODUCT

The most convincing demonstration of the relationship between genes and their product proteins came from two separate investigators, one by C. Yanofsky and his colleagues working on the tryptophan synthetase system in *E. coli*, the other from Sarabhai and his team who were studying bacteriophage.

Yanofsky *et al.* (1967) showed the existence of a colinear relationship between a gene and its protein structure. The tryptophan mutant in *E. coli* (trp^-) cannot grow on minimal medium until the amino acid is supplied from outside. In fact it was found that the trp^- mutants could be defective at one of a number of places in the *trp* biosynthetic pathway and the mutants were likewise named *trp A*, *trp B*, *trp C* and so on. The enzyme tryptophan synthetase is composed of two chains *A* and *B*, each produced by a different gene, that is gene *A* and gene *B* respectively.

Yanofsky *et al.* first focussed on gene *A* and mapped the sites of several mutations within gene A which specified the sequence of the polypeptide chains. They also determined the positions of the changed amino acid within the polypeptide chain of each of the mutant strains. Their earlier work had already established that the mutant polypeptide differed from the wild type protein by a change of a single amino acid. It was clear now that the sequence of the changed amino acids within the polypeptide chain was the same as the sequence of mutations that gave rise to the changed amino acids (Fig. 15.3). The genetic distances between the sites of mutations were proportional to the distance along the polypeptide chain between the altered amino acids. In other words, the positions of the mutant sites in gene A and of changed amino acids in the protein product were **colinear**.

Sarabhai *et al*. (1964) analysed colinearity using *amber* mutants in the bacteriophage T_4. In the class of mutants known as *amber,* translation of the genetic message is terminated at the site of mutation, so that only *fragments* of polypeptides are produced. When *amber* mutations occur in the genes specifying coat proteins, the new points of chain termination cause only *N*-terminal portions of the chain to be made. If *amber* mutants are grown on certain special host

bacteria designated as permissive, translation is not interrupted by chain termination and near normal phenotypes are produced.

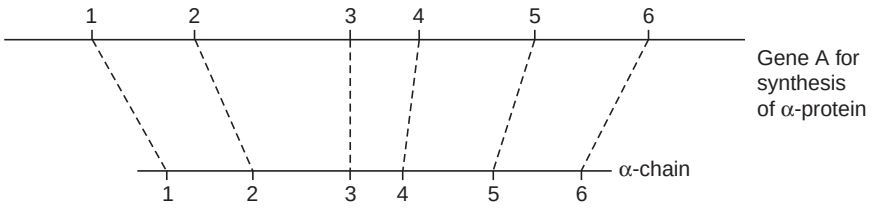


Fig. 15.3 Diagram illustrating colinearity between gene and protein.

Sarabhai *et al.* (1964) found that the lengths of the head polypeptide fragments produced by the various head *amber* mutations could be correlated with the genetic location of the mutation. Mutations near one end of the head gene produced short fragments. When the site of mutation was located further away from this end, the fragment length increased, thus confirming the concept of **colinearity** (Fig. 15.4).

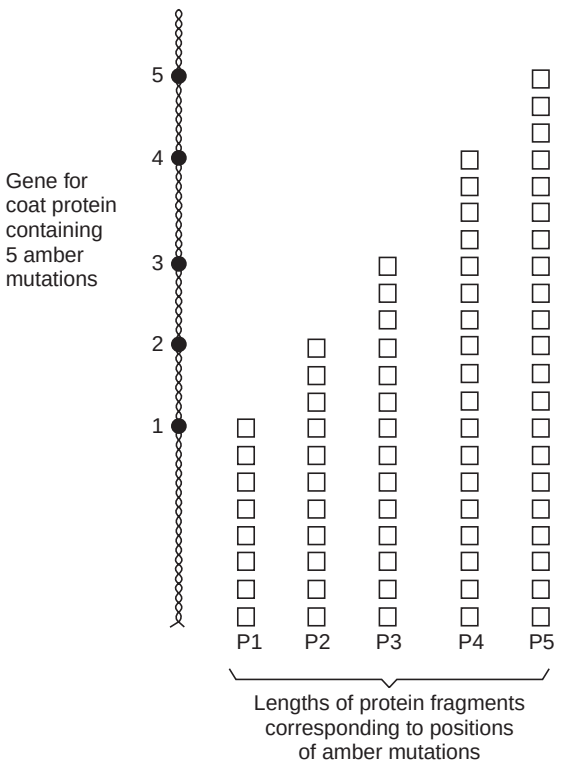


Fig. 15.4 Increasing lengths of coat protein fragments in relation to positions of amber mutations.

TRANSCRIPTION

The genetic information in the nucleotide sequence of DNA is passed on for protein synthesis through an intermediate messenger RNA (*m*RNA). A *m*RNA is a complementary copy of one of

the two DNA strands that make up a gene. The formation of an RNA copy from a DNA template is called **transcription** (Fig. 15.5).

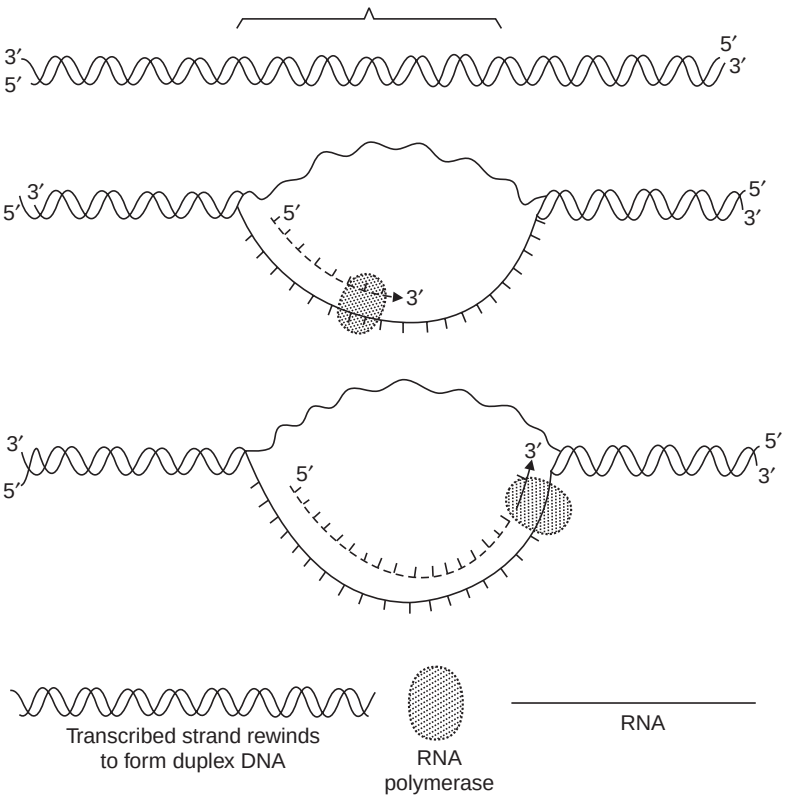


Fig. 15.5 Sequence of steps in the process of transcription.

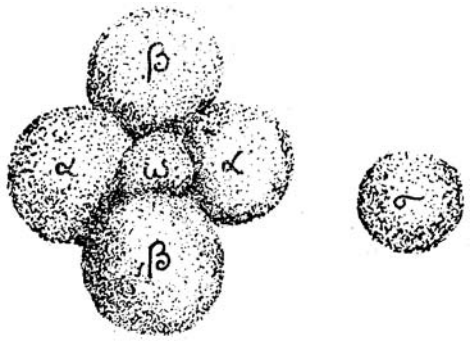
Since the nucleotide sequence of *m*RNA is complementary to that of the gene from which it is transcribed, the information in the *m*RNA is identical to that in the gene itself. The process of transcription occurs throughout interphase and continues up to early prophase of cell division. The enzyme that catalyses DNA to RNA transcription is called RNA polymerase and the RNA molecule produced is the transcript. The chemical synthesis of RNA transcript is similar to that of DNA. But there are following differences:

1. The RNA molecule produced in transcription is synthesised from a single strand of DNA, because in any particular stretch of DNA, usually only one strand of DNA serves as a template for RNA synthesis.
2. The synthesis of RNA uses the four ribonucleoside 5′-triphosphates (ATP, GTP, CTP, UTP) as precursors. They differ from DNA precursors in having a ribose sugar and uracil, instead of deoxyribose and thymine.
3. The sequence of bases in the RNA molecule is complementary to the sequence of bases in the DNA template. Thus, each base that is added to the growing end of the RNA chain is selected if it has the ability to base pair with the DNA template strand. Thus, the bases *T*, *C*, *A* and *G* in DNA template result in addition of *A*, *G*, *U* and *C* respectively to the growing chain of RNA.

4. In the synthesis of RNA, a sugar-phosphate bond is formed between the 3′-hydroxyl group of one nucleotide and the 5′-triphosphate of the next nucleotide in line. The chemical bond formed is the same as in the synthesis of DNA, but the enzyme is different. Transcription uses RNA polymerase rather than DNA polymerase.
5. Nucleotides are added only to the 3′-OH end of the growing chain, with the result that the 5′ end of the growing RNA molecule contains a triphosphate group. The 5′ to 3′ direction of chain growth is identical to that in DNA synthesis.
6. RNA polymerase (unlike DNA polymerase) does not require a preformed primer to initiate chain growth.

RNA Polymerase Enzyme

This is a complex enzyme which in *E. coli* is made up of 5 subunits or polypeptide chains designated β, β′, α, σ and ω with respective molecular weights of 160,000, 150,000, 90,000, 40,000 and 10,000. The α chain is present twice, others only once. The active form of the enzyme is called holoenzyme and has a total molecular weight of 500,000. The polypeptide chains are held together by secondary non-covalent bonds. Chains β, β′, α and ω form the core enzyme; σ (sigma factor) is weakly attached to the other chains and can easily become detached. The core enzyme catalyses the linkage of ribose nucleotides by phosphodiester bonds. The sigma factor recognises start sequences in the promoter region of DNA where transcription begins. In the presence of the sigma factor the holoenzyme binds to those nucleotides in promoter region which initiate transcription. Soon thereafter, the double helix unwinds and one strand serves as the template for transcription.



Diagrammatic view of RNA polymerase

The eukaryotic RNA polymerases I, II and III consist of 8 to 14 different subunits in each. They recognise different promoters and recognise different classes of genes. The two largest subunits of all three eukaryotic RNA polymerases are related to the β and β′ subunits of the *E. coli* polymerase. Five subunits of the eukaryotic RNA polymerases are common to all the three enzymes. These structural similarities allow eukaryotic polymerases to share several common functional properties.

The Site of Transcription, Promoter Recognition

The DNA sequence to which RNA polymerase binds to initiate transcription of a gene is called the **promoter**. The promoter is a specific site at the beginning of genes where transcription is initiated. The initiation process is important because this is the primary step at which transcription is regulated. Typical promoters are from 20 to 200 bases long. The base sequence among different promoters shows variation, and that relates with the strength in binding with the RNA polymerase.

Earlier studies on transcription raised the question as to whether one or both strands of DNA duplex are transcribed. In virus ϕX 174 which has only single-stranded DNA (unlike

most other DNA viruses), only one strand of DNA is transcribed into a complementary RNA sequence. In other viruses with duplex DNA, when we consider the entire genome, both strands of DNA duplex are transcribed, one strand serving as template for some genes, the other strand for remaining genes.

Transcription in Prokaryotes. Promoters in bacteria are located in the region of a DNA strand just preceding the initiation site of RNA transcription. The nucleotide at which transcription is initiated is denoted as + 1 and the preceding nucleotide as – 1. The portions of the

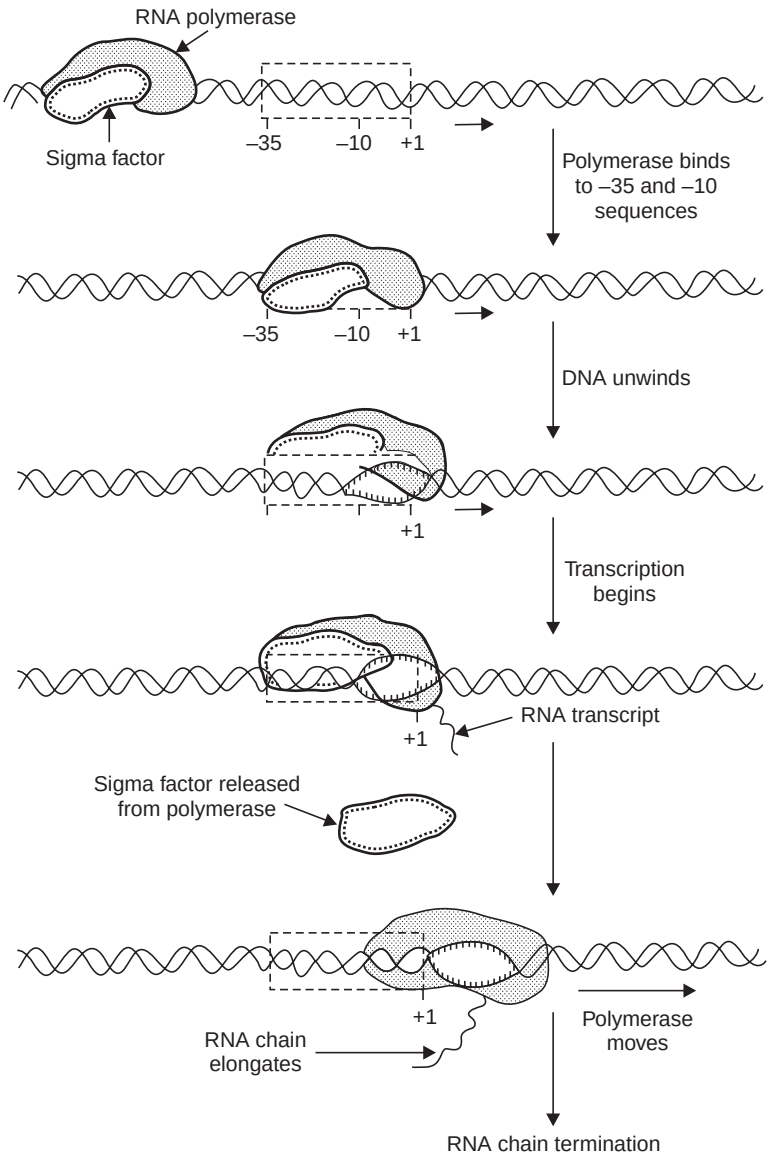


Fig. 15.6 Steps in the process of transcription in *E. coli*. After initial non-specific binding of RNA polymerase to DNA, polymerase moves and sigma subunit binds to promoter elements in – 35 and – 10 position; DNA unwinds and transcription is initiated; sigma subunit then dissociates from core polymerase which moves along DNA and elongates the RNA chain.

DNA preceding the initiation site, toward the 3′ end of the template are said to be **upstream** of that site. Those portions of the DNA succeeding it, toward the 5′ end of the template are said to be downstream of that site. Comparisons of promoter sequences of a series of different genes isolated from *E. coli* revealed that the region upstream of the transcription initiation site contains two sets of sequences that are similar in a variety of genes. These two common sequences, referred to as **consensus sequences**, contain 6 nucleotides each. Each base in the consensus sequence is the base most often observed at that position among any number of observed sequences. Consensus sequences are located approximately 10 and 35 base pairs upstream of the transcription start site. They are called the – 10 and – 35 elements, denoting their position in relation to the transcription initiation site, which is at + 1 position. (Fig. 15.6). The consensus sequence in position 35 in *E. coli* has TTGACA and that in position 10 has TATAAT. The sequences at the – 10 and – 35 positions are not identical in different promoters, but similar enough to establish consensus sequences. The – 10 sequence which is called the TATA box or *Pribnow box* (after the name of its discoverer) is similar to sequences found at corresponding positions in many eukaryotic promoters.

The positions of the promoter sequences determine where and on which strand the RNA polymerase begins synthesis. Experimental evidence supports the functional importance of

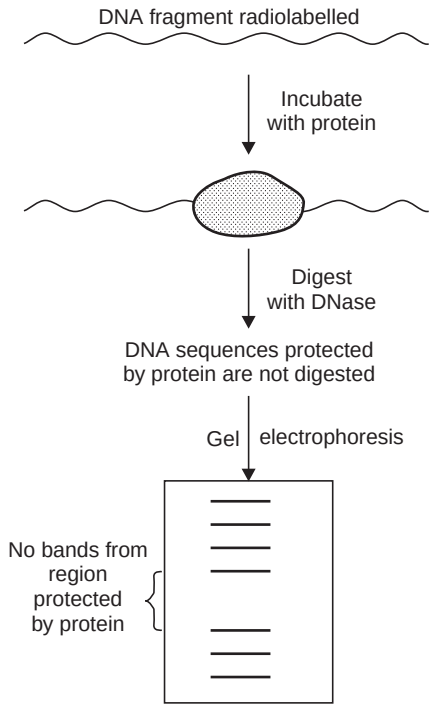


Fig. 15.7 Process of DNA footprinting. DNA fragments radiolabelled at one end are incubated with a protein that binds to a specific DNA sequence within the fragment. The sample is digested with DNase under conditions that cause DNase to produce one cut per molecule. The region of DNA to which protein is bound is protected from digestion by DNase. The DNA-protein complexes are denatured, and radiolabelled fragments of different sizes separated by gel electrophoresis. Fragments of DNA resulting from DNase cut within the region protected by protein binding are missing from the sample. Gel shows absence of bands at sites at which proteins bind to DNA.

the – 10 and – 35 promoter elements. First, genes with promoter elements that differ from the consensus sequences are transcribed less efficiently than genes whose promoters match the consensus sequence more closely. Second, mutations induced in either the – 35 or – 10 consensus sequences have strong effects on promoter function. Third, the sites at which RNA polymerase binds to promoters have been directly identified by **footprinting** experiments, widely used to determine the sites at which proteins bind to DNA (Fig. 15.7).

In these experiments, a DNA fragment is radiolabelled at one end. The labelled DNA is incubated with the protein, for example RNA polymerase, and then subjected to partial digestion with DNase. The method is based on the principle that the regions of DNA to which the protein binds are protected from DNase digestion. A parallel sample of DNA that was not incubated with protein is digested with DNase. The regions of DNA with bound protein can be identified by comparison of the digestion products from protein-bound DNA and DNA without protein. Such footprinting analysis has shown that RNA polymerase generally binds to promoters over approximately a 60 base pair region, extending from – 40 to + 20, that is, from 40 nucleotides upstream to 20 nucleotides downstream of the transcription start site. The sigma (σ factor) subunit of RNA polymerase binds specifically to sequences in both the – 35 and – 10 promoter regions, indicating the importance of these regions in promoter function.

Initiation of Transcription

In absence of sigma subunit, RNA polymerase can bind non-specifically to DNA with low affinity. Sigma plays a crucial role in directing the polymerase to promoters by binding specifically to both the – 35 and – 10 sequences, leading to initiation of transcription at the beginning of a gene. The initial binding between the polymerase and promoter is referred to as closed-promoter complex because the DNA is not unwound. The polymerase then unwinds about 15 bases of DNA around the initiation site to form an open-promoter complex and single stranded DNA is made available for transcription. The first nucleoside triphosphate is placed at the site. The next nucleotide in line is joined to the 3′ carbon of the ribose, and so forth. Transcription is initiated by the joining of two free NTPs at the + 1 site.

Chain Elongation

After addition of about the first ten nucleotides, sigma is released from the polymerase shown in Fig. 15.6. The polymerase then moves along the DNA template strand, adding nucleotides to the 3′ end of the growing RNA chain. Thus RNA chains grow in the 5′ to 3′ direction similar to what is observed in DNA synthesis. As it moves, the polymerase unwinds the template DNA over about 17 base pairs (less than two turns of the double helix) in the region of transcription. After RNA polymerase has passed, the DNA strands reform the duplex.

Chain Termination

The RNA chain continues to grow until the RNA polymerase encounters a termination signal. Then transcription stops, the RNA chain is released from the polymerase, and the enzyme dissociates from its DNA template. The most common type of termination signal in *E. coli* consists of a symmetrical inverted repeat of a *GC*-rich sequence followed by 4 or more A residues. Transcription of the *GC*-rich inverted repeat produces a segment in the growing RNA chain that can form a stem loop structure by complementary base pairing (Fig. 15.8). The formation of a self complementary structure dissociates the chain from the DNA template and terminates transcription. There are other types of transcription termination signals in prokaryotes as well as eukaryotes that involve binding of proteins to specific DNA sequences for chain termination, instead of formation of the stem-loop structure in RNA.

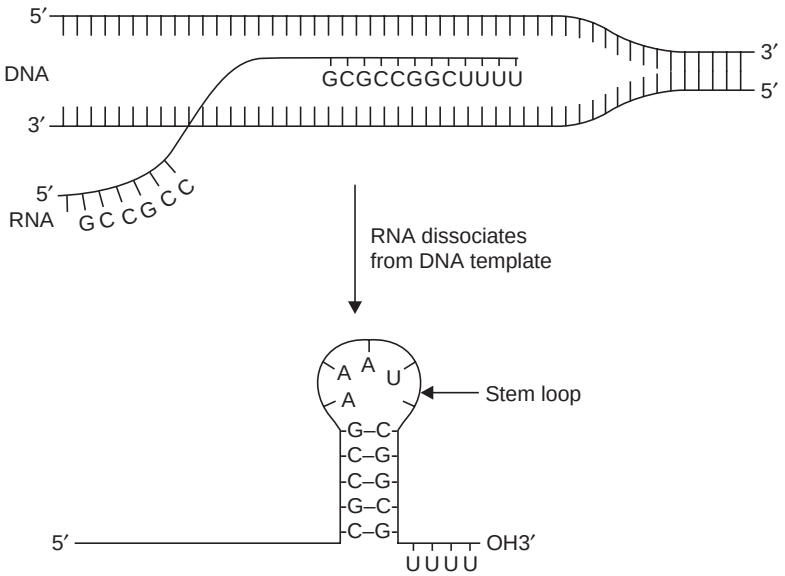


Fig. 15.8 Termination of transcription is signaled by a GC-rich inverted repeat followed by four A residues. The formation of a stem-loop structure by the inverted repeat in the RNA, causes RNA to dissociate from the DNA template.

EUKARYOTIC TRANSCRIPTION

There are two major differences between prokaryotic and eukaryotic transcription systems. First, a single RNA polymerase is able to transcribe all genes in bacteria, whereas eukaryotic cells have multiple different RNA polymerases that transcribe distinct classes of genes. Second, eukaryotic RNA polymerases do not bind directly to promoter sequences, but interact with a number of proteins to specifically initiate transcription. The complexity of the transcription process in eukaryotes is presumed to be related with the regulation of gene expression required to control activities of many different cell types in multicellular forms.

RNA Polymerases

Three distinct nuclear RNA polymerases transcribe different classes of genes in eukaryotic cells (Table).

Table. Classes of genes transcribed by different eukaryotic RNA polymerases

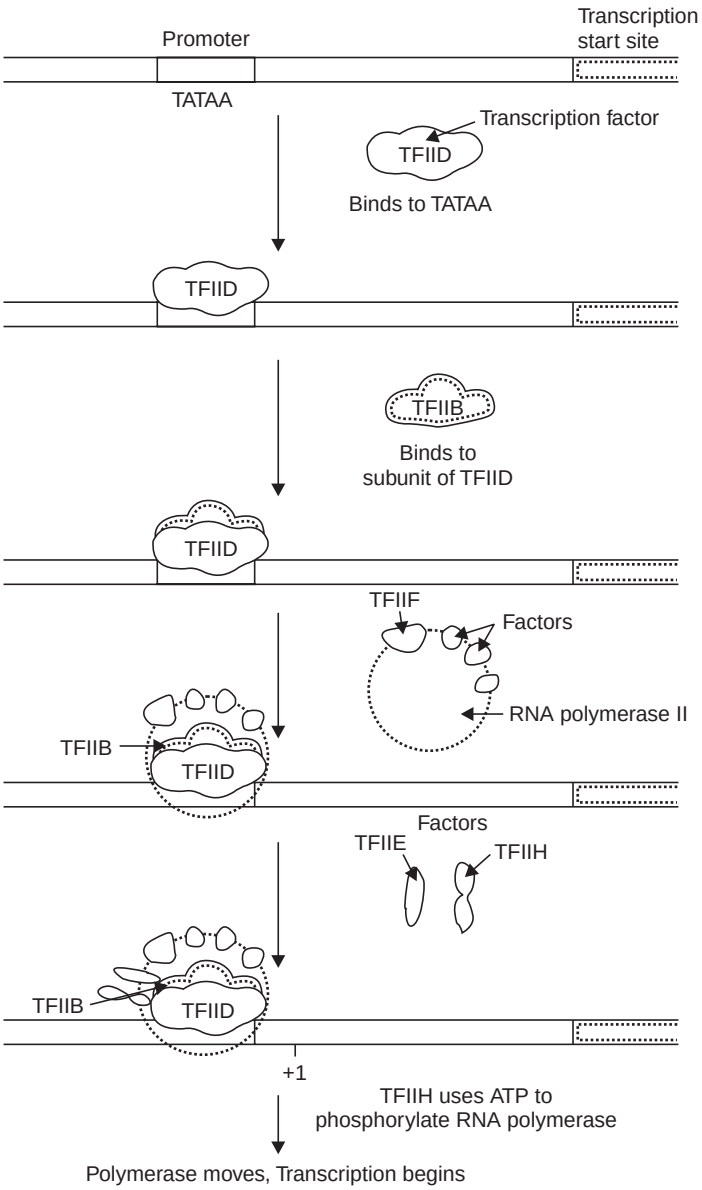
RNA Polymerase	*Type of RNA Synthesised*
II	*m*RNA, small nuclear snRNA
III	*t*RNA, 5S *r*RNA
I	*r*RNA 5.8*S*, 18*S*, 28*S*, small cytoplasmic scRNA

RNA polymerase II transcribes protein coding genes in nucleus to yield *m*RNAs; RNA polymerases I and III transcribe ribosomal RNAs (*r*RNAs) and transfer RNAs (*t*RNAs). The three largest species of *t*RNAs are transcribed by RNA polymerase I. The genes for transfer RNA and

the smallest species of ribosomal RNA (5S *r*RNA), as well as some small nuclear (*sn*RNAs) and cytoplasmic RNAs (*sc*RNAs) involved in splicing and protein transport are transcribed by RNA polymerase III. In addition, mitochondria and chloroplasts contain separate RNA polymerases similar to bacterial RNA polymerases that specifically transcribe DNA in these organelles.

Transcription by RNA Polymerase II

The different mode of action of transcription in eukaryotic cells was noted in 1979 when it was found that RNA polymerase II is able to initiate transcription only if additional proteins are added



Stages in initiation of transcription

to the reaction. In contrast with the bacterial sigma factors, transcription in eukaryotic cells requires distinct initiation factors that were not associated with the polymerase. Specific proteins acting as **transcription factors** have now been identified that are required by RNA polymerase II to initiate transcription. Two types of transcription factors have been defined: *general transcription factors* involved in transcription from all polymerase II promoters*; additional transcription factors* involved in control of expression of individual genes.

Experiments using *in vitro* systems have indicated that five general transcription factors are required for initiation of transcription by RNA polymerase II. The promoters of many genes transcribed by polymerase II contain a sequence similar to TATAA 25 to 30 nucleotides upstream of the transcription start site. This sequence referred to as the **TATA box** is similar to the –10 sequence of bacterial promoters and is involved in initiation of transcription as follows: first, a general transcription factor called TFIID (TF indicates transcription factor, II denotes polymerase II) binds to the TATA box. TFIID has multiple subunits including the **TATA-binding protein** (**TBP**). The TBP binds specifically to the TATAA consensus sequence and 10–12 other polypeptides called TBP-associated factors (**TAFs**). Second, TBP binds to a second general transcription factor (TFIIB) forming a TBP-TFIIB complex at the promoter. Following recruitment of RNA polymerase II to the promoter, two additional factors (TFIIE and TFIIH) are required for initiation of transcription. Two subunits of TFIIH are helicase that unwind DNA around initiation site, while another subunit is a protein kinase that phosphorylates repeated sequences in the largest subunit of RNA polymerase II. In spite of the development of *in vitro* systems, much remains to be elucidated about polymerase II transcription in eukaryotic cells.

Transcription by RNA Polymerases I and III

Like RNA polymerase II, the other two polymerases I and III also require additional transcription factors to associate with appropriate promoter sequences. Although the three eukaryotic polymerases recognise distinct types of promoters, a common transcription factor, the TATA-binding protein (TBP) seems to be required for initiation of transcription by all 3 polymerases. RNA polymerase *I* transcribes **ribosomal RNA genes** which are present in **tandem repeats**, to yield a large 45S pre-*r*RNA, which is then processed to derive the 28S, 18S and 5.8S *r*RNAs (Fig. 15.9). The promoter of *r*RNA genes consists of 150 base pairs just upstream of the transcription initiation site. These promoter sequences are recognised by two transcription factors, UBF (upstream binding factor) and SL1 (selectivity factor 1) which bind to the promoter and then recruit polymerase *I* to form an initiation complex. One of the four protein subunits of the SL1 transcription factor is TBP. Thus, TBP is a common transcription factor required by all 3 types of eukaryotic RNA polymerases. The promoter of ribosomal RNA genes does not contain TATA box, therefore, TBP does not bind to specific promoter sequences. Thus TBP associates with ribosomal RNA genes through the binding of other proteins in the SL1 complex to the promoter.

The genes for *t*RNAs, 5S *r*RNA and some of the small RNAs involved in splicing and protein transport are transcribed by **Polymerase III**. These genes are characterised by promoters that lie within, and not upstream of the transcribed sequence.

The process of termination of transcription was found out from experiments in prokaryotes in which nucleus is not bound by a membrane. Therefore, synthesis of proteins on ribosomes occurs simultaneously with synthesis of *m*RNA on DNA.

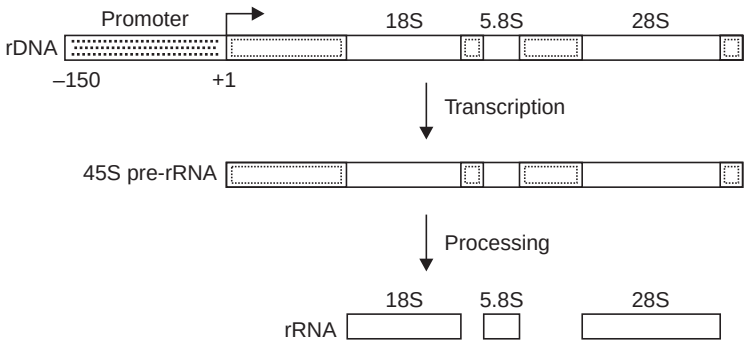
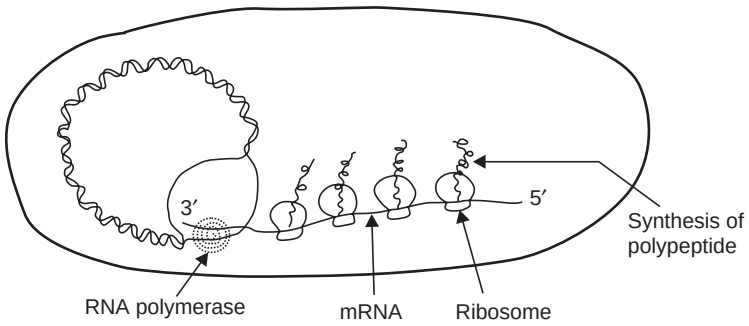


Fig. 15.9 The eukaryotic ribosomal RNA gene (rDNA) is transcribed into a large RNA molecule, the 45S pre-rRNA which is cleaved into the different rRNAs.

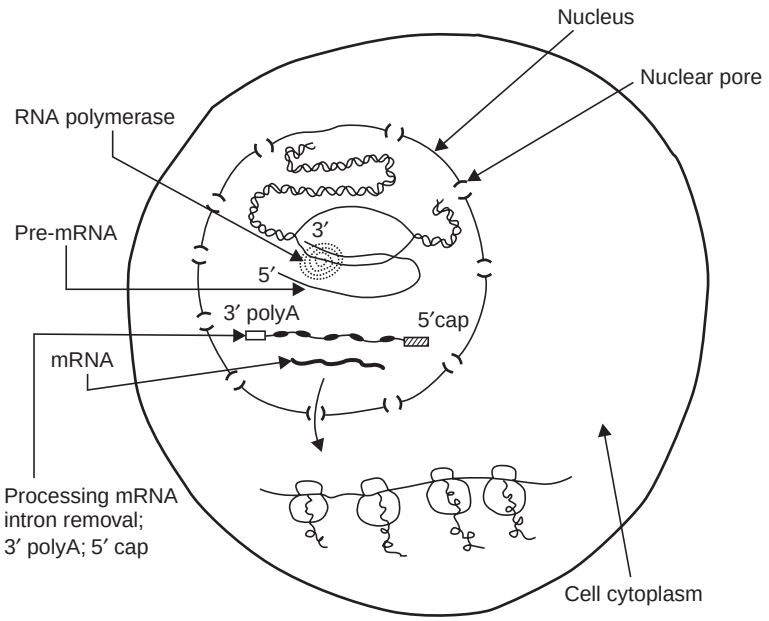
MESSENGER RNA

The existence of messenger RNA (*m*RNA) playing the intermediate role of RNA in protein synthesis was postulated by Jacob and Monod in 1961. Before that RNA in ribosomes was thought to perform the function of *m*RNA. The prevailing notion was, one gene – one ribosome – one protein. In **eukaryotes** the primary product of transcription consists of large RNA molecules of variable lengths and sedimentation coefficients between 20S and 200S, called **pre-mRNA** (earlier heterogeneous nuclear RNA, *hn*RNA). The complete RNA transcript or pre-*m*RNA is not translated into protein. Instead, only a fraction of RNA which has coding sequences is spliced off and modified to produce mature *m*RNA molecules that are transported through nuclear pores in the nuclear membrane to cytoplasmic ribosomes for protein synthesis. The coding sequences termed *exons* are interrupted by non-coding sequences, the *introns* that never leave the nucleus and are not translated. The introns are removed by splicing, that is they are cut, removed and the coding sequences (exons) are sealed with each other. In contrast, the prokaryotic RNA transcript functions directly as *m*RNA. The absence of nuclear membrane around chromatinic DNA in prokaryotes allows ribosomes access to newly formed *m*RNA and translation into protein begins on the nascent RNA chain, while transcription is still in progress (Fig. 15.10). In bacteria genes performing closely related functions are usually clustered together in the genome. They are transcribed into large sized *polycistronic m*RNAs which are translated into several polypeptide chains. Each polypeptide chain is synthesised independently from its own distinct initiation site. In contrast, eukaryotic *m*RNAs are *monocistronic*, encoding a single polypeptide chain.

Thus, in eukaryotes, pre-*m*RNA undergoes processing for 3 distinct modifications before export from the nucleus. Processing includes modifications at both ends of the RNA molecule as well as removal of introns. The 5′ end of pre-*m*RNA is modified by addition of a structure called a **7-methylguanosine cap**. Capping is initiated by the addition of a GTP to the 5′ terminal nucleotide of the pre-*m*RNA. Methyl groups are then added to this *G* residue and to the ribose of the 5′ nucleotide of the RNA chain. The components of a cap are a cluster of methyl groups. Quantitative studies suggest that there is about one cap for every 2,200 bases in He La cell *m*RNA. The 5′ cap aligns eukaryotic *m*RNAs on the ribosome during translation.



Prokaryotes: protein synthesis occurs simultaneously with synthesis of mRNA



Eukaryotes: Spatial separation of sites of synthesis of mRNA (nucleus) and protein synthesis (cell cytoplasm)

Fig. 15.10 Figure above shows simultaneous translation of message in RNA transcript in a prokaryotic cell. Lower figure displays spatial and temporal separation of transcription and translation in a eukaryotic cell; RNA transcript migrates out of nucleus into cytoplasm where it is translated by ribosomes.

The 3′ end of the eukaryotic *m*RNA undergoes a processing reaction for addition of a **poly-A tail**, called **polyadenylation**. The addition of a poly-A tail is signalled by the hexanucleotide sequence AAUAAA which is located 10 to 30 nucleotides upstream of the site of polyadenylation. This is the most conserved sequence for polyadenylation in mammalian cells, and other less conserved sequences also signal polyadenylation. These sequences are recognised by a complex of proteins, including a restriction endonuclease enzyme that cuts the RNA chain. A separate poly-A

polymerase adds a poly-A tail of about 200 nucleotides to the transcript. Polyadenylation signals the termination of transcription. Poly-A tails in most eukaryotes regulate translation and contribute to *m*RNA stability.

The third modification of pre-*m*RNA is the removal of introns by **splicing**. Most genes contain multiple introns. *In vitro* experiments have shown that splicing of pre-*m*RNAs proceeds in 2 steps (Fig. 15.11). First, the pre-*m*RNA is cleaved at the 5′ splice site, and the 5′ end of the intron is joined to an adenine nucleotide within the intron, near its 3′ end. In this step a bond forms between the 5′ end of the intron and the 2′ hydroxyl group of the adenine nucleotide that results in a closed loop structure. The intron is present in the loop. Second, a cut is formed at the 3′ splice site, and the two exons are joined by a ligase enzyme. The intron is excised in the form of a closed loop (lariat) that is linearised and degraded within the nucleus. Biochemical experiments have demonstrated that splicing takes place in large complexes called **spliceosomes**, composed of proteins and RNAs. The spliceosome contains five types of **small nuclear RNAs** (**snRNAs**) called *U*1, *U*2, *U*4, *U*5 and *U*6. The snRNAs range in size from 50 to 200 nucleotides, and are complexed with six to ten protein molecules to form small nuclear ribonucleoprotein molecules (**snRNPs**) which play a key role in the splicing process.

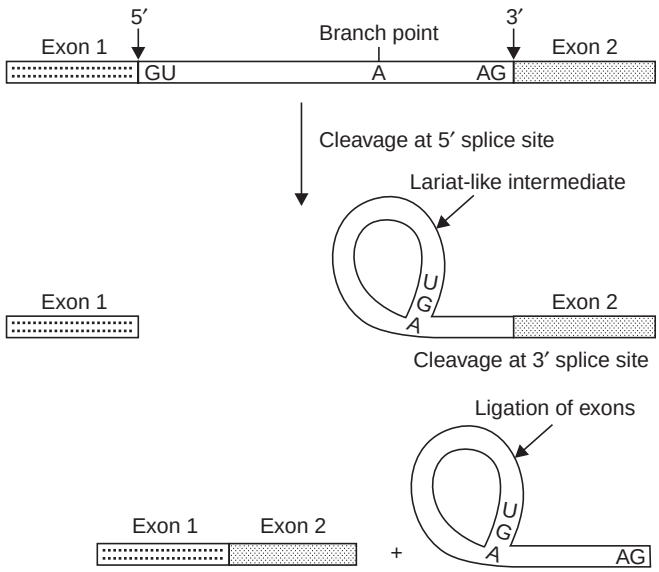


Fig. 15.11 Splicing of pre-mRNA occurs in two steps: First, there is cleavage at the 5′ splice site and joining of the 5′ end of the intron to an A position in the intron (branch point), leading to the formation of a closed loop (lariat-like) intermediate structure by the intron. In the second step, cleavage occurs at the 3′ splice site with simultaneous ligation of the exons. The intron is thus excised as a lariat-like structure.

Visualisation of Transcription

The idea of visualising gene transcription originally arose from light microscopic studies of lampbrush chromosomes in amphibian oocytes performed by Callan and Lloyd, and Gall in the 1960's; similar studies were done on the puffed polytene chromosomes of insects by Beermann and his colleagues. Oocyte-chromosomes are highly extended in the lampbrush state and contain thousands of chromosomal loci active in RNA synthesis (Chapter 19). Another favourable attribute

of oocytes is that there is amplification (manifold increase) of *r*RNA genes during early oogenesis giving rise to hundreds of extra nucleoli in a nucleolus. In this system, transcription of ribosomal cistrons has been visualised.

During transcription on oocyte lampbrush chromosomes, the DNA in the condensed, bead-like chromomere unravels and is spun out into a loop and transcribed. The loop axis becomes covered by the transcribed RNA fibrils embedded in a protein matrix (Fig. 15.12). At the base of each RNP (ribonucleoprotein) fibril, an RNA polymerase molecule is attached. In male meiosis and somatic cells transcription produces fine hair-like outgrowths from the chromosomes (Fig. 15.13).

Puffing in the giant polytene chromosomes in the salivary glands of insects represents a direct way of correlating chromosome structure with gene transcription. Puff formation indicates genes that are actively transcribing RNA which would be translated into salivary proteins. Further work of Grossbach (1969) and others made it possible to relate the synthesis of a specific protein with a specific puff.

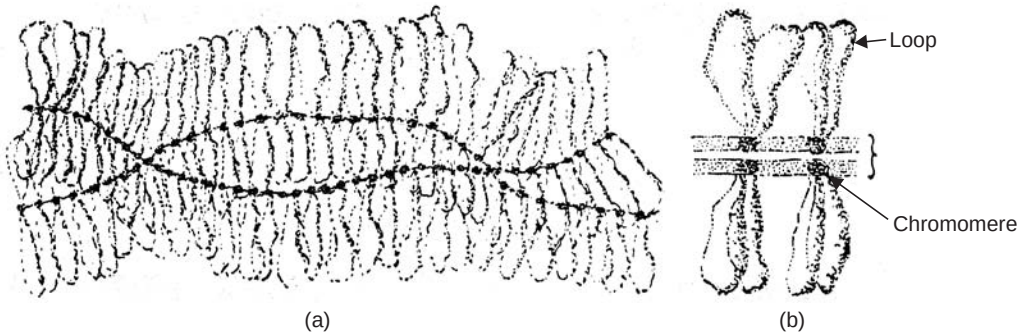


Fig. 15.12 (*a*) a portion from lampbrush chromosomes in amphibian oocytes and (*b*) schematic outline to show details of loop formation.

Fig. 15.13 The lampbrush state at meiotic prophase in the human male.

In 1969 Miller and Beatty developed a spreading technique for chromatin by which nascent RNP transcripts could be visualised in the electron microscope. The technique has since been applied to various materials. It allows us to visualise the spatial relationships between DNA, RNA polymerase and the RNA transcripts *in situ*.

EM studies have confirmed that most of the DNP (deoxyribonucleoprotein) fibrils are entangled in the chromomeric mass and only a small proportion is extended into lateral loops. The initiation and termination sites for transcription are located at the two ends of the loop *i.e.,* the thick and thin insertion sites of the loop. The lateral RNP fibrils which contain the nascent RNA transcripts are of increasing length. One loop is said to represent one transcriptional unit. Many times an active loop shows tandemly arranged transcription units separated by spacer regions (Fig. 15.14). The spacers represent non-transcribed regions. Up to 5 transcriptional units may be present on a loop (Scheer *et al.,* 1979); the loop is therefore a multigene structure. These authors also found initiation sites of two transcriptional units overlapping each other. The drug actinomycin-*D* which inhibits transcription removes RNP fibrils from the template and causes loops to collapse. The initial products of transcription observed in EM are longer than the average hnRNA molecules isolated by biochemical techniques. Similar studies on transcription have also been conducted on interphase nuclei of somatic cells in some organisms.

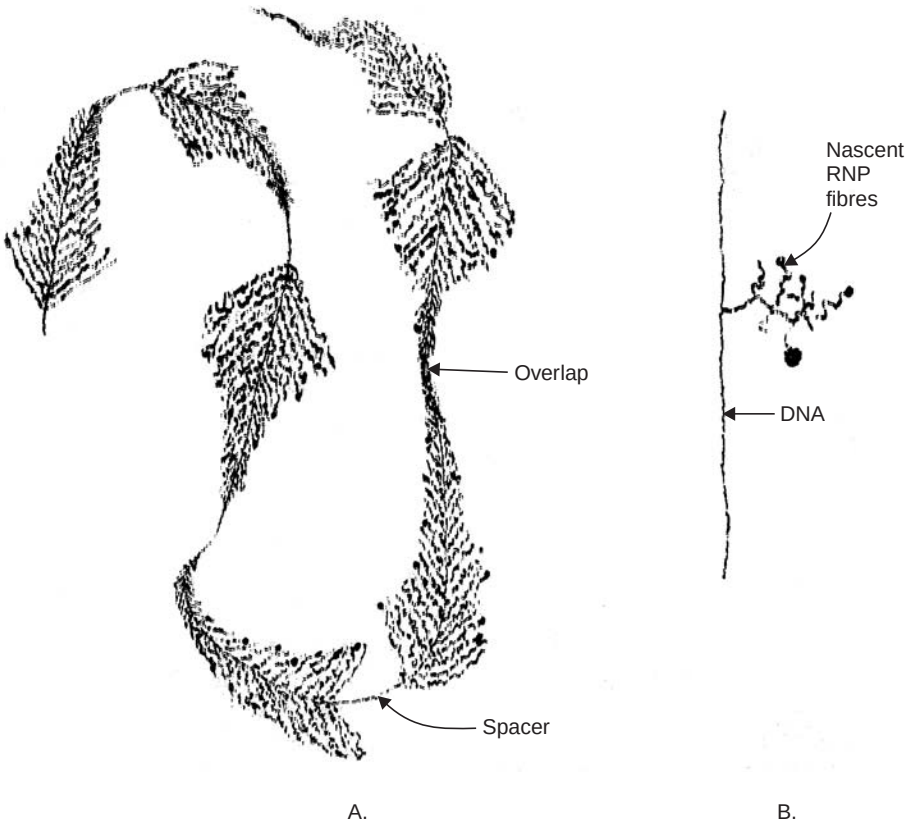


Fig. 15.14 Visualisation of transcribing genes. A—Transcription units are tandemly repeated on a DNA axis. Fig. A shows transcription of ribosomal cistrons and Fig. B exhibits non-ribosomal transcription.

Inhibitors of Transcription

Several compounds can inhibit transcription of DNA by RNA polymerase. One group of compounds acts by binding non-covalently to the DNA template and modifying its structure; the other group

binds to the RNA polymerase and inhibits its catalytic function. The most important inhibitor is actinomycin-*D* (AMD), an antibiotic produced by streptomyces. Its phenoxazene ring intercalates between two *GC* pairs, while its two peptide side chains form *H*-bonds with guanine bases and project into minor groove of the double helix. AMD does not interfere with the binding of RNA polymerase to DNA but inhibits chain elongation by preventing movement of core enzyme along the template. AMD does not interfere with replication of DNA. Aflatoxin, ethidium bromide and 2-acetylaminofluorine also inhibit transcription by binding to DNA.

A group of bacterial antibiotics called rifamycins act by inhibiting bacterial RNA polymerases. One such compound known as *rifampicin* binds non-covalently to the β subunit of RNA polymerase so that chain initiation is inhibited, but does not affect chain elongation. α-amanitin blocks one of the RNA polymerase enzymes present in eukaryotic cells; but bacterial mitochondrial or chloroplast RNA polymerases are not affected by it.

REVERSE TRANSCRIPTION

There are some cancer causing RNA viruses such as the Rous sarcoma virus *(RSV)* which produces cancer in chickens, avian myeloblastosis virus *(AMV)* causing leukemia in birds, and the mammary cancer and leukemia viruses in rodents. The viruses can cause heritable transformation of normal cells of animals and man into malignant cells *in vitro*. In the early 1960s H.M. Temin had suggested that RNA of cancer viruses must be transcribing a DNA intermediate in order to become integrated into the host cell genome. But this view was not accepted as it contradicted the central dogma, according to which information flows only from DNA to RNA to protein. This idea was substantiated later when Temin and D. Baltimore independently discovered *RNA-directed DNA polymerases* in RNA tumour viruses. The enzymes were also called reverse transcriptases. In 1975 Temin and Baltimore were awarded the Nobel Prize.

The purified *reverse transcriptases* from different RNA viruses have molecular weights ranging between 70,000 and 160,000. Like RNA polymerases and the DNA directed DNA polymerases, the reverse transcriptases also synthesise DNA chains in the 5′ to 3′ direction using deoxyribonucleoside triphosphate as precursors on an RNA template. They cannot initiate new DNA chains but can only add deoxynucleotides to a short RNA primer strand with a free 3′-OH end. The primer strand appears to be some *t*RNA molecule. Reverse transcription produces in the host cell cytoplasm RNA-DNA hybrid molecules in which the single DNA strands are complementary to the RNA template of the viral genome. The reverse transcriptase catalyses the breakdown of the RNA releasing the single stranded DNA. The DNA strands become converted into double helical and circular configurations forming the proviruses which become integrated into the host genome by recombination (Fig. 15.15). After integration the proviral DNA is transcribed by RNA polymerase II of the host cell.

It seems that reverse transcription occurs in eukaryotes both in normal cells and in cells infected with RNA tumour viruses. Reverse transcriptases have been isolated from malignant cells of some animals and from human patients of leukemia. Such transcriptases can utilise either RNA or DNA as template although RNA is preferred.

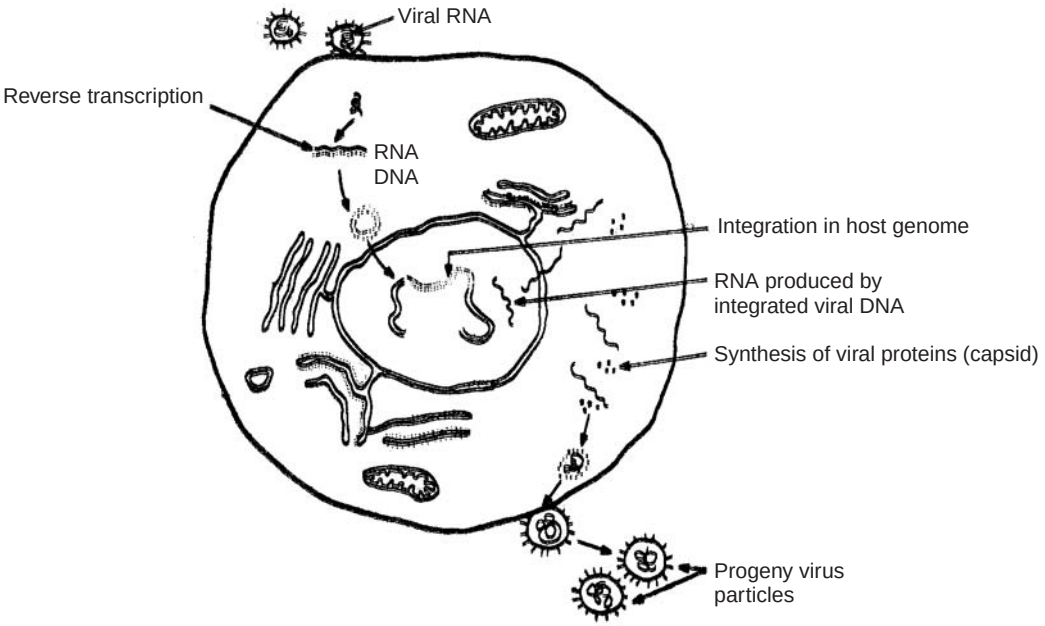


Fig. 15.15 Reverse transcription in an RNA tumour virus.

RNA AND PROTEIN SYNTHESIS

Structure of RNA

RNA is a long unbranched polynucleotide chain chemically similar to DNA except for two major differences. The sugar in DNA is a deoxyribose, in RNA a ribose. Out of the 4 nucleotides, three (G, C and A) are identical in DNA and RNA. The fourth is thymine in DNA, whereas in RNA the closely related pyrimidine uracil is present (Fig. 15.16). RNA molecules are usually single stranded and do not have complementary base ratios. The amount of adenine is often not equal to the amount of uracil, and the amounts of guanine and cytosine are also not always equal. Modified bases such pseudouridine (Ψ) and methylated bases are present; these are particularly abundant in *t*RNAs.

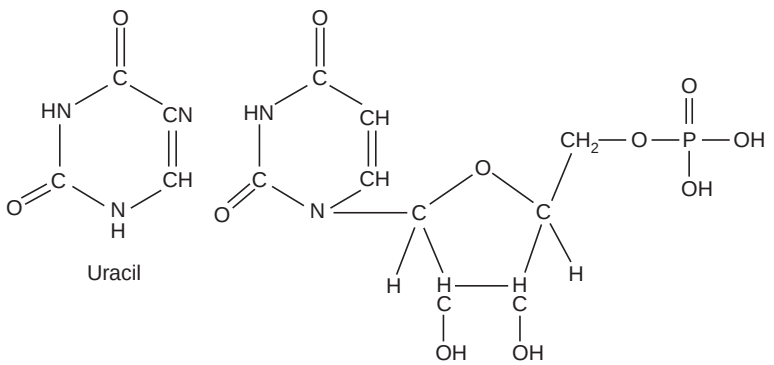


Fig. 15.16 The base uracil and its mononucleotide in RNA.

The products of transcription are 3 types of RNA, namely transfer RNA (*t*RNA), heterogeneous nuclear RNA (*hn*RNA), and ribosomal RNA (*r*RNA) all of which are involved in protein synthesis.

TRANSFER RNA

The genetic message contained in the nucleotide sequence of transcribed RNA cannot be read off directly by the amino acids in the polypeptide chain. Crick and Hoagland were first to postulate that an intermediate RNA molecule functions as an "adaptor" into which the amino acid is plugged so that it can be adapted to the nucleotide triplet language of the genetic code. When first discovered it was given the name soluble RNA (*s*RNA), but later it was called transfer RNA (*t*RNA) which described its function more accurately.

A particular *t*RNA first attaches to its specific amino acid and transports it to the ribosome against a specific nucleotide sequence on the *m*RNA. The *t*RNAs for 20 different amino acids all have different structures, and each is specifically adapted for *H*-bonding with the nucleotide sequences in *m*RNA. Each amino acid has at least one corresponding *t*RNA and some have multiple *t*RNAs. For example there are 5 distinctly different *t*RNAs specific for leucine in *E. coli* cells. Moreover there are different types of *t*RNAs for a given amino acid present in mitochondria and in the cytoplasm of eukaryotic cells.

The *t*RNA molecules are single-stranded, have about 73 to 93 nucleotides, a molecular weight of about 25,000 and sedimentation coefficient of about 4S. The single-stranded molecules can fold on themselves forming hairpins and many of the bases become linked by *H*-bonds. The adenine residues pair with uracils and guanines with cytosines. Due to this *intrachain* base pairing about 60–70 per cent of the *t*RNA structure exists in double stranded form. Although *t*RNAs for different amino acids have quite different base compositions, yet due to folding and intrachain base pairing, all of them basically form a four-leaved clover-like structure (rarely three-leaved as in alanyl-*t*RNA). The single-stranded regions in *t*RNA are represented by loops (Fig 15.17).

In higher organisms the genes for *t*RNA are reiterated (repeated). Thus bacteria have only 1 or 2 copies of each *t*RNA gene, yeasts have 5–7 copies, *Drosophila* has 13 copies, and *Xenopus* and mammalian cells about 200 (Min Jou *et al*., 1972). In eukaryotes, precursors to *t*RNAs reach the cytoplasm; they are longer than *t*RNA by about 30 nucleotides. Within an hour after reaching the cytoplasm they are cleaved to the dimensions of *t*RNA and some nucleotides become modified by methylation. Some of the uridine residues change into pseudouridine and dihydrouridine residues. Most *t*RNAs have many unusual bases such as inosine (*I*), pseudouracil (Ψ) and some methylated forms of normal bases.

The first complete nucleotide sequence of a *t*RNA was established by Holley and his colleagues in 1965 for yeast alanine *t*RNA. This was rewarded by a Nobel Prize 3 years later. All the *t*RNAs have the same terminal sequence CCA at the 3′ end of the *amino acid arm.* The last residue adenylic acid is the site to which the aminoacyl group is esterified by the enzyme aminoacyl synthetase. Besides the CCA terminus, *t*RNAs have three loops. The first *TC loop* of all *t*RNAs contains the pseudouridine (Ψ) residue; in the second or dihydrouridine (DHU) *loop,* a dihydrouridine is present. The third is the *anticodon* loop containing a specific triplet of bases which is different in *t*RNAs for different amino acids. This triplet is also complementary to the

corresponding codon triplets in *m*RNA with which it can form *H*-bonds; The anticodon arm ensures that the correct amino acid is selected and placed on the ribosome for addition to a growing polypeptide chain. The third base of the anticodon occupying the 5′ position is the wobble base which is not so specific as the other two bases in its interaction with the corresponding base in the *t*RNA codon.

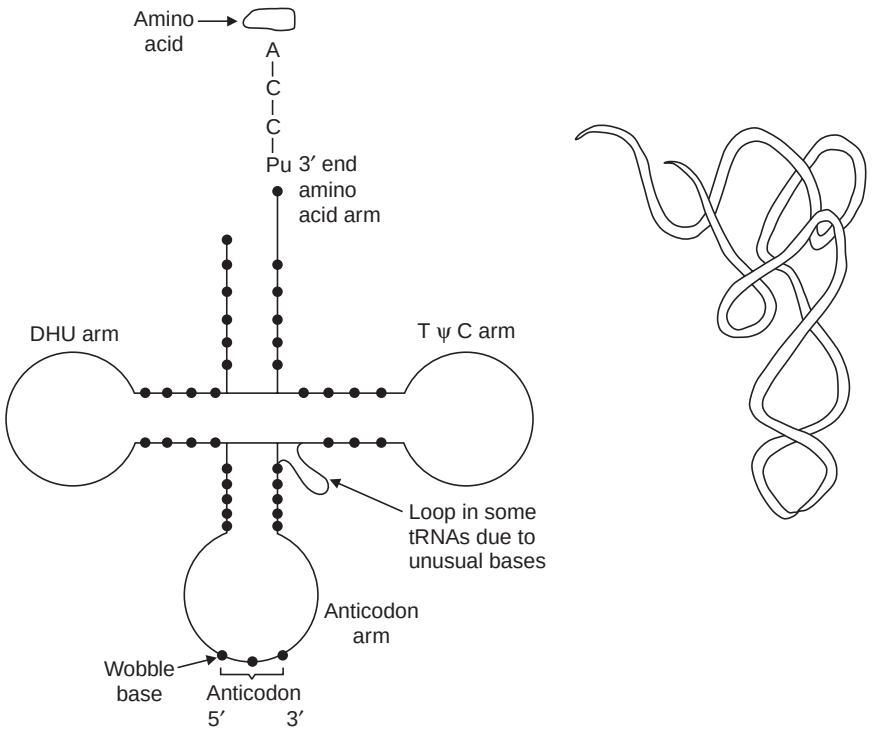


Fig. 15.17 Clover leaf model and three dimensional view of transfer RNA.

A *t*RNA contains at least two other specific sites, the *enzyme recognition* site, at which the *t*RNA is bound to the corresponding activating enzyme, and a second *ribosome attachment site.*

The presence of unusual nucleotides is a striking feature of most *t*RNAs. They probably arise by enzymic modifications of the already existing bases. Since unusual bases cannot form base pairs, they lie exposed in the single stranded loops of the *t*RNA molecule.

RIBOSOMES AND RIBOSOMAL RNA

Ribosomes are ribonucleoprotein particles present in all types of cells. They were first observed in EM by Claude in the cytoplasm of cells, and later on the surface of endoplasmic reticulum by Porter and Palade.

Ribosomes occur in 3 sizes: 70S in bacteria and chloroplasts, 60S in mitochondria, 80S in cytoplasm of eukaryotes. All ribosomes consist of two unequal subunits each containing RNA and protein in the ratio of 63 : 37. In bacteria the 70S ribosomes have 50S and 30S subunits and a diameter of 18 nm. An *E. coli* cell contains about 15,000 ribosomes constituting about 25 per cent

of the dry weight of the cell. The cytoplasmic ribosomes of eukaryotic cells are larger having a diameter of 20–22 nm (Fig. 15.18), and have a higher protein content and larger RNA molecules. They either lie free in the cytoplasm or are bound to the endoplasmic reticulum. In eukaryotes ribosomes are also present in the nucleus (they are assembled in the nucleolus) and in organelles like mitochondria and chloroplasts. In cells that are synthesising proteins, electron micrographs show ribosomes associated in bead-like strings called polyribosomes or polysomes formed by attachment of ribosomes to a single molecule of *m*RNA.

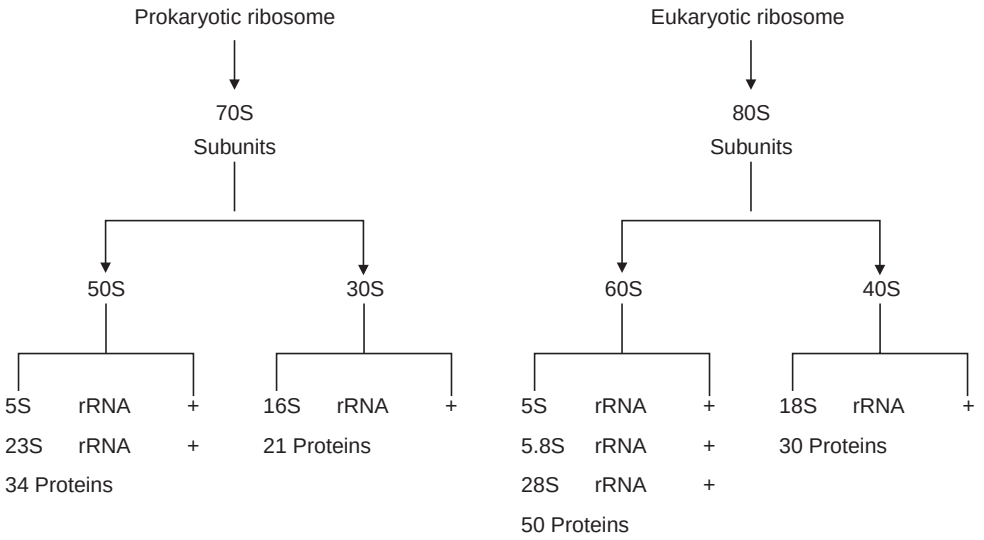


Fig. 15.18 Comparison between prokaryotic and eukaryotic ribosomes.

The 70S ribosomes of prokaryotes have in the larger 50S subunit, 23S and 5S types of *r*RNA and 34 different proteins. The 30S subunit contains a molecule of 16S *r*RNA and 21 different proteins. The eukaryotic cytoplasmic ribosomes show variation in sizes of subunits in different plants and animals and in size of *r*RNA. Plant ribosomes contain 25S and 16S or 18S *r*RNA whereas animal ribosomes contain 28S and 18S *r*RNA; 5S RNA is present in the larger subunit of plants and animals. The 28S RNA of eukaryotes has a small *r*RNA called 5.8S *r*RNA. It has about 160 bases and is bound to the 28S *r*RNA.

In 1964 Watson had suggested that ribosomes dissociate into their subunits and again reassociate during protein synthesis. This was later confirmed by Kaempfer in *E. coli.* It is necessary for the subunits to dissociate because *m*RNA and the initiating aminoacyl *t*RNA *cannot* bind directly to 70S ribosome, but are first bound to the 30S subunit.

Recently, high resolution studies on ribosomes have revealed new features in their fine structure. The three-dimensional structure of the large and small subunits, as well as the locations of their proteins have been determined by the techniques of immune electron microscopy and neutron diffraction. The first of these two techniques consists in preparing antibodies that bind to specific subunit proteins thus revealing their locations. The procedure involves purification of the ribosomal protein and injecting it into the bloodstream of an animal such as rabbit, to elicit the formation of antibodies (immunoglobulin G, IgG). The purified antibodies are mixed with whole ribosomes allowing them to bind to specific proteins that are exposed on the surface of the ribosome. As there are two binding sites in the antibody, therefore one antibody becomes bound to the subunit protein of two ribosomes. The subunits become associated in pairs and are visible in the

electron microscope. The analysis not only reveals the positions of the proteins, but also the three-dimensional structure of the ribosome. In the neutron diffraction technique, the subunits of the ribosomes are made to include a pair of proteins containing heavy hydrogen, deuterium. From the interference pattern made by a beam of neutrons which traverses the subunit, the distance between the two proteins (which contain deuterium) can be estimated. A number of such observations lead to mapping of the proteins.

In the three-dimensional model of an *E. coli* ribosome the smaller subunit is elongated and indented and fits into the concave surface of the larger subunit. The small subunit consists of 3 regions, a head, a base and a platform which is separated from the head by a cleft (Fig. 15.19). The larger subunit has a *central protuberance* and two projections, the larger called stalk, the shorter one ridge. The depression between the ridge and the central protuberance is called the valley.

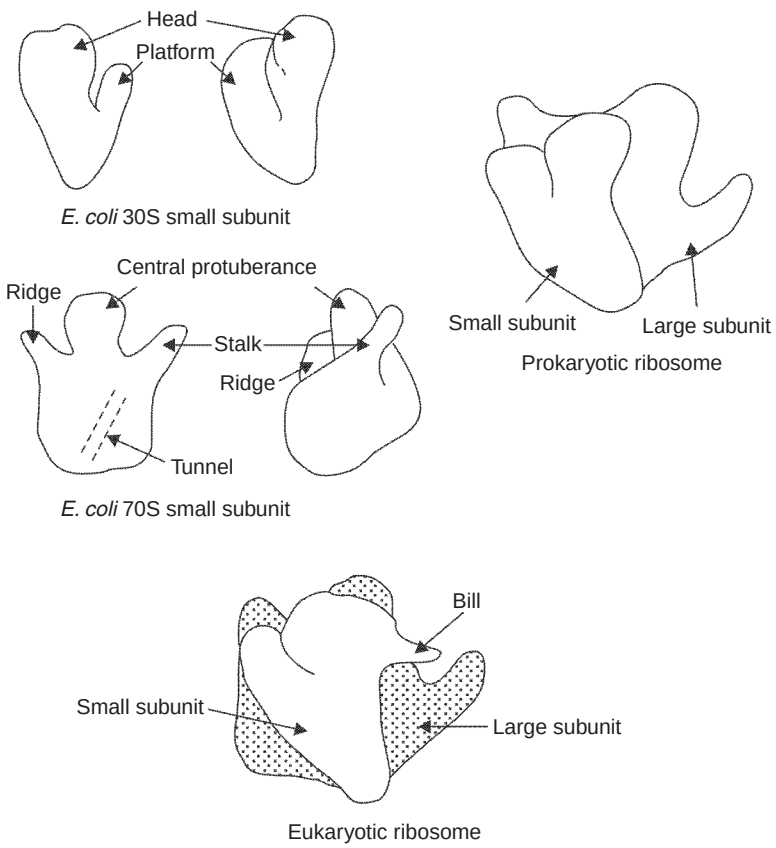


Fig. 15.19 The three-dimensional structure of the larger and smaller subunits of a ribosome.

The Ribosomal Proteins

The smaller subunit of the *E. coli* ribosome contains 21 proteins designated *S*1, *S*2 ..... *S*21; the 34 proteins of the large subunit are designated *L*1, *L*2 ..... *L*34. The proteins *S*20 and *L*26 are common to both subunits. Most of the proteins are rich in basic amino acids and their molecular weights range between 7,000 and 32,000 daltons, except *S*1 which has a molecular weight of 65,000. The various proteins have been characterised in relation to the ribosomal subunits. Some,

designated as *primary binding proteins* are attached to specific regions of the RNA strands. Six of the proteins bind to 16S RNA, 3 to 5S RNA and 11 to 23S RNA. Additional proteins are bound to the primary binding protein.

The Ribosomal RNA

About 70% of the ribosomal RNA appears double stranded due to intrachain base pairing between complementary bases. Within the subunits of the ribosome, *r*RNA is present in a highly folded form with the attached proteins. Besides being a structural component of ribosomes, *r*RNA is also involved in protein synthesis. The 3′ and of the 16S RNA has a sequence complementary to the ribosome binding site of *m*RNA (Shine and Dalgarno, 1974). The binding of 16S *r*RNA to *m*RNA allows the 30S subunit to recognise the starting end of the *m*RNA chain. The 5*S r*RNA present in the larger subunit has a sequence complementary to the sequence TCG which is present in the pseudouracil (Ψ) arm of all *t*RNAs. This allows the binding of *t*RNA to ribosome.

THE RECONSTITUTION OF RIBOSOMES

The dissociation and reconstitution of ribosomes has been a dramatic advance of recent times. When ribosomes are centrifuged in a gradient of 5M CsCl, about 30 to 40% of the protein is lost. Both the subunits dissociated into 2 inactive core particles which contain the RNA and some proteins *(core proteins)*. Simultaneously some other proteins called split proteins (SP) are released from the ribosomes. When the split proteins are added to their corresponding core particles the ribosome is reconstituted.

The reconstitution of a functional 30S subunit has been achieved by the addition of 16S *r*RNA to a mixture of proteins from the 30S subunit. Following a similar procedure the 50S subunit has also been reconstituted. Along with the subunits the whole ribosome is also assembled.

When the 30S subunit is treated with 4M urea and 2M lithium chloride it dissociates and releases all the proteins. The 16S *r*RNA can be extracted from the 30S subunit by phenol. If now the 16S RNA is put together with the 2 proteins of the same subunit, reconstitution takes place.

Ribosomal RNA constitutes about 80–90% of the total cellular RNA. Due to high RNA content, ribosomes present in the cell cytoplasm stain with basic dyes such as toluidine blue. For producing *r*RNA in such large quantities, the genes for *r*RNA are repeated a few hundred times in both prokaryotes and eukaryotes. In eukaryotes the genes for *r*RNA are transcribed in the nucleolus by RNA polymerase I; only the 5S *r*RNA present in the larger subunit is transcribed by RNA polymerase III. When transcription of ribosomal cistrons is observed in *EM*, the nascent transcripts are seen to be present in increasing length from the site of initiation to termination (Fig. 15.14); they are called christmas trees. Two successive christmas trees are separated by non-transcribed spacer segments. The primary transcripts become methylated to produce mature *r*RNAs; the spacers are degraded.

The synthesis of *r*RNA and the assembly of ribosomes take place in the nucleolus. The proteins of ribosomes are synthesized on the cytoplasmic ribosomes from where they pass into the nucleolus and become associated with *r*RNA.

In higher eukaryotes the genes for *r*RNA are clustered in the nucleolus organising regions of a few specific chromosomes. This region is recognised by the presence of a secondary constriction

at metaphase. Mouse has six such chromosomes, whereas man and chimpanzee have five. Each cluster means one transcription unit and there could be 40–50 units in a nucleolus organiser region. In *Xenopus laevis* and *D. melanogaster* the *r*RNA genes are present in a single cluster.

A mutant in *X. laevis* does not have a nucleolus (*anucleolate* mutant). In animals homozygous for this mutation there is no synthesis of 18S and 28S *r*RNA. The bobbed mutants in *D. melanogaster* are deficient in a portion of *X* chromosome which has genes for *r*RNA.

SPECIAL FEATURES OF rRNA GENES

In some lower eukaryotes ribosomal genes consist of inverted repeated sequences called *palindromic dimers* (for palindromes see Chapter 14). Blackburn and Gall (1978) found palindromic dimers in the polyploid macronucleus of *Tetrahymena*. In some slime molds, palindromic dimers are extrachromosomal and passed on to daughter cells during cell division (Vogt and Braun, 1977).

In some amphibian oocytes, the genes for *r*RNA are amplified to form a very large number of extrachromosomal nucleoli. The increase takes place during early stages of meiosis; the previous work of Gall had described a cap of dense material which was later proved to be amplified DNA by the hybridisation technique (Doggins and Gall, 1972). Amplification takes place by the rolling circle method of DNA replication (Chapter 14).

THE GENETIC CODE

The four letter language of nucleic acids is translated into the twenty letter language of the proteins viz the genetic code. The early genetic experiments indicated that each amino acid is coded for by a certain number of successive nucleotides in DNA. The best way of deciphering the genetic code therefore would be to compare the nucleotide sequence of a gene containing segment of DNA with the amino acid sequence of its specific protein. This could not be achieved until the early 1960's due to lack of knowledge about the existence of *m*RNA. Once it became established that *m*RNA carries information from DNA to a specific protein, the problem was simplified. Thereafter the study of the genetic code was approached biochemically, and consisted in analysing the relationship between the nucleotide sequence of *m*RNA and the amino acid sequence of its protein. Gradually the code words for all the amino acids were discovered and confirmed by genetic and biochemical evidence. The elucidation of the genetic code is one of the greatest scientific achievements in recent times.

The Triplet Code

The 4 code letters of DNA specifying the sequence of 20 amino acids are the 4 bases *A*, *T*, *G* and *C*. If the 4 bases are arranged in groups of 2, it is possible to have $4^2 = 16$ pairs or combinations of code words. Since 16 pairs are not enough for the 20 amino acids, the code for each amino acid must contain more than 2 bases. The 4 bases taken 3 at a time can specify $4^3 = 64$ different amino acids. The genetic code for amino acids therefore, consists of triplets of bases.

Identifying Code Words for Amino Acids

Nirenberg and Matthaei in 1961 devised a cell free system for protein synthesis. They could break open cells and utilise a mixture of the cell components for synthesising proteins. When radioactively

labelled amino acids were supplied to the cell extracts, they were detected in the newly synthesised proteins. Nirenberg and Matthaei also used artificially synthesised *m*RNA in the cell free system. When they added a synthetic *m*RNA consisting only of one base uracil (Poly *U*), the protein synthesised consisted only of one amino acid phenylalanine. Since *U* codes for polyphenylalanine, it suggests that the triplet *UUU* codes for the amino acid phenylalanine. By similar experiments it was found out that the synthetic polyribonucleotide poly *C* codes for polyproline, that is, the triplet *CCC* codes for the amino acid proline. The synthesis of polyribonucleotide (*m*RNA) such as poly *U* in the laboratory was possible due to the earlier discovery of the enzyme polynucleotide phosphorylase. This enzyme was first isolated by Ochoa (Nobel Laureate) and his colleagues from *Azobacter vinelandii.* Ochoa and his colleagues also devised a cell free system that could be used for protein synthesis. They found out that the triplet *AAA* codes for lysine.

Attempts were now made to synthesise random RNA copolymers in cell free systems using 5′-ribonucleoside diphosphates and the enzyme polynucleotide phosphorylase. In this method the enzyme catalyses the addition of the 5′-ribonuclease diphosphates to the 3′ ends; the availability of the diphosphates depending on their concentration. That is, if the concentration of ADP is 5 times that of CDP, then the copolymer synthesised will also contain 5 times the number of Cs as compared to As, but in random order. The coding triplets in such a copolymer lead to the synthesis of a polypeptide chain containing mainly lysine, threonine and proline. That is because *C* and *A* can form various possible triplets such as *CCA*, *CAC*, *AAC*, *ACA*, *CCC*, *AAA*. As the proportion of *As* is five times more than *Cs* in the copolymer, the frequency of *AAA*, *AAC*, and *CAA* codons would be correspondingly higher than the remaining codons. By use of synthetic copolymers most of the possible codons could be assigned to amino acids. However, this technique could only determine the composition of the code words. The sequence of nucleotides in the triplets that is, the spelling of a codon still remained unknown.

The Triplet Binding Assay

In 1964 Nirenberg and Leder discovered a direct method for determining the sequence of nucleotides in the code. They found that synthetic homopolymers such as poly *U* stimulated the binding of its specific aminoacyl-*t*RNA (*i.e.* Phenylalanine-*t*RNAphe) to ribosomes; simultaneously these ribosomes bind to the synthetic *m*RNA (poly *U*). No other aminoacyl-*t*RNA will bind to ribosomes and to poly *U*. Similarly when the synthetic *m*RNA consisted of poly *A* and ribosomes were added, it induced the binding of lysine-*t*RNA and not of any other aminoacyl-*t*RNA. This is called triplet *binding assay* or *t*RNA-binding technique. Later on this test was also applied to polyribonucleotide copolymers used as synthetic *m*RNA (Fig. 15.20). Nirenberg and Leder also found that triplet code had a polarity. Thus the triplet *GUU* stimulates the binding of valine *t*RNAval, whereas *UUG* binds leucine *t*RNALeu. They next proceeded to synthesise trinucleotides of known base sequence. By the triplet binding method they could determine which aminoacyl-*t*RNA was specifically bound to ribosomes in the presence of a trinucleotide of known sequence. In this way the sequence of nucleotides in a triplet was established for many of the amino acids.

H.G. Khorana and his colleagues made some very outstanding contributions in this field. Khorana was recipient of the Nobel Prize in 1968. They provided experimental evidence for codon sequences using synthetic *m*RNAs of known nucleotide sequences. They synthesised various copolymers with two, three and four kinds of bases. For example, when copolymer consisting of poly *GUA* was prepared, it had the sequence *GUAGUAGUAGUAGUAGUA* . First they synthesised

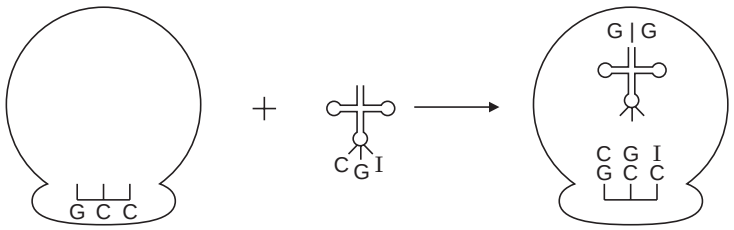
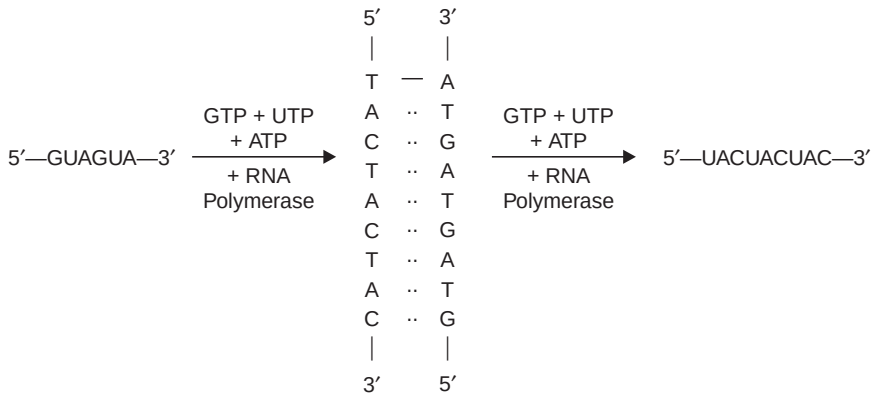


Fig. 15.20 The triplet binding technique. When triplet GCC is used as a synthetic *m*RNA, it binds alanine-*t*RNA and no other aminoacyl *t*RNA to the ribosome. The triplet GCC therefore codes for alanine.

two complementary deoxyribonucleotides each containing nine residues that is *d*(*TAC*)3 and *d*(*GTA*)3. Then they used these two oligonucleotide chains as templates on which long DNA chains were synthesized from the four deoxyribonucleoside triphosphates by the enzyme DNA polymerase I. Neither of the two oligonucleotide chains when used *alone* could serve as templates. But when both chains were present, *d*(*TAC*)3 acted as template for synthesis of poly *d*(*GTA*), and *d*(*GTA*)3 served as template for synthesis of poly *d*(*TAC*). These two long complementary chains formed a double helical structure.

After this Khorana and colleagues prepared long polyribonucleotide chains with a sequence corresponding to poly *d*(*TAC*) and poly *d*(*GTA*). This was achieved by using the poly *d*(*TAC*) and poly *d*(*GTA*) double helix as template for synthesis by RNA polymerase. The choice as to which DNA strand would be transcribed by RNA polymerase depended upon the ribonucleoside triphosphates. When *GTP*, *UTP* and *ATP* were added, the polyribonucleotide synthesised from the poly *d*(*TAC*) template strand was poly *GUA*. The other strand was not transcribed due to absence of *CTP* from incubation mixture. But when *CTP*, *UTP* and *ATP* were supplied, poly *UAC* was synthesised from the second template strand. By this method two long polyribonucleotides were synthesised each having defined repeating sequences.



By 1965 the work of Khorana, Nirenberg and Mathaei, and of Ochoa had resulted in identification of the coding triplets for all the amino acids (Table I by Crick).

Table I

First Base 5′ end	*Second Base*				*Third Base 3′ end*
	U	*C*	*A*	*G*	
U	Phe	Ser	Tyr	Cys	*U*
	Phe	Ser	Tyr	Cys	*C*
	Leu	Ser	Nonsense	Nonsense	*A*
	Leu	Ser	Nonsense	Trp	*G*
C	Leu	Pro	His	Arg	*U*
	Leu	Pro	His	Arg	*C*
	Leu	Pro	Glun	Arg	*A*
	Leu	Pro	Glun	Arg	*G*
A	Ileu	Thr	Asp	Ser	*U*
	Ileu	Thr	Asp	Ser	*C*
	Ileu	Thr	Lys	Arg	*A*
	Met	Thr	Lys	Arg	*G*
G	Val	Ala	Aspn	Gly	*U*
	Val	Ala	Aspn	Gly	*C*
	Val	Ala	Glu	Gly	*A*
	Val	Ala	Glu	Gly	*G*

Features of the Genetic Code

1. The code is commaless, that is there is no nucleotide reserved for punctuations in order to indicate the end of one codon and the beginning of the next. The triplets are read successively one after the other.

2. It is a degenerate code in the sense that most of the amino acids are coded for by more than one triplet. For example, the amino acids tyrosine, histidine, glutamic acid and some others are coded by two triplets each; and amino acids arginine, serine and leucine are each coded by 6 triplets. There are only 2 amino acids tryptophan and methionine which are coded by a single triplet.

3. Out of the 64 triplets, 3 do not code for any amino acid. These are, *UAG*, *UAA* and *UGA*. These triplets serve as signals for the termination of polypeptide chains.

4. The first two bases in the triplet specify the amino acid; the third base is less specific.

According to Crick (1966) the third base tends to "wobble" on the basis of which he proposed his *wobble hypothesis.* Crick noticed that the degeneracy of the code involves only the third base in the codon in most cases. Thus *UUU* and *UUC* both code for phenylalanine; *UUA* and *UUG* for leucine; and *GCU*, *GCC*, *GCA* and *GCG* for alanine. Crick explained the wobble hypothesis on the basis of codon-anticodon base pairing. The pairing of *m*RNA codon and *t*RNA anticodon is antiparallel. Conventionally the nucleoside at the 5′ end is written to the left. Therefore the third position of the codon is the first position of the anticodon. This leads to the following type of pairing:

1st base of tRNA	Pairs with	3rd base of mRNA
U	”	*A* or *G*
C	”	*G*
A	”	*U*
G	”	*U* or *C*
I	”	*U*, *C* or *A*

Crick proposed these pairing rules *(wobble rules)* for the codon position. According to these rules it is possible to relate more than one species of *t*RNA each with a specific anticodon, to the reading of specific codons. Thus the codons for glutamine *CAA* and *CAG* would be expected to pair with a single anticodon *UUG*. However, the hydrogen bonds between the bases of the codon and anticodon are loose enough to allow the complex to dissociate; this is required during active protein synthesis.

5. The genetic code is *universal*. It is applicable to tobacco mosaic virus, bacteriophages, *E. coli,* animals and man. In all the species tested the code triplets are identical.

GENETIC EVIDENCE FOR THE CODE

Comparative studies of mutations (changes in gene structure) and corresponding alterations in amino acid sequence of a specific protein have confirmed the validity of the genetic code. Some of these studies relate to the effects of mutation on the amino acid sequence of haemoglobin, the protein present in red blood corpuscles. A molecule of haemoglobin contains 4 polypeptide chains, two identical α chains consisting each of 141 amino acids, and two β chains, also identical and containing 146 amino acids in each. In 1956 Ingram investigated the abnormal haemoglobin (*S*) from patients having sickle cell anaemia. He found that in the β chain, the amino acid in the sixth position is valine, whereas in normal hemoglobin (*A*) it is glutamic acid. There is already a valine in the first position. The two valines in positions 1 and 6 become associated and presumably lead to a conformation that causes the erythrocyte to become sickle-shaped (Fig. 15.21). That hemoglobin *S* results from mutation in the gene can be observed from the inheritance of the protein. Individuals with sickle-cell anemia are homozygous for the gene controlling haemoglobin *S*. When such a person marries a normal person, all their children are heterozygotes as they carry one gene for haemoglobin *S* and one gene for normal haemoglobin *A*. The red blood cell of heterozygous individuals contain roughly half normal haemoglobin and half sickle-cell haemoglobin. Such persons are said to be carriers of the sickle-cell trait and lead almost normal lives. Most homozygotes die of sickle cell anaemia before the age of 30. Under conditions of low oxygen, sickle cells tend to clump together. The clumps block the capillaries so that blood supply to some regions is affected.

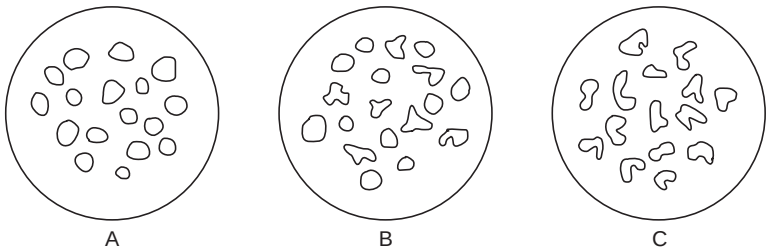


Fig. 15.21 The shape of normal and sickled red cells. *A*—in normal homozygote, *B*—a heterozygote, *C*—shows sickle-shaped red cells in a double homozygote.

Many other kinds of mutant haemoglobins with one abnormal amino acid in the β chain, some in the α chain have been investigated. The substitution of the amino acid is explained on the basis that only a single nucleotide in the coding triplet is altered. The changed triplet becomes a codon for another amino acid.

Single point mutations have been studied for their effect on the sequence of viral coat proteins. The tobacco mosaic virus consists only of RNA and protein. Mutations were induced in viral RNA by treatment with nitrous acid before infecting tobacco leaves. The lesions produced by mutant strains of virus are different from those of the normal virus. The proteins of the mutant strains were isolated by Wittmann. He could analyse 29 different amino acid substitutions in the polypeptide chains which consisted of 157 amino acids.

The genetic code has also been confirmed from the study of frameshift mutations, so called because the normal reading frame of nucleotide triplets becomes changed due to addition or deletion of a single base. Frameshift mutations produce defective proteins with altered amino acid sequence in corresponding positions thus verifying the genetic code. Crick, Brenner and their colleagues in 1961 conducted genetic experiments with such mutations and established the following: (*a*) triplets of bases code for each amino acid; (*b*) the code is non-overlapping; (*c*) the sequence of bases is read from a fixed starting point, the first nucleotide of the sequence; if the starting point is displaced by one base, the reading of the subsequent triplets becomes incorrect; (*d*) the code is degenerate.

Crick and his colleagues experimented with proflavine-induced mutations in *rII* mutants of bacteriophage *T*4 (details on *rII* locus in Chapter 22). Proflavine is an acridine mutagen that either adds or deletes one or more bases in DNA. The wild type *T*4 grows on both *E. coli* strain *B* and on *E. coli* K 12 (λ). The *rII* mutants of *T*4 do not grow on strain *K*, and produce *r* type plaques on *E. coli B*. Crick *et al.,* (1961) worked with a mutant *P* 13 renamed *FCO* in the *B* cistron of *rII* region.

Crick argued that if a mutation is due to addition of a base, then deletion of a base could revert the mutant to the wild type. They found that reversion was actually due to a second mutation involving a suppressor in the same gene. The wild type so produced was a double mutant. They could map 18 suppressors in the *B* region of mutant *FCO*. The suppressors produce an *r* mutation; the phages do not grow on *E. coli K* and produce *r* plaques on *B* strain. The point to note was that although the *rII* mutants looked alike, each had a structural defect in a different region of the *B* cistron.

The earlier experiments of Brenner had already shown, that by double infection of *E. coli* the defects in genes of two viruses can be recombined to yield progeny having both defects in one gene. The conclusion that a nucleotide had been added or deleted in a mutant was arrived at in the following way. The viruses were divided into two groups. When the defects in any two viruses of the *same* group were recombined, the resulting *rII* mutants would not grow on *E. coli K*. But if defects in viruses belonging to the 2 groups were combined, the resulting progeny viruses would grow on strain *K* and produced wild type plaques on *E. coli B*.

Crick and his colleagues assumed and later proved by further experiments that the 3 nucleotides of a single triplet or codon are determined solely by their position relative to the first nucleotide of the gene. Suppose for the sake of simplicity that the *rII* mutants form *r* plaques due to a sequence of triplets repeated as follows

CATCATCATCATCAT .....

Suppose that one *rII* mutant had a deletion of the fourth base in the sequence and that another mutant got an extra base *G* inserted between the 6th and 7th bases, the sequence would then be

C A T A T C A T C A T C A T .....
↑
C A T C A T G C A T C A T C A T .....
↑

It can be seen that due to deletion in the fourth position the triplets now read as

CAT ATC ATC ATC ATC ATC ....

The addition of *G* in the second sequence has produced the following triplets

CAT CAT GCA TCA TCA TCA ....

These triplets code for different amino acids and the resulting protein produces *r* type plaques. If by double infection bases 1 to 5 of the first mutant virus were combined with bases 6 onwards of the second, the resulting virus would have DNA of original length and sequence

C A T A T G C A T C A T C A T C A T ....

Thus after the second change the triplets *C A T* are restored. Such a sequence could produce a protein with one incorrect amino acid (coded by *ATG*). It may be assumed that if the deletion and insertion are not too far apart, they could produce a protein normal enough to enable the virus to grow on *E. coli K*. Indeed, by applying the mapping technique of Benzer, and increasing the distance between the deletion and the insertion, this assumption was found to be true.

Uncoded Amino Acids in Proteins

Some proteins contain amino acids that are not coded for by a triplet of nucleotides. For example, hydroxyproline and hydroxylysine in collagen, desmosine in elastin, *N*-methyllysine in some muscle proteins, and a few others. Such amino acids arise due to post-translational modification of amino acids that have been inserted into the polypeptide after instructions from their specific triplets.

TRANSLATION AND PROTEIN BIOSYNTHESIS

The translation of the encoded genetic message leads to synthesis of polypeptide chains of correct sequence and length. The process involves *m*RNA, ribosomes, 20 kinds of amino acids and their specific *t*RNAs. The amino acids are inserted into a growing polypeptide chain on a ribosome that moves along the *m*RNA in a 5′ to 3′ direction. Each amino acid has an α-amino group at one end and an α-carboxyl group at the other; the chain thus has an amino terminal and a carboxy terminal.

A *m*RNA may be associated with several ribosomes (polysomes) in a cell actively engaged in protein synthesis. The process takes place in 4 major steps namely activation of amino acids, chain initiation, elongation and termination. Each step is controlled by specific enzymes and cofactors.

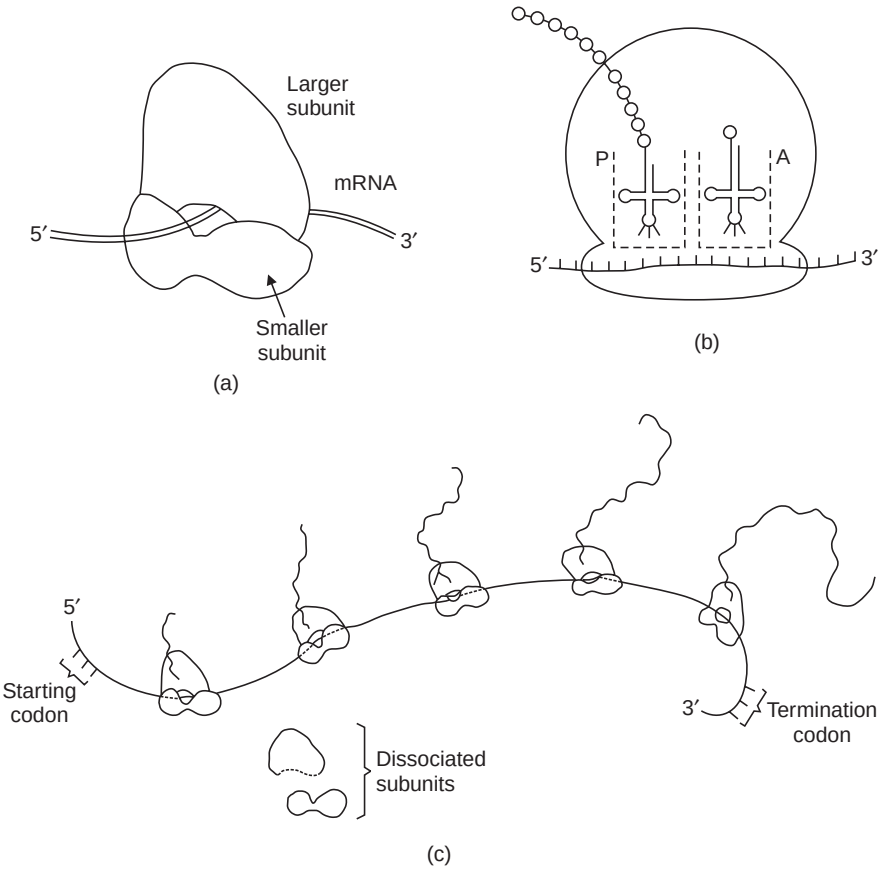


Fig. 15.22 (*a*)—three dimensional view of a ribosome with *m*RNA passing through the groove; (*b*)—ribosome showing *t*RNA binding sites and (*c*)—polysome containing 5 ribosomes on a *m*RNA strand. The growing chains are of increasing lengths.

1. ACTIVATION OF AMINO ACIDS

The amino acids are activated by a class of enzymes known as amino acid activating enzymes, or specifically *aminoacyl tRNA synthetases* of which there is one specific for each amino acid and its *t*RNA. The overall reaction is as follows:

$$\text{ATP + Amino acid + } t\text{RNA} \xrightarrow{Mg^{++}} \text{Aminoacyl-}t\text{RNA + AMP + PPi}$$

The activating enzymes or synthetases range in molecular weight between 100,000 to 240,000; some have a single chain, others are multichain enzymes. All of them require Mg^{++} for their activity.

The amino acid is then attached (esterified) to the terminal adenosine residue at the 3′ end of its *t*RNA in 2 steps. In the first the amino acid is activated when ATP reacts with the amino acid catalysed by aminoacyl synthetase to form an intermediate aminoacyl adenylic acid-AMP-synthetase complex and pyrophosphate is released. In the second stage the aminoacyl

group is transferred from AMP complex to the 3′ end of its *t*RNA. Following this adenylic acid and aminoacyl-*t*RNA are released.

ATP + amino acid $\rightleftharpoons$ aminoacyl adenylic acid + pyrophosphate aminoacyl adenylic acid + *t*RNA aminoacyl-*t*RNA + adenylic acid

The aminoacyl now becomes esterified by its carboxyl group to the free hydroxyl group at the second position of the terminal adenylic residue at the 3′ end (*C-C-A*) of *t*RNA (Fig. 15.23)

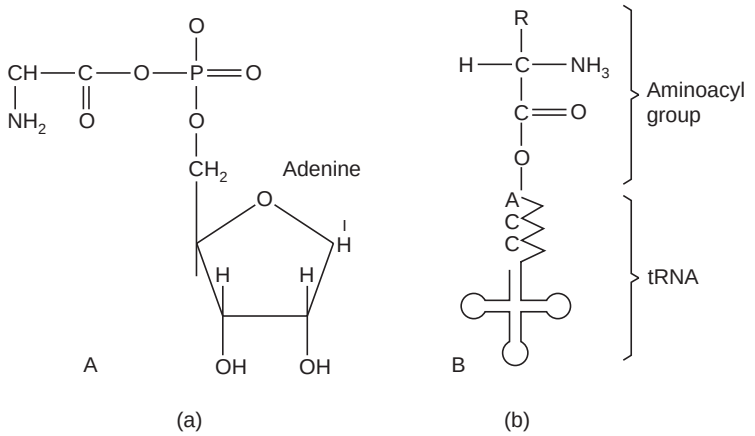


Fig. 15.23 *A*—General structure of aminoacyl adenylate and *B*—An aminoacyl *t*RNA.

The aminoacyl-*t*RNA synthetases are highly specific for both the amino acid and its *t*RNA. Consequently, these enzymes must possess at least two binding sites, one for the amino acid, the second for its corresponding *t*RNA molecule. It is obvious that each aminoacyl-*t*RNA synthetase is a highly specific enzyme which can select only one amino acid out of 20; and then it selects the correct *t*RNA for that particular amino acid and no other. How does an aminoacyl-*t*RNA synthetase recognize a *t*RNA? The genetic and biochemical evidence points to a general condition whereby a *t*RNA and an aminoacyl-*t*RNA synthetase 'fit' each other. In addition, a synthetase might be able to recognise a few recognition points of two or three bases in the *t*RNA structure.

Events on the Ribosome

As mentioned earlier, the *m*RNA after its formation on DNA, migrates out of the nucleus to the cytoplasm where it becomes associated with the 30S subunit of the ribosome. The *t*RNA molecules loaded with their amino acids (called charged *t*RNAs) also reach the ribosomes. The 50S subunit of the ribosome has two binding sites for two *t*RNA molecules called *peptidyl* (*P*) and *aminoacyl* (*A*) sites. It is here that the amino acid molecules will be polymerised into a polypeptide chain. In bacteria about 10 amino acids are added to the chain per second, in animals only about 2 per second. The *m*RNA binds to the 30*S* subunit at the ribosome-binding site. Within this site are nucleotide sequences which ensure the correct lining up of *m*RNA molecules on the ribosomal surface. Every *m*RNA molecule thus has one *ribosomal-binding site* for each of its independently synthesized polypeptide chains (Fig. 15.22*a*).

A ribosome contains only one nascent polypeptide chain bound at its carboxyl end to the *t*RNA molecule which is attached to its specific site on the ribosome. When the next amino acid is brought by a second *t*RNA to be linked to the carboxyl end of the chain by peptide bond formation, the first *t*RNA is released. The ribosome moves over the *m*RNA and the next triplet of nucleotides comes to lie in position for the next amino acid (Fig. 15.23).

The successive ribosomes of such a polysome carry polypeptide chains at successive stages of completion (Fig. 15.22*C*). The 5′ end of the *m*RNA is the first to become attached to the 30S subunit of a ribosome. The number of ribosomes in polysomes is variable. There may be 5 or 6 as in the case of each polypeptide chain of hemoglobin containing about 150 amino acids. Larger polypeptide chains of about 500 amino acids may have 20 or more ribosomes in a polysome. In *E. coli* up to 40 ribosomes have been observed in a polysome.

In both prokaryotic and eukaryotic cells, translation begins with the amino acid methionine encoded by the nucleotide triplet *AUG*. Bacteria may initiate translation using alternative initiation codons such as *GUG*. When it is present at the beginning of a polypeptide chain, *GUG* codes for methionine. *GUG* normally encodes valine. In bacteria, polypeptide synthesis is initiated with a modified methionine, that is *N*-formylmethionine, whereas in eukaryotes, unmodified methionine can initiate protein synthesis. Mitochondria and chloroplasts whose ribosomes are similar to those of prokaryotes use formylated methionine to initiate translation.

Prokaryotes and eukaryotes have different signals that identify initiation codons. The initiation codons in bacterial *m*RNA are preceded by a specific Shine-Dalgarno sequence that aligns the *m*RNA on the ribosome for translation. Alignment takes place through base pairing with a complementary sequence near the 3′ terminus of 16S *r*RNA in the small subunit of the ribosome. The base pairing mechanism allows bacterial ribosomes to initiate transcription not only at the 5′ end of the *m*RNA, but also at the multiple internal initiation sites of polycistronic messages. In eukaruotes in contrast, ribosomes recognise *m*RNAs by binding to the 7-methylguanosine cap at the 5′ terminus. The ribosomes then traverse the *m*RNA downstream of the 5′ cap until they encounter the AUG initiation codon. Eukaryotic *m*RNAs do not have the Shine-Dalgarno sequence, and translation is initiated at a site determined by the ribosome traversing the *m*RNA from the 5′ terminus.

The First Step in Translation

The first step in translation in bacteria involves the binding of three initiation factors, *IF*-1, *IF*-2 and *IF*-3 to the 30S small subunit of the ribosome (Figs. 15.24, 15.25). The *m*RNA and the *N*-formylmethionyl *t*RNA join the complex. *IF*-2 recognises the initiator *t*RNA. *IF*-3 is released allowing a large 50*S* ribosomal subunit to associate with the complex. The association triggers release of *IF*-1 and *IF*-2, resulting in the formation of a 70S initiation complex having *m*RNA and initiator *t*RNA bound to the ribosome. The complex is ready to begin peptide bond formation. Initiation in eukaryotes is more complicated and requires at least 10 proteins designated *eIF*-1 to *eIF*-10 (eukaryotic initiation factors).

2. INITIATION OF POLYPEPTIDE CHAINS

In *E. coli* and other prokaryotes the first amino acid in a polypeptide chain is methionine with a formyl group (CHO) attached to the free amino group. Initiation of protein synthesis therefore requires *N*-formylmethionyl-*t*RNA (designated *f*Met-*t*RNAf), and not methionyl-*t*RNA. The transformylation takes place when a formyl group is transferred from N^{10}-formyltetrahydrofolate

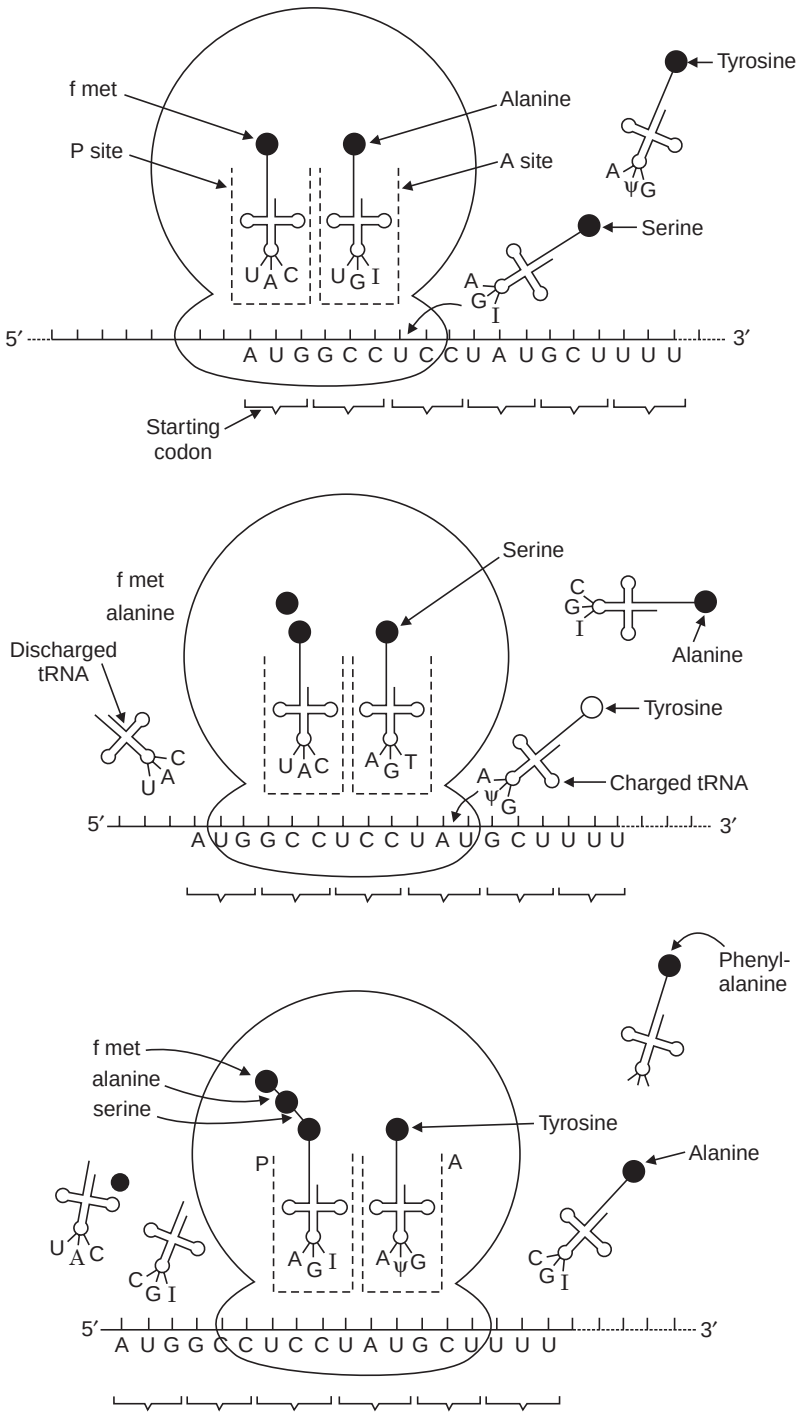


Fig. 15.24 Generalised pattern for the synthesis of a polypeptide chain.

to the α-amino group of methionyl-*t*RNA. Thus the amino acid methionine in *E. coli* has *t*RNAs of two distinct species only one of which can initiate chain synthesis: methionyl $t\text{RNA}_{\text{Met}}$ and

methionyl-tRNAf$_{Met}$. Only the methionyl-tRNA of second species (tRNAf$_{Met}$) can accept the N-formyl group and become formylated. Both species of tRNA have the same anticodon sequence *UAC*, but have slight difference in the remaining nucleotide sequence. Formylation takes place only after the amino acid has become attached to the tRNA molecule, and is catalysed by the enzyme *transformylase*. The reaction is N^{10}-Formyl-tetrahydrofolate + Met-tRNAf → tetrahydrofolate + fMet + tRNAf. The other species of tRNAf$_{Met}$-tRNA$_{Met}$ cannot be formylated nor is recognised by transformylase.

The starting codon *AUG* for methionine is read both by tRNA$_{fMet}$ and by tRNA$_{Met}$. The triplet *GUG* which codes for valine is read also by tRNA$_{fMet}$ but not by tRNA$_{Met}$. Thus both *AUG* and *GUG* are starting codons and can lead to the placement of formylmethionine in the first position of the polypeptide chain. When *AUG* is present in an internal position in mRNA, a methionine is inserted in an internal position in the chain; when *GUG* is present internally in mRNA it codes for valine.

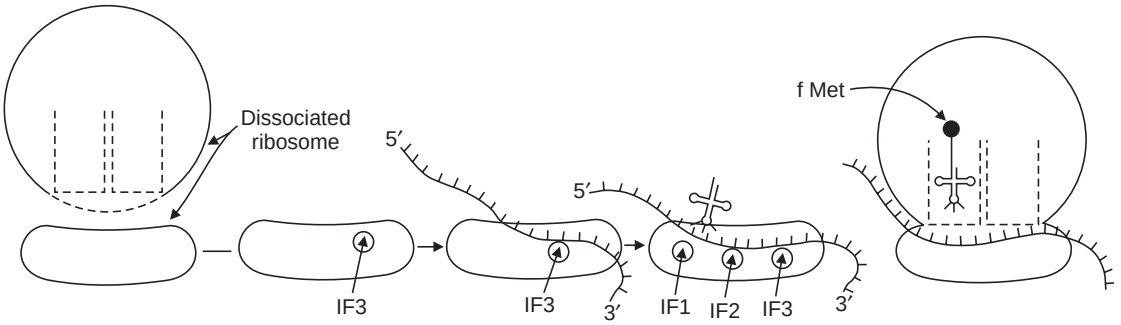


Fig. 15.25 Stages in the formation of the initiation complex.

It now appears likely that not only in bacteria but also in mitochondria of eukaryotic cells, fMet-tRNAf initiates protein synthesis.

3. CHAIN ELONGATION

Elongation begins when the initiating aminocyl-tRNA is positioned on the peptidyl site (*P*) opposite the starting codon on mRNA and aminocyl site (*A*) is free. Elongation is achieved in 3 steps as follows:

(*a*) *Binding at site A:* The aminoacyl-tRNA corresponding to the first triplet after the staring codon is bound to the *aminoacyl site* (*A*) in the following way. First the aminocyl tRNAx (x denotes any of the 20 amino acids) binds to a specific protein, an *elongation factor* (*EF–T*). Actually *EF–T* contains 2 subunits namely *EF–*T_8 and *EF–*T_u. Now *EF–T* combines with *GTP* to form an *EF–*T_u*–GTP* complex and *EF–*T_8 is released. The *EF–Tu–GTP* complex now binds the aminoacyl–tRNAx to form *EF–*T_u*–GTP*–aminoacyl–tRNAx complex. This complex now binds to the ribosome in such a way that the aminoacyl-tRNAx is placed on site *A*, while the anticodon triplet becomes hydrogen bonded to its specific codon in mRNA. In eukaryotes the elongation factors are designaged *EF*–1 and *EF*–2 and they act differently.

(*b*) *Peptide bond formation:* When the *P* site is occupied by the $_f$Met-tRNA$_f$ and the *A* site by the aminoacyl-tRNAx, a peptide bond is formed between the amino group of the amino-acyl-tRNAx and the carboxyl carbon of the $_f$Met-tRNA. The enzyme *peptidyl transferase* present in the 50S subunit of the ribosome catalyses this reaction (Fig. 15.26). The result is a dipeptide attached

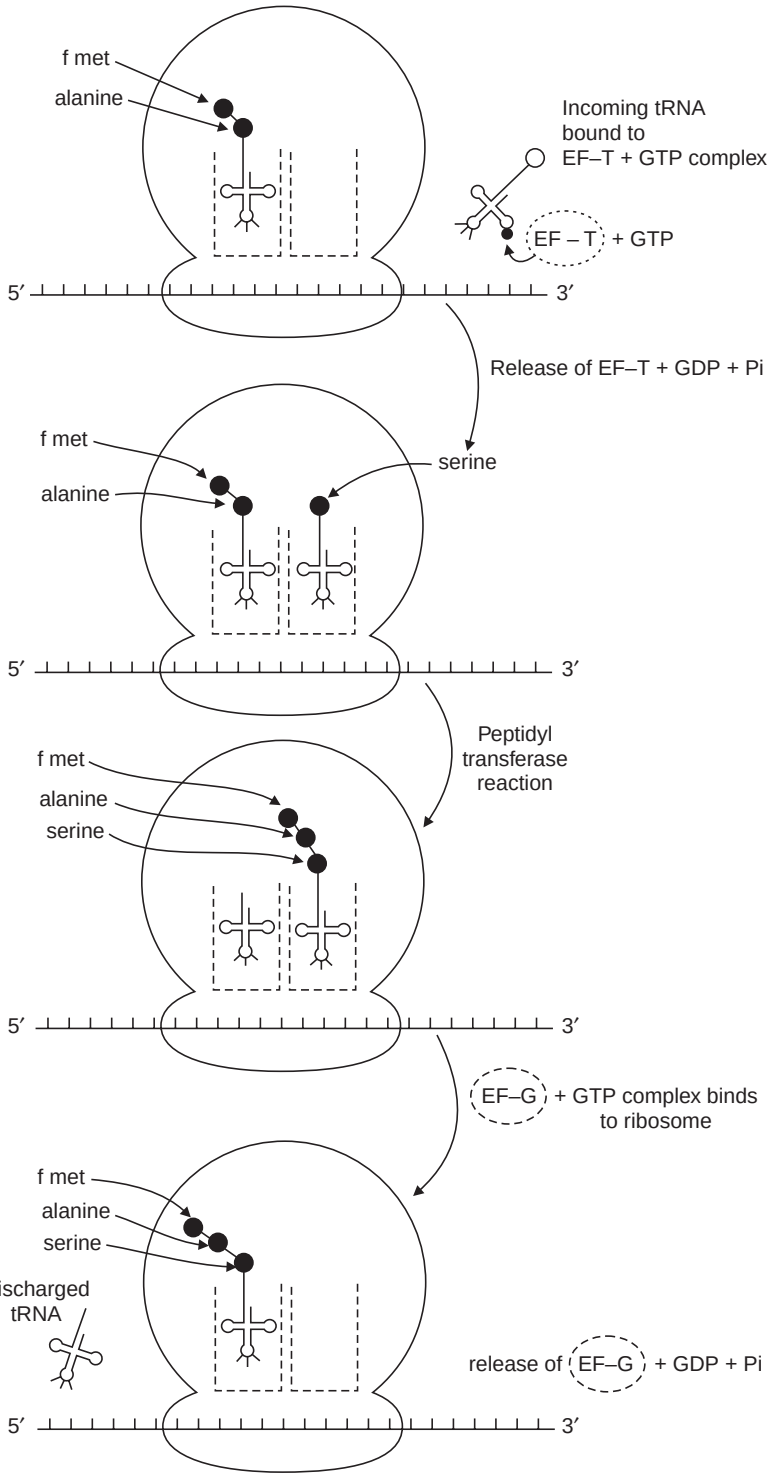


Fig. 15.26 Steps in the elongation of the polypeptide chain.

to the *t*RNAx at site *A*; the *t*RNA at the *P* site has discharged its amino acid (*f*met) and is itself empty, but is bound to *P* site.

(*c*) *Translocation:* While the *t*RNA carrying the dipeptide is still bound to the *A* site opposite the *m*RNA codon of the newly added amino acid *x*, the ribosome moves to the next codon on *m*RNA. Simultaneously, the *t*RNA with dipeptide shifts from the *A* site to *P* site. This movement is called translocation and requires a specific protein, an elongation factor *G*(*EF–G*) and *GTP*. The *GTP* first binds to *EF–G* to form a complex which binds to the ribosome. *GTP* hydrolyses to form *GDP* and *P*. The energy released during hydrolysis is utilised for a conformational change which causes the ribosome to move to the next codon on *m*RNA carrying the chain from the *A* site to the *P* site. This brings the *A* site now opposite to a new codon, and a new aminoacyl-*t*RNA will now come and occupy the *A* site. After translocation *EF–G* dissociates from the ribosome.

4. CHAIN TERMINATION

If there was no signal to stop chain growth, polypeptide chain would continue to elongate indefinitely. This is not so. Each one of 3 special termination codons in *m*RNA (*UAA*, *UAG*, *UGA*) can signal the termination of chain growth. These codons are read off by specific proteins, release factors R_1 (responding to *UAA* and *UAG*) and R_2 (responding to *UAA* and *UGA*).

When the termination codon is reached the polypeptide is still attached to the *t*RNA which is on the *A* site of the 50S subunit. For releasing the chain from the *t*RNA, release factors R^1 and R^2 are required (Fig. 15.27). They bind to the ribosome and cause the *t*RNA at A site to shift to the *P* site. The enzyme peptidyl transferase hydrolyses the ester bond between the chain and the *t*RNA, aided by the bound releasing factors. When the polypeptide is released, the last *t*RNA and the *m*RNA are set free. This is then repeated by another ribosome which would associate with the *m*RNA as soon as the initiation point is free.

DESTINATIONS OF PROTEINS

Proteins Synthesised on Free Ribosomes: Ribosomes involved in protein synthesis could lie free in the cytoplasm or are bound to membranes of endoplasmic reticulum. **Free ribosomes** synthesise protein without a signal sequence, and continue synthesis until the polypeptide is completed. Such a protein may assume its three-dimensional conformation with the help of chaperones present in the cytosol. Some of the important destinations for proteins synthesised on free ribosomes are:

1. *The Cytosol*. These are proteins with housekeeping responsibilities, take part in metabolic activities, constituting various enzymes for the metabolic machinery. Include tubulins for generating microtubules, actin for microfilaments. They are released into the cytosol for performing their function.
2. *The Nucleus*. Some proteins move from the cytosol into the interior of the nucleus. These include the histones, transcriptional factors and ribosomal proteins. They are targeted to the nucleus by their **nuclear localisation sequence**, consisting of about 7 to 41 amino acids.

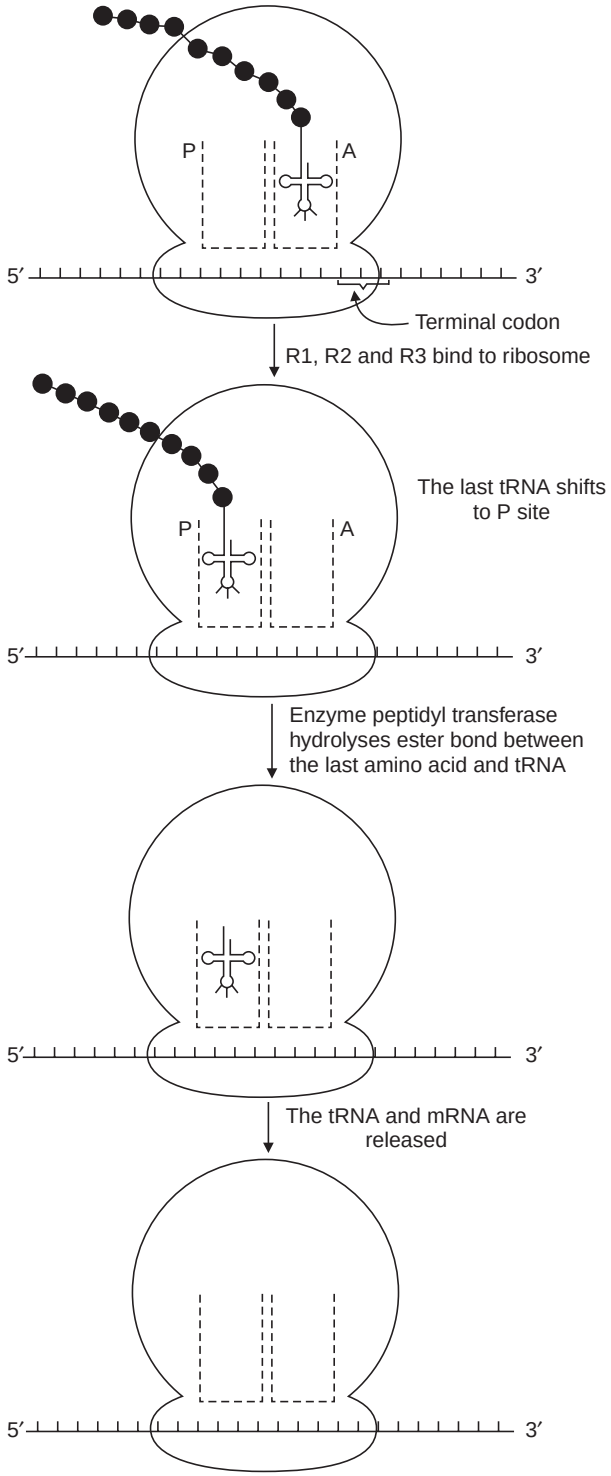


Fig. 15.27 Steps in the termination of the protein chain.

3. *Mitochondria*. Even though mitochondria have their own genome, ribosomes, and protein synthesising machinery, most of the proteins used by mitochondria are encoded by nuclear genes, synthesised in the cytosol on free ribosomes and transported into mitochondria. These proteins possess a characteristic signal sequence which is recognized and bound by a chaperone called **mitochondrial stimulation factor** (**MSF**). MSF targets the protein to a receptor embedded in the outer membrane of the mitochondrion. Other factors and receptors direct the proteins through the intermembrane space to the inner mitochondrial membrane and the matrix.
4. *Chloroplasts*. Like mitochondria, chloroplasts also have their own genome and protein synthesising machinery. Yet, most of the proteins used in chloroplasts are encoded by the host cell genome. These proteins are synthesised on free cytosolic ribosomes, then transported into the chloroplast. These proteins have a characteristic **transit sequence** by which they are recognised, and chaperones are also needed for their transport to their final destination, which could be the chloroplast inner membrane, thylakoid membrane or the stroma.
5. *Peroxisomes*. Proteins synthesised for peroxisomes carry a **peroxisomal targeting signal** (**PTS**) that binds to a receptor molecule, which internalises the protein into the peroxisome. The signal sequence then returns to carry another protein load to the peroxisome. Two peroxisome targeting signals have been identified, each having its own specific receptor to take it to the peroxisome.

Proteins Synthesised on Ribosomes Bound to Endoplasmic Reticulum: Proteins used in membrane-bound organelles such as lysosomes, vesicles and plasma membrane on cell boundary, are synthesised on ribosomes bound to the membranes of endoplasmic reticulum (*ER*). These are called **secretory proteins**. The synthesis of these proteins actually begins on free cytosolic ribosomes that are in no way different from ribosomes bound to *ER* membranes. During synthesis on free ribosomes, the polypeptide chain is translocated (cotranslational translocation) to *ER*. The free ribosomes engaged in synthesis of proteins destined for secretion, are targeted to *ER* by a **signal sequence** at the amino terminus of the growing polypeptide chain. The signal sequences contain about 20 amino acids, some of which being hydrophobic enable them to cross the *ER* membrane into the lumen of *ER*.

The process of cotranslational translocation takes place as follows. As the polypeptide chain is emerging from the free cytoplasmic ribosome, the signal sequences are recognised and become bound to a **signal recognition particle** (**SRP**). The binding produces a complex consisting of ribosome, SRP and growing polypeptide chain, and further translation is inhibited. The entire complex is targeted to the *ER* and binding takes place with the SRP receptor on *ER* membrane. SRP is released. The signal sequence is inserted into a membrane channel in ER and translocates the polypeptide chain into the lumen of *ER*.

The proteins then begin their journey in the **secretory pathway** starting from lumen of *ER*. They become included in **secretory vesicles** that are budded off from *ER* membrane. They reach the Golgi apparatus, travel through the stacked Golgi membranes, and are then budded out in vesicles. These vesicles target the organelles that are final destinations of the protein, fuse with the organelle membrane and deliver the protein. The proteins undergo their required modifications (for example glycosylation) during their journey in the secretory pathway.

OPEN READING FRAME

As already mentioned, of the 64 codons, only 61 specify amino acids, called **sense codons**. The remaining three UAG, UAA and UGA do not specify any amino acid, and no *t*RNAs in normal cells carry the appropriate anticodons. These 3 codons are the stop codons or **nonsense codons**, or **chain terminating codons**. They are used singly or in tandem (end-to-end) groups (UAG UAA for example) to specify end of translation of a polypeptide chain. Thus, when we read a particular *m*RNA sequence, we look for a stop codon located at a multiple of three nucleotides (triplets) – in the same reading frame – beginning with the UAG start codon to determine where the amino-acid-coding sequence for the polypeptide ends. This is called an **open reading frame** or **ORF**. Open reading frames can be detected in a genome by identifying potential protein coding regions by computer analysis of the DNA sequence. ORFs are long stretches of nucleotide sequence that can encode polypeptides because they do not contain any of the three chain-terminating codons (UAA, UAG and UGA).

FUNCTIONAL PROTEINS

The genetic information contained in the nucleotide sequence of DNA has been translated into the amino acid sequence of the protein chain. Proteins have the primary responsibility to perform activities as directed by their information content. Among the several thousand different proteins in a cell, some are structural components of cells and tissues, acting in the transport and storage of small molecules, for example, the transport of oxygen by hemoglobin; transmission of information between cells, example protein hormones; providing defense against infection by antibodies; to act as enzymes and catalyse chemical reactions in biological systems. Functional proteins are far more complex than linear chains of amino acids. To produce functional proteins, polypeptide chains have to undergo modifications, such as folding into a specific three-dimensional conformation. In many cases multiple polypeptide chains might associate into a functional complex. Additional modifications such as cleavage and covalent attachment of carbohydrates and lipids, may be critical for correct functioning and localisation of the protein within the cell.

Protein folding and Chaperones

The experiments of Anfinsen and collaborators half a century ago, demonstrated that information for protein folding into a three-dimensional conformation is provided by its amino acid sequence. Working with the enzyme ribonuclease, they found that denatured proteins could spontaneously refold into an active conformation. This implied that the information required to specify the three-dimensional conformation of a protein is contained in its primary amino acid sequence. Thus, the three-dimensional structure of a protein corresponds to its thermodynamically most stable conformation, determined by interactions between its constituent amino acids. Protein folding thus appeared to be a self-assembly process that did not require additional cellular factors. Since the order of nucleotides in DNA specifies the sequence of amino acids in the polypeptide chain, it follows that the nucleotide sequence of a gene contains all the information needed to determine the three-dimensional structure of its protein product. Recent studies have however, shown that protein folding is a complex process and requires mediation by additional cellular proteins.

Proteins which act as catalysts that facilitate folding of other proteins are called molecular **chaperones**. Chaperones catalyse protein folding by assisting the self-assembly process. Chaperones do not convey information required for folding, and the folded conformation is determined

solely by the amino acid sequence of the protein. They appear to function by binding to unfolded or partially folded polypeptides that are intermediates along the pathway to the final folded state. In the absence of chaperones, unfolded or partially folded polypeptide chains remain unstable within the cell, would either fold incorrectly or aggregate into insoluble complexes. In some cases, chaperones bind to nascent polypeptide chains that are still being translated on ribosomes, thus preventing incorrect folding. In the case of cytosolic proteins that are transported into mitochondria, chaperones within the mitochondrion facilitate transfer of the polypeptide chain across the membrane and its subsequent folding within the organelle. In addition, chaperones are involved in the assembly of proteins that comprise multiple polypeptide chains, in the assembly of macromolecular structures such as nucleoplasmin, and in the regulation of protein degradation.

Many of the molecular chaperones have been identified to be **heat-shock proteins (Hsp)**, that are expressed in cells subjected to high temperatures or other forms of environmental stress. Heat-shock proteins are thought to stabilise the refolding of proteins that have been partially denatured in response to elevated temperature. Many of the heat-shock proteins are also expressed in cells under normal conditions and are believed to be essential for cellular functions. These proteins act as molecular chaperones both under normal conditions and in cells under environmental stress.

Enzymes in Protein Folding

At least two types of enzymes are known to catalyse protein folding by breaking and reforming covalent bonds. **Protein disulphide isomerase** catalyses the breakage and reformation of disulphide bonds between cysteine residues in many proteins. Disulphide bonds are generally restricted to secreted proteins and some membrane proteins that are synthesised in ribosomes bound to endoplasmic reticulum, because the cytosol contains reducing agents that maintain cysteine residues in their reduced, -SH form, thereby preventing the formation of disulphide, *S-S* linkages.

Peptidyl prolyl isomerase is the second enzyme that plays a role in protein folding by catalysing the isomerisation of peptide bonds that involve proline residues. In contrast with peptide bonds between most amino acids that are almost always in the *trans* form, proline is unusual in that the equilibrium between the *cis* and *trans* conformations of peptide bonds that precede proline residues is only slightly in favour of the *trans* form. Isomerisation between the *cis* and *trans* configurations of prolyl peptide bonds is catalysed by peptidyl prolyl isomerase.

Protein Cleavage

Cleavage, also called **proteolysis**, is an important step in maturation of many proteins. During translation the initiator methionine from the amino acid terminal of the newly formed chain is removed by cleavage, soon after the amino terminus of the polypeptide chain emerges from the ribosome. After cleavage, chemical groups such as acetyl groups or fatty acid chains are added to the amino terminal residues.

The amino terminus of secreted proteins undergoes proteolytic transformation in order to facilitate translocation of these proteins across membranes in both prokaryotes and eukaryotes. The same is true of proteins destined for incorporation into the plasma membrane, lysosomes, mitochondria and chloroplasts. These proteins contain **signal sequences** about 20 amino acids long at their amino terminus that direct proteins to their destinations, and are removed by proteolytic cleavage after crossing the membrane. The signal sequence contains mostly hydrophobic amino acids which are able to cross the hydrophobic interior of the membrane. The remainder of the polypeptide chain passes through a channel in the membrane. After traversing the membrane,

the signal sequence is cleaved by a specific membrane protease (**signal peptidase**) and the mature protein is released.

The formation of active enzymes and hormones also takes place by cleavage of larger precursors. The precursor of insulin is a longer polypeptide from which insulin is derived by two cleavages. The initial precursor (preproinsulin) contains an amino terminal signal sequence that directs the polypeptide chain to the endoplasmic reticulum. Removal of the signal sequence yields the second precursor, proinsulin which is converted to insulin. The two chains of insulin are held together by disulphide bonds following proteolytic removal of an internal peptide bond. The proteins of many animal viruses are derived from larger precursors. In replication of HIV, a virus-encoded protease cleaves precursor polypeptides to form the virus structural proteins. Since the HIV protease (as also reverse transcriptase) has a central role in virus multiplication, it is an important target for development of drugs (protease inhibitors) for treatment of AIDS.

Glycosylation

The addition of carbohydrates to proteins to form **glycoproteins** is called **glycosylation.** Most glycoproteins are secreted or incorporated into the plasma membrane. Some nuclear and cytosolic proteins are also glycosylated. Glycosylation is either *N*-linked or *O*-linked. In *N*-linked glycoproteins, the carbohydrate is attached to the nitrogen atom in the side chain of asparagines. In *O*-linked glycoproteins, the oxygen atom in the side chain of serine or threonine is the site of carbohydrate attachment. Glycosylation of proteins takes place while they are traversing the endoplasmic reticulum-Golgi network.

Modification by Lipids

The attachment of lipids to some proteins in eukaryotic cells facilitates their incorporation into the plasma membrane. A fatty acid, myristic acid (a 14 carbon fatty acid) is attached to the amino terminus (*N*-terminal glycine residue) of the polypeptide chain during translation, called **N-myristoylation**. The initiator methionine is first removed by proteolysis before fatty acid addition. Proteins modified by *N*-myristoylation are usually associated with the inner face of the plasma membrane. When lipids are attached to the side chains of cysteine, serine or threonine it is called **prenylation**. The prenyl groups of lipids are attached to the sulfur atoms in the side chains of cysteine. Importantly, many plasma membrane-associated proteins that are involved in cell growth and differentiation are prenylated. For example, the ras oncogene proteins responsible for uncontrolled growth of many human cancers. In another type of fatty acid modification, palmitic acid (a 16-carbon fatty acid) is added to sulphur atoms of the side chains of internal cysteine residues, called **palmitoylation**. Proteins modified by palmitoylation associate with the cytosolic face of the plasma membrane. Lipids linked to oligosaccharides, that is **glycolipids** are added to the *C*-terminal carboxyl groups of some proteins, where they serve as anchors for attachment of proteins to the external face of the plasma membrane. The glycolipids attached to these proteins contain phosphatidylinositol, hence they are referred to as **glycosylphosphatidylinositol** or **GPI anchors.**

Protein Degradation

Besides synthesis, degradation of proteins takes place that maintains appropriate levels of protein in cells. The half-life of proteins within cells could vary from minutes to several days. Some of the rapidly degrading proteins include the transcription factors whose rapid turnover allows their

levels to change quickly in response to external stimuli, and to function as regulatory molecules. Some proteins are rapidly degraded in response to specific signals and play a role in regulation of intracellular enzyme activity. Undesirable proteins are also degraded. Eukaryotic cells have two major pathways for degradation of proteins, the ubiquitin-proteasome pathway and the lysosomal proteolytic pathway.

The Ubiquitin-Proteasome Pathway

The ubiquitin polypeptide contains 76 amino acids. Nuclear and cytosolic proteins destined for proteolyses bind to ubiquitin to the amino group of the side chain of a lysine residue. The addition of further ubiquitins forms a multiubiquitin chain which is recognised and degraded by a large protease complex called the **proteasome**. Ubiquitin-mediated proteolysis is called ubiquitination, the process takes place in several steps and requires energy in the form of ATP. Ubiquitin is then released and can be reused in another cycle.

Lysosome Mediated Proteolysis

Lysosomes are single membrane-bound organelles in eukaryotic cells that contain several digestive enzymes, all of which are active at low pH not higher than 4.5. The acidic environment within lysosomes is maintained by proton pumps in the lysosomal membrane. Lysosomes play a significant role in degrading both externally acquired proteins (by endocytosis) as well as in turnover of organelles and cytosolic proteins. Proteins targeted for lysosomes become enclosed in **autophagic vesicles** derived from membranes of the endoplasmic reticulum. The vesicles fuse with the membrane of the lysosome and release the protein that is acted upon by the digestive enzymes and degraded.

QUESTIONS

1. How does the DNA of $\phi X174$ differ from that of man? Why is the enzyme DNA dependent RNA polymerase so-named?
2. If the sequence of nucleotides in a transcribing gene is *CATGTAGC*, what would be the nucleotide sequence of its transcript?
3. What makes oocytes of amphibians a favourite material for study of gene transcription? State three reasons.
4. Draw diagrams to bring out the differences between the lampbrush chromosomes of the toad *Xenopus laevis* as seen in the light microscope, and the lampbrush state in the cell of a mouse embryo as seen in the electron microscope.
5. The sigma factor is not firmly bound to the other subunits of the RNA polymerase enzyme. What purpose does this serve?
6. Match correctly the discoveries with the investigators below:

Ochoa	visualisation of transcription
H.G. Khorana	colinearity of genes and enzymes
Miller and Beatty	the triplet code
Beadle	reverse transcriptase
Temin	nutritional mutants in *Neurospora*
Yanofsky	RNA synthesis

SELECTED READINGS

Adesnik, M. *et al.,* 1972. Evidence that *m*RNA Molecules (except histone *m*RNA) contain poly (*A*) sequences and that the poly (*A*) has a nuclear function. *J. Mol. Biol.* **71** : 21.

Angelier, N. and Lacroix, J.C. 1975. Complexes de transcription d' origines nucleolaire et chromosomique *d*' ovocytes de Pleurodeles waltlii et.P. poireti. *Chromosoma* **51** : 323.

Arnheim, N. and Southern, E.M. 1977. Heterogeneity of the Ribosomal Genes in Mice and Men. *Cell* **11** : 363.

Beerman, W. 1965. Differentiation at the Level of the Chromosomes. *In* Cell Differentiation and Morphogenesis. Ed. Beerman, W. North Holland, Amsterdam.

Beerman, *W.* 1972. Developmental Studies on Giant Chromosomes. Springer-Verlag, Berlin.

Benzer, S. 1955. Fine Structure of a Genetic region in Bacteriophage. *Proc. Natl. Acad. Sci. U.S.* **41** : 344.

Beyer, A.L. *et al.,* 1979. Transcriptional Units in Eukaryotic Chromosomes. *In* Molecular Genetics, Part III. Chromosome Structure. Ed. Taylor, J. H. Academic Press, New York.

Bonner, J.J. and Pardue, M.L. 1977. Polytene Chromosome Puffing and *in situ* Hybridisation Measure Different Aspects of RNA Metabolism. *Cell* **12** : 227.

Busby, S. and Ebright, R.H. 1994. Promoter Structure, Promoter Recognition, and Transcription Activation in Prokaryotes. *Cell* **79** : 743–746.

Fiers, W. and Grosjeau, H. 1979. On Codon Usage. *Nature* **277** : 328.

Gilbert, W. *et al.,* 1974. Sequences of Controlling Regions of the Lactose Operon. *Cold Spring Harb. Symp. Quant. Biol.* **38** : 845.

Goodpasture, C. and Bloom, S.E. 1975. Visualisation of Nucleolar Organiser Regions in Mammalian Chromosomes Using Silver Stain. *Chromosoma* **53** : 37.

Gray, N.K. and Wickens, M. 1998. Control of Translation Initiation in Animals. *Annu Rev. Cell Dev. Biol.* **14** : 399–458.

Green, R. and Noller, H.F. 1997. Ribosomes and translation. *Annu. Rev. Biochem*. 66: 679–716.

Hahn, S. 1998. The Role of TAFs in RNA Polymerase II Transcription. *Cell* **95** : 579–582.

Hernandez, N. 1993. TBP, A Universal Eukaryotic Transcription Factor. *Genes Dev*. **7** : 1291–1308.

Hess, O. and Meyer, G.F. 1968. Genetic Activities of the *Y* Chromosome in *Drosophila* during Spermatogenesis. *Adv. Genetics* **14** : 171.

Jacob, F. and Monod, J. 1961. Genetic Regulation Mechanisms in the Synthesis of Proteins. *J.Mol. Biol.* **3** : 318.

Lagerkvist, U. 1978. Two Out of Three: An Alternative Method for Reading. *Proc. Natl. Acad. Sci. U.S.* **75** : 1759.

Lake, J.A. 1977. Ribosome Structure and Functional Sites. *In* Gene expression. *FEBS 11th Meeting Copenhagen.* Eds. Clark, B.F.C. *et al.,* Vol. **43.** Pergamon Press, Oxford. p. 121.

Mathis, D.J. and Chambon, P. 1981. The SV40 Region TATA Box is Required for Accurate *in vitro* Initiation of Transcription. *Nature* **290** : 310.

McClain, W. H. 1993. Transfer RNA Identity. *FASEB J.* **7** : 72–78.

McKnight, S.L. and Miller, O. L. Jr. 1976. Silk Fibroin Gene in *Bombyx mori. In* Progress in Nucleic Acids. Res. & Mol. Biol. Eds. Cohn *et al.,* Vol. **19**: p. 313.

Miller, O.L. Jr. and Beatty, B.R. 1969. Visualisation of Nucleolar Genes. *Science* **164** : 955.

Nakamura, Y. *et al*., 1996. Emerging Understanding of Translation Termination. *Cell* **87** : 147–150.

Nirenberg, M.W. and Matthaei, J.H. 1961. The Dependence of Cell Free Protein Synthesis in *E. coli* upon Naturally Occurring or Synthetic Polyribonucleotides *Proc. Natl. Acad. Sci. U.S.* **47** : 1588.

Pelling, C. 1972. Developmental Studies on Giant Chromosomes. Springer-Verlag, Berlin.

Pukilla. Patricia, J. 1975. Identification of Lampbrush Chromosome Loops which Transcribe 5S Ribosomal RNA in *Notophthalmus (Triturus) viridescens. Chromosoma* **53** : 71.

Ptashne, M. and Gilbert, W. 1970. Genetic Repressors. *Sc. Amer.* **222** : 36.

Pain, V.M. 1996. Initiation of Protein Synthesis in Eukaryotic Cells. *Eur.J. Biochem*. **236** : 747–771.

Sadgopal, A. 1968. The Genetic Code after the Excitement. *Adv. Genetics* **14** : 325.

Sarabhai, A.S. *et al*., 1964. *Nature* **201** : 13.

Shine, J. and Dalgarno, L. 1974. The 3 Terminal Sequence of *E. coli* 16S *r*RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci. U.S.* **71** : 1342.

Smith, C.L. *et al*., 1978. Promoter-specific Inhibition of Transcription by Antibiotics which Act on DNA Gyrase, *Nature* **275** : 420.

Uptain, S.M., Cane, C. M., and Chamberlin, M.J. 1997. Basic Mechanisms of Transcript Elongation and its Regulation. *Annu. Rev. Biochem*. **66** : 117–172.

Varley, J.M. *et al*., 1980. Satellite DNA is Transcribed on Lampbrush Chromosomes. *Nature* **283** : 686.

Yanofsky, C. *et al*., 1964. On the Colinearity of Gene Structure and Protein Structure. *Proc. Natl. Acad. Sci. U.S.,* **51** : 266.

Wain-Hobson, S. *et al*., 1981. Preferential Codon Usage in Genes. *Gene* **13** : 355.

16

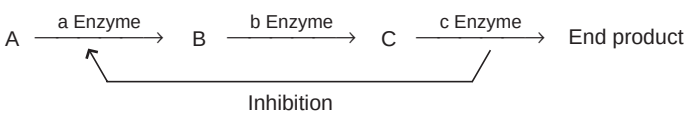
Genetic Regulation in Prokaryotes and Eukaryotes

Various questions concerning control of gene action remain unanswered till today. There is ample evidence that all cells of a higher organism inherit the *same* genetic constitution from a single-celled zygote; yet what specifically causes some genes to be active, others to be inactive at any time is not fully known. A specialized cell may produce large quantities of a single protein at a certain time. For example, the single gene for the production of silk fibroin in *Bombyx mori* produces 10^4 molecules of *m*RNA which are stable for a few days. Each *m*RNA has codon message for the synthesis of 10^5 proteins. It follows that a single gene produces 10^9 protein molecules (silk fibroin) in four days. Even in a one-celled organism all genes are not active at the same time. Obviously some mechanisms exist which regulate activity of genes. Regulation of genetic material requires that different proteins be made at different times and in different amounts. In eukaryotes regulation of protein synthesis plays a fundamental role in differentiation of cells. The first insight into genetic regulatory mechanisms came from induction and repression experiments on micro-organisms.

WHAT ARE INDUCTION AND REPRESSION?

Sometimes enzyme synthesis can be induced in cells by substrates called inducers. If yeast cells are grown for several generations in a medium containing lactose, fermentation occurs. The enzyme lactase is formed in high concentrations and it hydrolyses lactose into glucose and galactose. But if yeast cells are grown in a medium without lactose, they are found *not* to contain the enzyme lactase. When such cells are grown on a lactose medium, after about fourteen hours, the breakdown of lactose starts yielding glucose and galactose. Thus, we say that the presence of lactose in the medium induces the cells to form lactase. Therefore, lactase is an *inducible* enzyme.

Sometimes enzyme activity is repressed leading to inhibition of protein synthesis. The mechanism is called *negative feedback.* In a negative feedback system, the presence of the end product inhibits one of its own precursors by inhibiting the catalytic action of the first enzyme in the pathway. Thus the signal to stop making the end product is provided by the end product. When the end product has been used up, the blocked enzyme is freed to initiate synthesis once again.



REGULATION IN PROKARYOTES

The genetic basis for induction and repression was studied for several years by Jacob and Monod at the Pasteur Institute in Paris. They investigated regulation of the activities of genes which control fermentation of lactose through synthesis of the enzyme β-galactosidase in *E. coli.* They were awarded Nobel Prize in 1965.

If wild type *E. coli* cells are grown on a medium containing glucose, the cells are not able to utilise lactose and contain very small quantities of the enzyme β-galactosidase. But if wild type *E. coli* are grown on a medium devoid of glucose, but containing lactose as the only carbon source, within two minutes they start synthesising β-galactosidase. The synthesis of enzyme continues until very large amounts (about 3000 molecules per cell) have been produced. It was found that along with β-galactosidase, lactose induces the synthesis of two other enzymes *viz.* β-galactoside permease, which facilitates entry of lactose into the cells and β-galactoside transacetylase, whose function is obscure. The three collectively are known as *lac* enzymes.

Jacob and Monod studied gene regulation by isolating lactose mutants of *E. coli* which had one defect or the other in this regulation. The mutants revealed following different types of genes performing different functions in regulation.

(*a*) There are mutants which on growing on lactose medium, do not have *one* of the three enzymes synthesised on induction. Mapping techniques have shown that they have defects in three adjacent genes, each of which directs the synthesis of one of the enzymes. These are called *structural genes* and were shown by Lederberg and his colleagues to be arranged continuously on the chromosome in the order β-galactosidase (denoted *z* gene), permease (*y*) and transacetylase (*a*).

(*b*) *Constitutive mutants:* Enzymes may be constitutive or induced. Constitutive enzymes are those made in constant amounts in a cell, without regard to the metabolic state of the cell. Induced enzymes are made when required in response to the presence of their substrates in a cell.

Constitutive mutants of *E. coli* studied by Jacob and Monod are those that synthesise the three enzymes regardless of the presence or absence of the inducer. The gene showing this defect was called the regulator gene (denoted by *i*) and was found by mapping techniques to lie before the *z* gene.

(*c*) *E. coli* cells which were diploid as they had one complete chromosome and a second chromosome fragment homologous with a portion of the first chromosome. Such a bacterial cell is partially diploid for some genes and is called meroploid. Mutants with two chromosomes in a cell were analysed as follows: one chromosome had an active *i* but a defective *z* gene ($i^+ z^-$); the other chromosome having active *z* but defective *i* gene ($i^- z^+$). Such mutants produce β-galactosidase only in presence of inducer. It means that the active regulator gene (i^+) on one chromosome can regulate the active structural gene (z^+) on the other chromosome. Obviously, the regulator gene must be controlling the synthesis of an intermediary molecule which diffuses through the cytoplasm.

Some other experiments showed that the regulator gene codes for the amino acid sequence of a specific protein called repressor. The repressor molecule diffuses from the ribosomes where it is formed and becomes physically bound to a specific site on DNA near the structural gene.

(*d*) Further understanding of the repressor molecule came from mutants which were constitutive even though they had an active regulator gene. Such mutants failed to respond to the repressor because of a defect in a small specific region of the chromosome to which the repressor becomes bound. This was called the operator (denoted *O*) situated near the beginning of the β-galactoside structural gene (*z*).

The existence of operators was first revealed by genetic analysis. A mutation in the operator can make it inactive, preventing the binding of the repressor. When this happens, then constitutive enzyme synthesis occurs on the *z*, *y* and *a* genes. These mutants are therefore called operator constitutive O^c mutants. The operator constitutive mutants can be distinguished from mutations in repressor genes by measuring enzyme synthesis in partially diploid cells for certain chromosomal regions. If such a partially diploid cell contains one mutant and one functional repressor gene, repression occurs because repressor molecules produced by one functional locus can bind to both operators. But if there is one non-functional operator locus, the cells would always be constitutive.

From genetic studies in mutants combined with biochemical evidence, Jacob and Monod derived the following conclusions: the *lac* operon regulates the metabolism of lactose. When *E. coli* cells are grown on a medium containing lactose, the *lac* operon becomes functional and synthesises enzymes required for the transport and breakdown of lactose. The *lac* operon does not function when glucose is present or when lactose is absent from the medium. The *lac* operon contains a promoter (*p*), an operator (*o*), and three structural genes (*z*, *y* and *a*). It also has a transcription terminator gene (*t*) which gives the chain termination signal during *m*RNA synthesis. The regulator gene directs the formation of a repressor protein. This protein has affinity for the sequence of nucleotides of the operator and can bind to the operator. When the repressor is bound to the operator it prevents movement of RNA polymerase towards the three structural genes; no *m*RNA

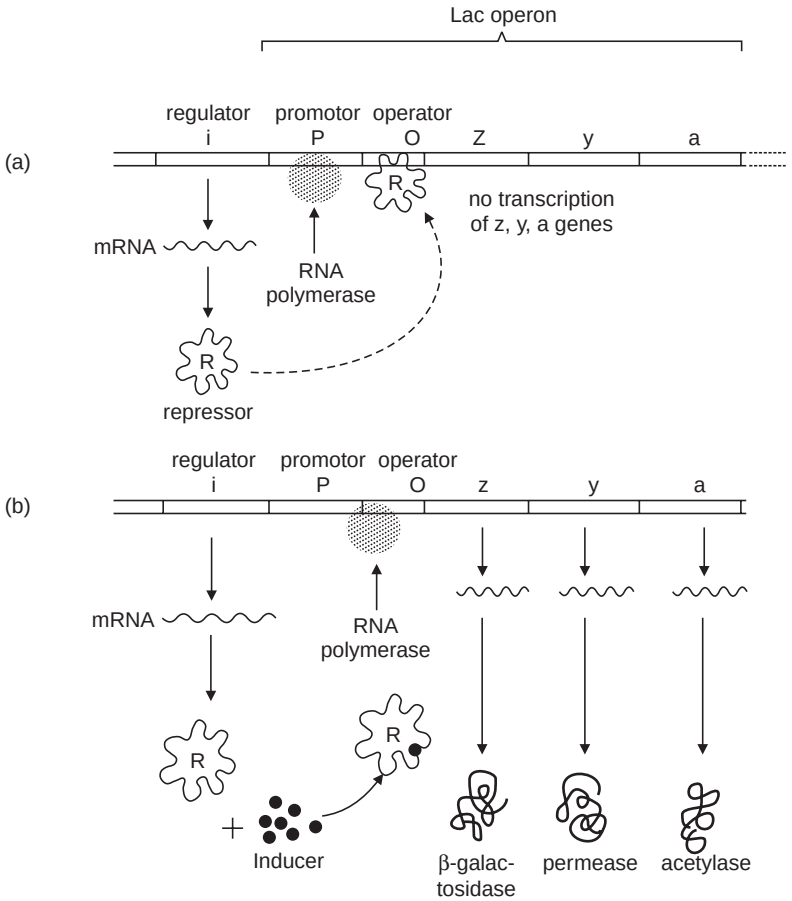


Fig. 16.1 Diagrams to illustrate functioning of the *lac* operon.

is synthesised, and therefore the three proteins are not formed. When inducer (lactose) is present, its molecules can bind to another active site of the repressor protein. This binding changes the three dimensional conformation of the protein, so that it loses its affinity for the operator. The operator is made free, *m*RNA is transcribed by the structural genes and all the three enzymes are synthesised (Fig. 16.1).

Enzyme Repression

Jacob and Monad also postulated repression of enzyme synthesis. For example, if histidine is added to the culture medium in which *E. coli* cells are growing, the enzymes leading to the formation of histidine become repressed, and histidine is not synthesised. The process is called feedback inhibition or end-product repression. By itself the repressor molecule is inactive. But when a corepressor binds with it, the repressor-corepressor complex binds with the operator gene that is specific for the structural genes of this operon, and prevents transcription. Thus there are two types of repressor molecules, one which binds with the inducer and promotes synthesis of enzymes; the other binds the corepressor resulting in end-product repression.

DETAILS OF GENES

Structural Genes

They control amino acid sequence of a protein by producing *m*RNA. There are as many structural genes as there are proteins which are regulated. Either a single *m*RNA is transcribed from each structural gene or, all the structural genes of an operon form a single polycistronic *m*RNA.

The Operator

It is the binding site for the repressor and controls the structural genes. Itself it is under the negative control of the repressor protein. The operator determines whether or not structural genes are to be repressed by the repressor produced by regulatory gene. If there is a mutation in the operator region rendering it functionless, the repressor is not able to bind and structural genes are transcribed. Mutants of the operator locus are called operator-constitutive mutants (O^c).

Gilbert and his colleagues isolated the operator region by breaking DNA of *lac* region into fragments 1,000 base pairs long. A property of the operator region is that when it is complexed to the repressor it is protected from digestion by DNAase. The operator region has double helical DNA about 27 base pairs long.

The Promoter

It consists of a small segment of DNA, less than 100 nucleotides in length which lies between the regulatory gene and the operator gene. It has a binding site which is recognised by the enzyme DNA directed RNA polymerase and which gives the initiation signal for transcription of *m*RNA. RNA polymerase moves along the operator locus and over to the structural genes which transcribe *m*RNA. Besides the *lac* operon, in other operons, the promoter region has a second binding site for a specific protein called the cyclic AMP receptor protein (described later).

The Lac Operon

The structural genes *z, y* and *a,* together with the operator *o* constitute the *lactose* operon. The operon provides a mechanism for the coordinate expression of structural genes (under the control of the operator and the regulator) resulting in the induction of enzymes due to a single inducer.

The operator locus controls the transcription of the entire group of coordinately induced genes so that a single large polycistronic messenger RNA is formed. The enzyme RNA polymerase binds to the promoter and all the genes in the operon are transcribed in a sequence. When repressor produced by the regulator is bound to the operator, transcription is not initiated and expression of all the genes in the operon is inhibited simultaneously.

Another feature of the operon is *polarity*. The genes *z*, *y* and *a* synthesise equal quantities of the three enzymes β-galactosidase, permease and acetylase and in the same order as their respective genes are located in DNA. Thus β-galactosidase (product of *z* gene) is synthesised first, followed by permease (*y* gene product), and last of all acetylase (produced by *a* gene). This is the effect of polarity. If there is mutation in *z* gene, there is no transcription of all three structural genes. If the mutation is in the *y* gene, then both *y* and *a* genes are inhibited, whereas the *z* gene synthesises β-galactosidase. These are called *polar* mutations.

Generally speaking we can say that an operon is a group of functionally related structural genes which can be turned on or off coordinately under the control of the same regulatory gene. The clustering of genes for various enzymes of a single metabolic pathway is probably necessary to facilitate operon function.

The Regulator Gene

It determines whether structural genes will be transcribed or not. It codes for the amino acid sequence of a specific *repressor protein*. The repressor molecule diffuses from the ribosomes where it is formed and becomes bound to the operator. Due to the diffusible nature of its product repressor, the regulator gene does not always lie adjacent to the structural genes it regulates. When the regulator gene undergoes mutation, it can no longer inhibit the transcription of the structural genes. The genes *z, y* and *a* then synthesise the three enzymes whether the inducer is present or not.

For a long time the chemical nature of the repressor molecules postulated by Jacob and Monod could not be identified. In 1967, W. Gilbert and Muller-Hill succeeded in isolating the *lac* repressor. They produced mutant *E. coli* cells which contained almost ten times the amount of *lac* repressor present in normal cells. The repressor protein has now been crystallised. It has a molecular weight of about 150,000 with a high affinity for its specific locus in *E. coli*. Besides the lac repressor, the galactose and tryptophan repressor have also been isolated in a pure form.

The lac repressor is a tetramer consisting of four units of a protein coded by the regulator gene, each of molecular weight about 37000. An *E. coli* cell contains about 10 such repressor molecules. There are almost 347 amino acid residues in a repressor. About 50 amino acids in the *lac* repressor bind to a sequence of about 12 base pairs in the operator region. This binding of the repressor prevents transcription of *m*RNA by RNA polymerase.

When lactose is present in the medium its uptake into the *E. coli* cell is followed by trans-glycosylation, a slight molecular rearrangement to form allolactose (Fig. 16.2). The synthesis of allolactose is catalysed by the few β-galactosidase molecules that are present prior to induction. Allolactose binds to the *lac* repressor to form an *inducer-repressor complex*. When the inducer becomes bound to the repressor, the binding of the repressor to the operator region is released due to a change in the 3-dimensional conformation of the repressor protein called *allosteric effect*. The unbound operator then permits RNA polymerase to transcribe. In this case allolactose is the effector molecule which prevents the regulatory protein from binding to the operator.

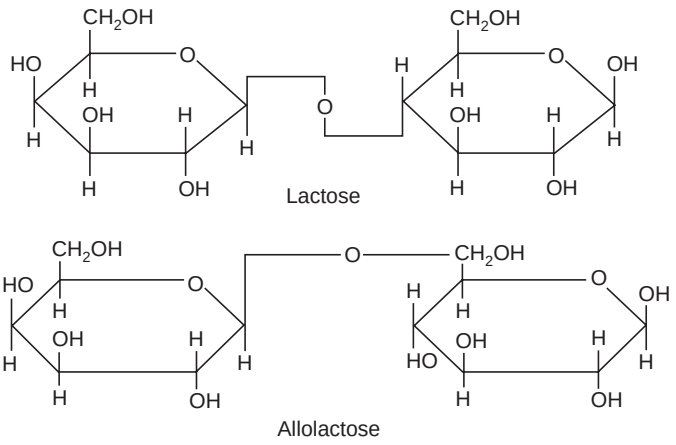


Fig. 16.2

THE INDUCIBLE AND REPRESSIBLE SYSTEMS

In the *inducible* system when *inducer* is absent, the repressor binds to the operator and blocks it; RNA polymerase cannot move along DNA so that very small amounts of *m*RNA if at all, are synthesised by the structural genes.

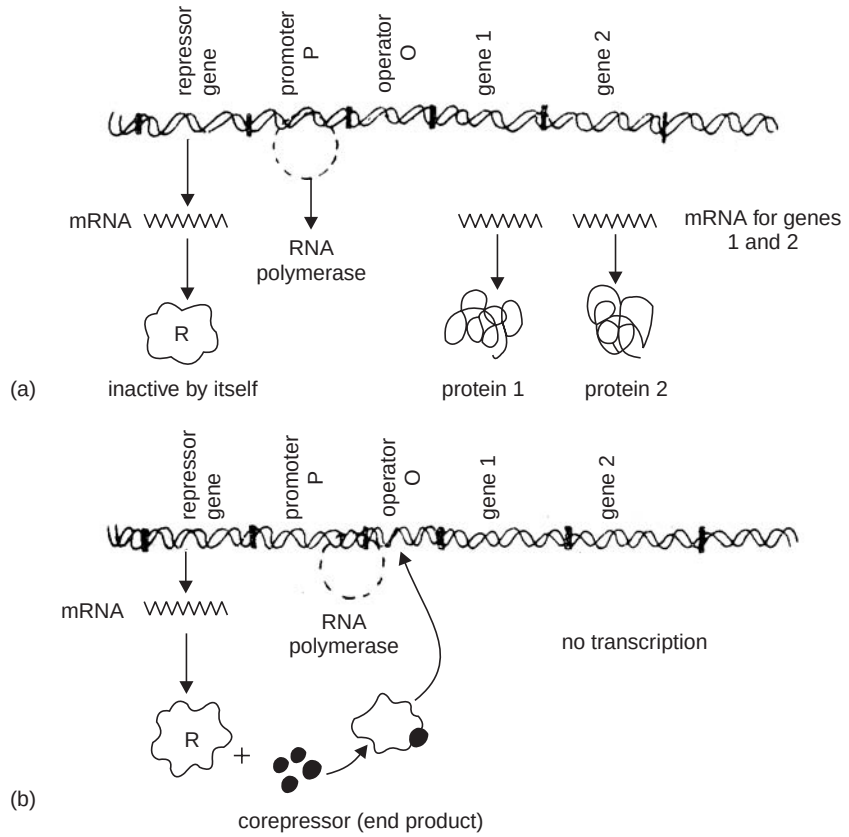


Fig. 16.3 *a*, *b* Diagrams to illustrate genetic regulation in a repressible system.

But when lactose is present in the medium as an *inducer,* the operon is induced to synthesise large quantities of the enzymes required in the transport and catabolism of lactose. In this case the operon functions because the repressor gets bound to the inducer molecule, the operator becomes free, and structural genes synthesise *m*RNA. The lac operon provides one example of an inducible system in which the existence of the nutrient in the medium induces synthesis of large amounts of enzymes needed for the catabolism of that nutrient. Such systems therefore, operate in the degradation of exogenous substrates in catabolic processes.

In repressible systems the operon controls synthesis of proteins or enzymes needed for anabolic reactions. Such an operon continues to function normally until there is an excess of the products. When the end product is in excess it functions as a repressing metabolite called corepressor. In this system the repressor is inactive by itself. But when the corepressor binds to the repressor to form a *repressor-corepressor* complex which attaches to the operator, the structural genes cannot transcribe *m*RNA (Fig. 16.3). In a repressible system the operon is negatively controlled. In addition to the histidine operon already described, the tryptophan operon in *E. coli* also functions as a repressible operon. When there are normal concentrations of tryptophan in a cell, the operon is functional or *derepressed.* But when tryptophan is in excess, it acts as a corepressor and binds to the inactive repressor. The complex attaches to the operator and prevents *m*RNA synthesis by structural genes. As the concentration of tryptophan decreases, it causes the repressor molecules to remain free, the operator becomes unbound, and the structural genes transcribe *m*RNA. Thus the operon again becomes derepressed.

TRANSCRIPTIONAL AND TRANSLATIONAL CONTROL

The lactose operon demonstrates that the control of transcription involves interaction of regulatory proteins with specific DNA sequences, and this is broadly applicable to eukaryotes as well. Regulatory sequences such as the operator are called ***cis*-acting control elements** because they bring about expression of only linked genes on the same DNA molecule. In contrast, repressor proteins are called ***trans*-acting factors** because they can influence the expression of genes located on other chromosomes within the cell. Furthermore, the *lac* operon is considered as an example of negative control of gene expression because binding of the repressor *inhibits* transcription. There are however, other examples where *trans*-acting factors are activators (positive control) of transcription.

Positive control of transcription has been demonstrated in *E. coli* through studies on the effect of glucose on the expression of genes encoding enzymes that lead to breakdown (catabolism) of other sugars, such as lactose. Lactose provides an alternative source of carbon and energy. As long as glucose is available, it is preferentially utilised, with the result that enzymes involved in catabolism of alternative energy sources are not expressed. That means, if *E. coli* are grown on a medium that provides both glucose and lactose, the *lac* operon is not induced, and only glucose is used by cells of *E.coli*. Thus, glucose represses the *lac* operon even in the presence of the normal inducer, lactose.

Repression by glucose, also called **catabolite repression** is actually mediated by a positive control system which is determined by levels of cyclic AMP. (*c*AMP) (Fig. 16.4). In bacteria cAMP is produced from ATP by enzyme adenylyl cyclase. The conversion of ATP to *c*AMP is regulated in

such a way that levels of *c*AMP increase when glucose levels drop. Cyclic AMP then binds to a transcriptional regulatory protein called catabolite activator protein (**CAP**). The binding of *c*AMP to CAP stimulates the binding of CAP to its specific DNA sequences. In the *lac* operon this specific DNA sequence is located approximately 60 bases upstream of the transcription start site. CAP then interacts with the alpha subunit of RNA polymerase, and that facilitates the binding of polymerase to the promoter and activating transcription.

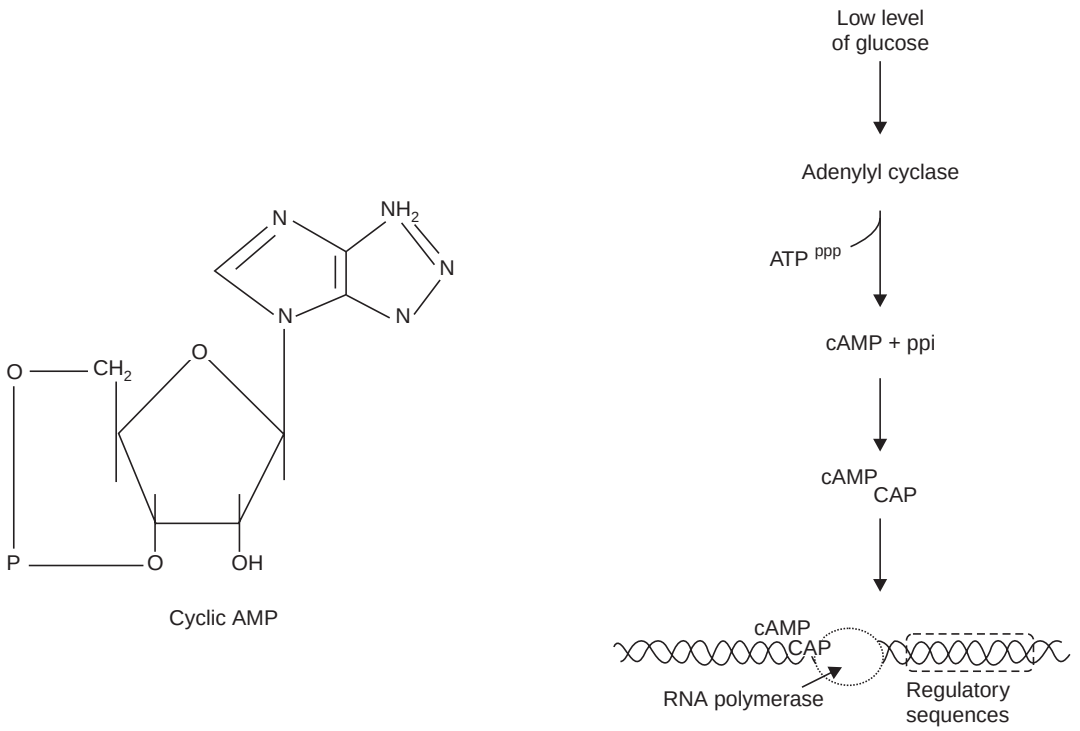


Fig. 16.4 Positive control of *lac* operon by glucose. Activated by low levels of glucose, adenylyl cyclase converts ATP to cyclic AMP. CAMP then binds to the catabolite activator protein (CAP) which stimulates its binding to regulatory sequences of various operons that are related with metabolism of other sugars such as lactose. CAP interacts with the α subunit of RNA polymerase to initiate transcription.

Regulation of E. coli Tryptophan Operon

Genes of the amino acid tryptophan (*trp* genes) are considered as repressible genes in which the presence of the metabolite (*trp*) in the environment turns off the expression of its structural genes. Tryptophan acts as a **co-repressor**. Regulation of the *trp* operon occurs in two ways. In the first, the expression of the five structural genes *E*, *D*, *C*, *B*, and *A* that code for enzymes involved in the synthesis of tryptophan, is controlled by a specific **regulatory gene**. The regulatory gene codes for a specific protein called **repressor.** The repressor by itself is inactive, but when it becomes complexed with tryptophan (co-repressor) it becomes activated. The activated repressor-co-repressor complex binds to a specific region of DNA, the operator situated adjacent to the structural genes that are being regulated. This blocks the movement of RNA polymerase towards structural genes. The structural genes, operator and promoter regions together constitute the

operon. Thus, when tryptophan is present in the environment, the repressor forms a complex with tryptophan, then binds with **operator** and prevents transcription of the structural genes. On the contrary, when tryptophan is lacking in the environment, the repressor remains free and inactive, does not bind with operator, resulting in transcription of structural genes and synthesis of tryptophan. This is referred to as a **negative control system** because the repressor which is product of the regulatory gene acts to turn off transcription of structural genes (Fig. 16.5).

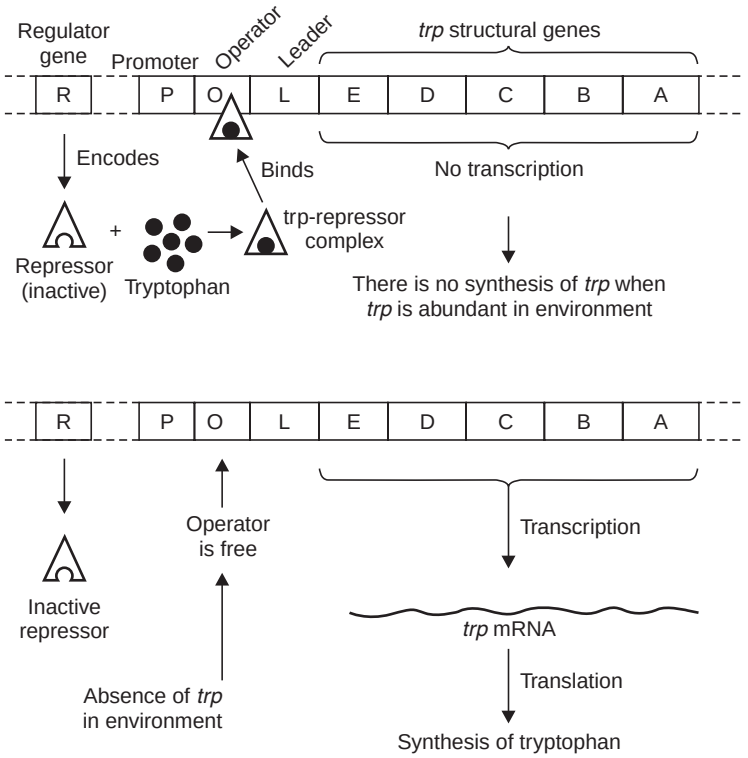


Fig. 16.5 Negative control system for regulation of the tryptophan gene.

The second mechanism called **attenuation**, regulates the expression of tryptophan structural genes by controlling the ability of RNA polymerase to continue elongation over a specific nucleotide sequence. This mechanism operates when high levels of tryptophan are available. There is attenuation of regulation by a sequence that terminates transcription prematurely. This sequence or region of attenuation is located 162 nucleotides downstream of the transcription start site, that is the first structural gene. Transcription terminates in this region if tryptophan is available, before RNA polymerase reaches the first structural gene. In other words, attenuation occurs if the specific aminoacylated *t*RNA is available. If not, transcription continues, producing a functional *trp m*RNA.

Transcription is initiated at the promoter region producing what is referred to as the leader transcript. The leader RNA contains a start and a stop signal for protein synthesis. As described in Chapter 15, since prokaryotes lack a nuclear membrane, transcription and translation can occur simultaneously, unlike eukaryotes where there is spatial separation of transcription (in nucleus) and translation (in cytoplasm). Therefore, while the leader RNA is being synthesised, ribosomes begin translation at the 5′ end. This results in a short peptide chain while the RNA

polymerase is transcribing the leader region. If tryptophan-*t*RNA is available, synthesis of the peptide chain will continue, until the ribosome reaches the stop signal present in the leader RNA. However, if there is not sufficient tryptophan-*t*RNA, the leader RNA will not be translated into peptide, and the ribosome will be arrested at the tryptophan codons in the leader RNA, without reaching the stop signal.

Besides the stop and start sequences, the leader RNA contains 4 regions which have *complementary* sequences which enable formation of stem and loop structures by base pairing. Region 1 can form base pairs with region 2; region 2 can form base pairs on both sides either with region 1 or with region 3; region 3 can likewise form base pairs with region 2 or with region 4; region 4 can base pair only with region 3. Therefore, 3 possible stem/loop structures can form in the RNA transcript (Fig. 16.6).

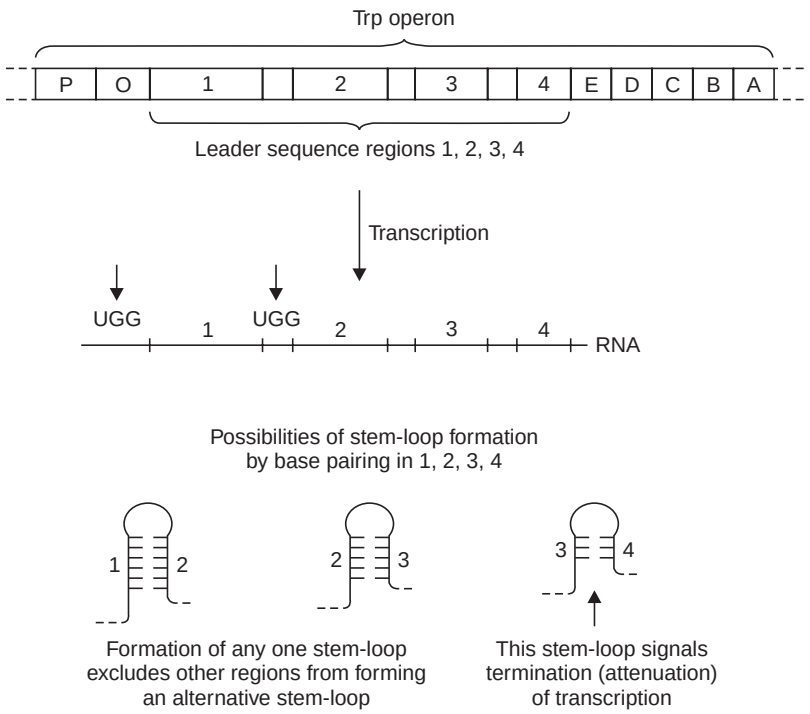
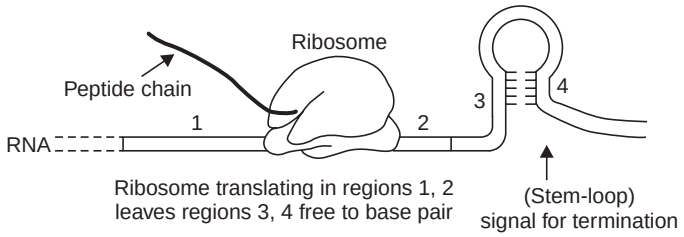


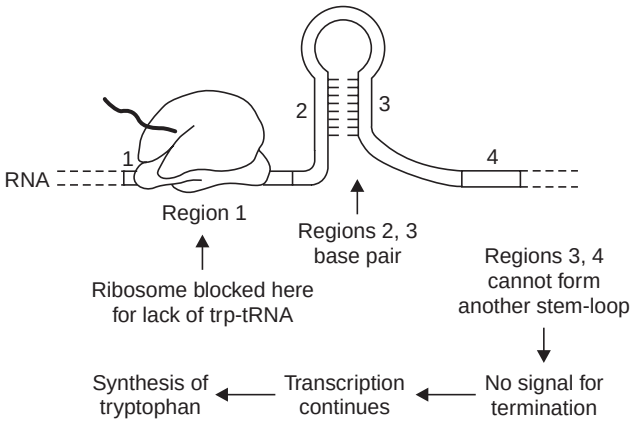
Fig. 16.6 Formation of stem-loop structures in RNA transcription of trp operon.

When region 3 base pairs with region 4, it generates a signal for attenuation, that is, premature termination of transcription. However, it must be noted that if stem/loop has already been formed in the region preceding region 3, then region 3 will not be available to base pair with region 4. Another important point to note is that, if the ribosome is translating in region 2, then region 2 would not be available for base pairing with region 1 or with region 3. In that situation region 3 will be free to base pair with region 4. As stated above, base pairing only between regions 3 and 4 to form stem/loop signals RNA polymerase to terminate transcription. This implies that when sufficient amount of tryptophan-*t*RNA is available to translate the leader RNA, it will stop transcription (attenuation) prematurely, and structural genes will not be transcribed. In contrast, if tryptophan-*t*RNA is lacking or insufficient to translate the leader RNA, there will be no attenuation. In that case the ribosome will stop at the two *trp* codons in region 1, thus leaving

region 2 available to base pair with region 3. That means region 3 would not be available to base pair with region 4, which is an essential requirement for signaling RNA polymerase to terminate transcription. In absence of attenuation then structural genes will be transcribed (Fig. 16.7).



A. High tryptophan levels – attenuation



B. Low tryptophan levels – transcription

Fig. 16.7 The attenuation mechanism for termination of transcription in the *trp* operon.

Regulation of Transcription in Eukaryotes

The control of expression of eukaryotic genes is more complex than in prokaryotes, and is primarily at the level of initiation of transcription. In general, transcription in eukaryotic cells is controlled by proteins that bind to specific regulatory sequences and modulate the activity of RNA polymerase. In the many different cell types of multicellular eukaryotic organisms, regulation of gene expression is accomplished by the combined actions of multiple different transcriptional regulatory proteins, by methylation of DNA, and packaging of DNA into chromatin.

Promoters and Enhancers

In bacteria, transcription is regulated by the binding of proteins to *cis*-acting sequences, as in the *lac* operon, that control transcription of adjacent genes (*z, y, a*). Similar *cis*-acting sequences regulate the expression of eukaryotic genes. The method of identifying these sequences is based on the use of gene transfer assays by which the activity of supposed regulatory regions of cloned genes are studied (Fig. 16.8). The regulatory sequence is ligated to a **reporter gene** that encodes an easily detectable enzyme. The reporter gene is transferred into cultured cells (**transfection**).

The expression of the reporter gene indicates biological activity of the regulatory sequence and provides a sensitive assay for the ability of the cloned regulatory sequences to direct transcription.

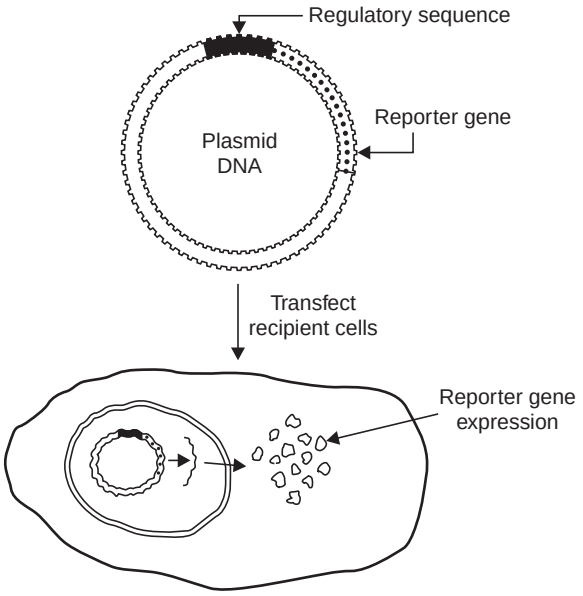


Fig. 16.8 Gene transfer assay; method using reporter gene.

The two core promoter elements, the **TATA box** and the **Inr sequence** in genes transcribed by RNA polymerase II serve as specific binding sites for transcription factors. Inr is the initiator sequence spanning the transcription start site in promoters of many genes transcribed by RNA polymerase II. Other *cis*-acting sequences function as binding sites for a variety of regulatory factors that control expression of individual genes. These *cis*-acting regulatory sequences are usually located upstream of the TATA box. Interestingly, two regulatory sequences commonly found in eukaryotic gene were found to be present in the promoter of the herpes simplex virus gene that encodes thymidine kinase. These two sequences are located about 100 base pairs upstream of the TATA box, and their consensus sequences are CCAAT and GGGCGG (called the GC box). The binding of specific proteins to these sequences has been shown to initiate transcription.

In contrast to the CCAAT and GC boxes in the thymidine kinase promoter of herpes simplex virus, the regulatory sequences of several mammalian genes are located further away, upto 10 kilobases , from the transcription start site. These sequences are called **enhancers** and were first described in the virus SV40. Like promoters, enhancers function by binding transcription factors that act by regulating RNA polymerase. Transcription factors bound to distant enhancers function by the same mechanisms as those bound adjacent to promoters, that is, the *cis*-acting regulatory sequences. The binding of specific transcriptional regulatory proteins to enhancers is responsible for the control of gene expression during development, differentiation and in response of cells to hormones and growth factors.

Transcriptional Activators

Among the most thoroughly studied transcription factors are the **transcription activators**, which bind to regulatory DNA sequences and stimulate transcription. Transcriptional activators consist of two domains, one region binding to DNA to anchor the factor to the proper site on DNA; the other activates transcription by interacting with other components of the transcriptional

system. Detailed studies have revealed that the DNA-binding domains of many of these proteins are related to one another. The **zinc finger** domains contain repeats of cysteine and histidine residues that bind zinc ions and fold into finger-like loops that bind DNA. Transcription factors of the **steroid hormone** receptors contain zinc finger domains. The steroid hormone receptors regulate gene transcription in response to hormones estrogen and testosterone. The activation domains of transcription factors are not as well characterised as their DNA binding domains.

Regulation by Alternative Splicing of RNA Transcript

The primary transcript of some genes could be spliced in alternative ways to yield different products. Even when the same promoter is used to transcribe a gene, different cell types can produce different quantities of a protein, or even a different protein. This could result from differences in the *m*RNA produced or from processing of *m*RNA. This can be achieved when the same transcript from one cell type is spliced differently from the transcript in another type of cell. The protein-coding exons may be the same in the different cell types, but the splicing pattern of the transcript may be different. In that case the protein is identical, but the rate of synthesis is different, because the *m*RNA molecules are not translated with the same efficiency. In other cases, the protein-coding part of the transcript has a different splicing pattern in each cell type, with the result that the *m*RNA molecules produced code for proteins that are not identical even though they share certain exons.

Transcripts of the human genome are frequently spliced in alternative ways. Owing to this, the approximately 30,000 human genes may encode 64,000 to 96,000 different proteins.

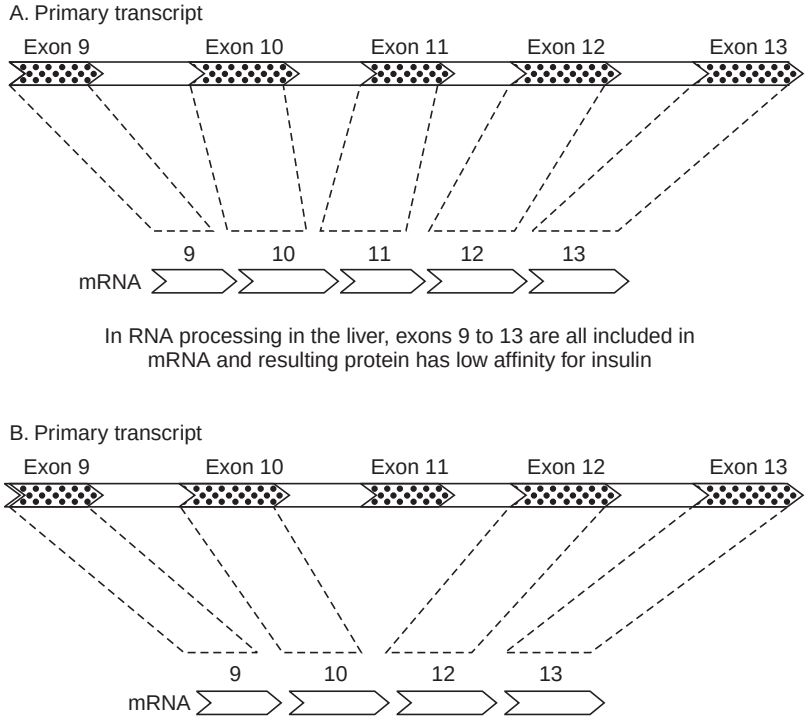


Fig. 16.9 Alternative splicing of human insulin gene mRNA.

Alternative RNA processing is considered to be one of the principle sources of human genetic complexity. For example, the human insulin receptor gene undergoes alternative splicing that results in the inclusion or exclusion of exon number11 in the *m*RNA (Fig. 16.9). The resulting forms of the polypeptide chain differ in length by 12 amino acids. In liver cells, all the 20 exons are present in *m*RNA for the long form of the receptor protein (Fig. 16.9, Part *A*), whereas in skeletal muscle exon 11 is eliminated along with the flanking introns and excluded from the *m*RNA for the short form (Fig. 16.9, Part *B*). The long form of the receptor shows low affinity for insulin and is expressed in tissues such as the liver that are exposed to relatively higher concentrations of insulin. The short form of the protein has a high affinity for insulin and is expressed preferentially in tissues such as the skeletal muscle that are normally exposed to lower levels of insulin.

Thus, alternative splicing provides a mechanism for generating proteins with different properties from the *same* gene. The *Dscam* gene of *Drosophila* could give rise to approximately 38016 different proteins by alternative splicing. The actual number of proteins synthesised is not known. The human genome that contains 30,000 to 40,000 genes produces different proteins whose number is several times greater, by alternative splicing.

Unlike genes of *Drosophila* and lower worms, human genes are distributed over a large region of the genome, and the primary *m*RNA transcripts are very long. Alternative splicing of most human genes leads to multiple protein products. About one third of *all* human genes are believed to undergo alternative splicing. Among genes that are alternatively spliced, the average number of distinct *m*RNAs produced from the primary transcript ranges between 2 and 7. The average number of different *m*RNAs per gene across the genome is in the range of 2 to 3, which includes genes that produce a single *m*RNA as well as those that produce multiple different *m*RNAs. Thus alternative splicing greatly increases the number of protein products that can be encoded from a relatively small number of genes.

Stability of Messenger RNA

Different *m*RNAs persist for different time periods in a cell. Since short-lived *m*RNAs would produce fewer protein molecules as compared with long-lived *m*RNAs, factors affecting the **stability of mRNA** determine the level of gene expression. *m*RNA is degraded by one of two pathways. One route of degradation is the **deadenylation-dependent pathway** which begins with enzymatic cutting down of the length of the poly-*A* tail. When the poly-*A* tail is trimmed to a length of 25 to 60 nucleotides, the *m*RNA becomes susceptible to a decapping enzyme that removes the 5′ cap so that the molecule is not able to initiate translation. *m*RNA in this state is rapidly degraded by exonuclease. The other degradative pathway is the **deadenylation-independent pathway**, which is initiated either by decapping or with endonuclease cleavage of the *m*RNA, after which exonuclease activity degrades the molecule completely. Thus, the deadenylation-independent pathway prevents accumulation of truncated (broken) polypeptides in the cell.

The stability of RNA could be affected by a phenomenon known as **RNA interference** (**RNAi**) which leads to destruction of RNA transcripts. In this, RNA interference introduces a few hundred nucleotide pairs of **double-stranded RNA** that triggers degradation of RNA transcripts containing *homologous sequences*. RNA interference has been seen to occur in a wide variety of organisms including *Drosophila*, trypanosomes, amphibians, mammals and humans. It seems that the genes in question are transcribed at the normal level, the transcripts are degraded *before* they leave the nucleus, and are not translated. This effect of **gene silencing** is highly specific and requires only a few double-stranded RNA molecules per cell to trigger the silencing. Importantly, the RNAi effect can be transmitted from cell to cell.

Regulation at the Level of Translation

In eukaryotic cells, transcription and translation are uncoupled in the process of gene expression. This permits regulation at the level of translation independently from transcription. The major types of translational control are: inability of a *m*RNA molecule to be translated under certain conditions; regulation of the overall rate of protein synthesis; inhibition or activation of translation by **small regulatory RNAs** that undergo base pairing with the *m*RNA; activation of previously untranslated cytoplasmic *m*RNAs. In the case of translational control by small regulatory RNAs, usually the regulatory RNAs are complementary in sequence to part of the *m*RNA whose translation they control. An RNA sequence that is complementary to a *m*RNA is called an **antisense RNA**. The antisense regulating RNAs act by pairing with the *m*RNA. It can either activate or inhibit translation. Small regulatory RNAs can also regulate translation.

Epigenetic Control of Genetic Regulation

Epigenetic phenomena are alternative states of gene activity that are heritable, but do not follow Mendelian rules of inheritance, are not explained by mutation, changes in gene sequence or normal developmental regulation. They are changes brought about in gene expression by heritable chemical modifications in DNA. The prefix *epi* means "besides" or "in addition to". Therefore, **epigenetic** refers to heritable changes in gene expression that are not associated with changes in the DNA sequence, but with something "in addition to" the DNA sequence, usually either chemical modification of the DNA bases or proteins bound with DNA. It is now clear that many epigenetic phenomena occur largely via changes in chromatin structure. In general, methylation of DNA is associated with turning off gene expression. However, some organisms that clearly exhibit epigenetic effects, for example *Drosophila*, do not have DNA methylation. Modifications of histones and non-histone chromosomal proteins have also been implicated in epigenetics.

DNA METHYLATION

The DNA of many higher eukaryotes is modified after replication by methylation of cytosine bases at the 5-carbon position. The cytosines are incorporated in their normal unmodified form in the course of DNA replication; the methyl group (CH_3) is then added by an enzyme DNA methylase. In mammals DNA is methylated specifically at the cytosines that precede guanosine residues, that is CpG dinucleotides (5′-CG-3′). Many mammalian genes have CG-rich regions upstream of the coding region that provide multiple sites for methylation. These are called *CpG islands* where "*p*" represents the phosphate group in the polynucleotide backbone. In higher plants 3–7% of the bases contain methylated cytosine, while in animal DNA the proportion is less. In some protozoa, the adenine groups are methylated, while in the green alga *Chlamydomonas* both adenine and cytosine are methylated. There is a higher concentration of methylated cytosines in highly repetitious eukaryotic DNA. Methylation is found to correlate with reduced transcriptional activity of genes. Genes that contain high frequency of CpG dinucleotides in the vicinity of their promoters show a low rate of transcriptional activity. Methylation inhibits transcription of these genes through the action of a protein, MeCP2, that specifically binds to methylated DNA and brings about repression of transcription.

Heavy methylation associated with reduced transcription is seen in the inactive *X* chromosome in mammalian cells which is extensively methylated. In fact, in adult mammals, the majority of CpG dinucleotides in all chromosomes are methylated in adult cells. The

unmethylated CpGs are usually associated with the promoters of housekeeping genes that are transcriptionally active.

Treatment of cells with the cytosine analog azacytidine reverses methylation and can restore transcriptional activity. For example, in cell culture, some lineages of rat pituitary tumour cells are able to express the gene for prolactin, whereas other related lineages cannot do so. The reason being that the gene is methylated in the non-producing cells but is not methylated in cells that produce prolactin. Treatment of non-producing cells with azacytidine reverses methylation, resulting in prolactin expression in these cells.

The regulatory role of DNA methylation has been determined through an unusual type of epigenetic silencing known as **genomic imprinting**, which controls the expression of some genes involved in development of embryos in mammals. Usually, both paternal and maternal alleles of a gene are expressed in diploid cells. However, a few imprinted genes have been found in mouse and humans whose expression depends upon whether they have been transmitted (inherited) from the mother or from the father. In some cases, only the paternal allele of an imprinted gene is expressed, whereas the maternal allele is transcriptionally inactive. In other imprinted genes, the maternal allele is expressed while the paternal allele is inactive. It seems that DNA methylation is able to distinguish between paternal and maternal alleles of imprinted genes. For example, the gene *H19* is specifically methylated during development of male germ cells, but not female germ cells. Thus *H19* is transcribed only from the maternal copy. After fertilisation, the zygote develops into an embryo containing a methylated paternal allele and an unmethylated maternal allele of this gene. The paternal *H19* allele therefore, remains methylated and transcriptionally inactive in embryonic cells and somatic tissues.

The process of genomic imprinting displays following characteristics. Imprinting occurs in the germ line; it affects at the most a few hundred genes, many of them located in clusters; it is accompanied by heavy methylation; imprinted genes are differentially methylated in the female and male germ lines; once imprinted and methylated, a **silenced gene** remains transcriptionally inactive during embryogenesis.

Although mammalian gametes are extensively methylated, most of the DNA is demethylated in preimplantation development, except for imprinted genes that have a sex-specific **gene silencing** pattern. After implantation, embryonic DNA is remethylated to the heavy methylation levels that are present in adult somatic cells. In the germ line of both males and females, all remethylated genes acquire methylation to an identical level. Except for a few genes that have sex-specific patterns of imprinting and different levels of methylation in female and male in females, imprinted genes undergo methylation during growth of oocytes prior to ovulation. Because methylation associated with imprinting is retained throughout embryonic development, any gene that is imprinted in either the female or male germ line has, therefore, only one active copy in the embryo.

Epigenetic sex-specific gene silencing is seen in a pair of human syndromes called *Prader-Willi syndrome* and *Angelman syndrome* that display neuromuscular defects, mental retardation and some other abnormalities. Both conditions result from rare spontaneous deletions involving the chromosomal region *15q11*. If the deletion takes place in the father, it results in Prader-Willi syndrome, whereas if the deletion occurs in the mother, it produces Angelman syndrome. The reason is that, the DNA in the deleted region *15q11* includes at least three genes that are imprinted and differentially methylated in the male and female gametes.

Ribosomal RNA Gene Silencing

Eukaryotes have hundreds of *r*RNA genes whose transcription by RNA polymerase I maintains ribosome production and protein synthesis. However, only a subset of the total *r*RNA gene pool is active at any one time, thus maintaining a dynamic balance between gene silencing and activation. Since *r*RNA genes are essentially identical in sequence in a pure species, it is not possible to distinguish between active and inactive genes. The explanation is provided by the epigenetic phenomenon called **nucleolar dominance**. Epigenetic silencing of one parental set of *r*RNA genes and transcription of *r*RNA genes from the other parent is associated and is called nucleolar dominance, commonly observed in interspecific hybrids. Parental genes in hybrids typically differ in sequence as well as expression, and have allowed understanding of chromatin modifications of *r*RNA genes in active and inactive states. Since only active *r*RNA genes initiate formation of a nucleolus in the nucleolus organiser region of the chromosome, hence the name nucleolar dominance. The genes that encode ribosomal RNA are present in two types of chromatin, one that permits transcription, the other that is transcriptionally inactive. As in other epigenetic phenomena, chromatin modifications result in selective gene silencing. Although the epigenetic mechanisms that discriminate between parental sets of *r*RNA genes are not known, it seems that the level of cytosine methylation and histone modifications (acetylation) that alter chromatin structure effectively turn *r*RNA genes on and off resulting in active and inactive *r*RNA genes.

QUESTIONS

1. What are constitutive mutants? How did Jacob and Monod discover the role of the regulator gene?
2. Describe briefly how the *lac* operon functions in *E. coli.* Which component/s of the *lac* operon have not been found in higher organisms as yet?
3. State whether β-galactosidase synthesis is inducible or constitutive in the following meroploid strains of *E. coli:*

   i^+/i^+, $0^+/0^c$

   i^+/i^- $0^c/0^c$

   i^-/i^-, $0^+/0^+$
4. What effects would be observed if a mutation occurred in the following regions of the lac operon: (*a*) *lac* promoter; (*b*) *lac* operator; (*c*) the gene *z;* (*d*) the gene *y;* (*e*) the gene *i*.
5. What do you understand by

   (*a*) allosteric control,

   (*b*) repressor,

   (*c*) sensor region.
6. Give 5 reasons to explain why Jacob and Monod's model for gene regulation cannot be directly applied to eukaryotes.
7. Gene amplification might be one of the possible ways of regulating gene expression. Explain.
8. Using the example of the steroidal hormone ecdysone, comment on the involvement of hormones in genetic induction.
9. Which of the several types of proteins found in the cell nucleus are more suitable in your opinion, as candidates for regulating gene expression. Give reasons.

SELECTED READINGS

Akam, M.E. *et al.,* 1978. Drosophila: The Genetics of Two Major Larval Proteins. *Cell* **13** :215.

Allfrey, V.G. and Mirsky, A. E. 1963. Mechanisms of Synthesis and Control of Protein and RNA Synthesis in the Cell Nucleus. *Cold Spring Harb. Symp. Quant. Biol.* **28**:247.

Alt, F.W. *et al.,* 1976. Synthesis and Degradation of Folate Reductase in Sensitive and Methotrexate Resistant Lines of SI80 Cells. *J. Biol. Chem.* **251**:3063.

Ashburner, M. 1970. Patterns of Puffing Activity in the Salivary Gland Chromosomes of *Drosophila* V.

Response to Environmental Treatments. *Chromosoma* **31**:356.

Ashburner, M. 1970. Formation and Structure of Polytene Chromosomes during Insect Development. *Adv. Insect Physiol.* **7**:1.

Ashburner, M. 1972. Ecdysone Induction of Puffing in Polytene Chromosomes during Insect Development. *Adv. Insect Physiol.* **7**:1.

Atchison, M.L. 1988. Enhancers. Mechanisms of Action and Cell Specificity. *Annu. Rev. Cell Biol.* **4** : 127–153.

Banhmann, K. 1972. Genome Size in Mammals, *Chromosoma* **37**:85.

Balhorn, R. *et al.,* 1975. Phosphorylation of the Lysine-rich Histones Throughout the Cell Cycle. *Biochemistry.* **14**:2504.

Baltimore, D. and Huang, A. S. 1970. Interaction of HeLa Cell Proteins with RNA. *J. Biol.* **47**:263.

Barlow, D.P. 1995. DNA Methylation and Genomic Imprinting. *Cell* **77** : 473–476.

Beard, P. 1978. Mobility of Histones on the Chromosome of SV40. *Cell* **15**:955.

Beerman, W. 1971. Effect of α-amanitin on Puffing and Intranuclear RNA Synthesis in *Chironomus* Salivary Glands. *Chromosoma* **34**:152.

Beerman, W. and Pelling, C. 1965. H^3-thymidine Markierung Einzelner Chromatiden in Riesenchromosomen. *Chromosoma* **16** : 1.

Berendes, H.D. 1968. Factors Involved in the Expression of Gene Activity in Polytene Chromosomes. *Chromosoma* **24**:418.

Bonner, J, 1979. Properties and Composition of Isolated Chromatin. *In* Chromatin Structure and Function. Part A. Ed. Nicolini. Plenum Press, New York.

Bonner, J. *et al.,* 1968. The Biology of Isolated Chromatin. *Science* **159***:*47.

Brown, D.D. 1981. Gene Expression in Eukaryotes. *Science* **211***:*667.

Brown, D.D. and David, I. B. 1968. Specific Gene Amplification in Oocytes. *Science* **160**:272.

Buratowski, S.1995. Mechanisms of Gene Activation. *Science* **270** : 1773–1774.

Chang, J.C. *et al.,* 1978. Effect of Histone Acetylation on Structure and *in vitro* Transcription of Chromatin. *Nucleic Acids Res*. **5**:3523.

Clever, U. and Karlson, P. 1960. Induktion von puffveränderungen in den speicheldrüsen chromosomen von. C *tentans* durch ecdysone. *Exp. Cell Res*. **20**:623.

Davidson, E.H. and Britten, R.J. 1979. Regulation of Gene Expression: Possible Role of Repetitive Sequences. *Science* **204**:1052.

Douvas, A. S. and Bonner, J. 1977. Contractile Proteins in Eukaryotic Chromatin. *In* Mechanisms and Control of Cell Division. Eds. Rost and Gifford. Dowden, Hutchinson and Ross. New York.

Davie, J. R. and Candido, P.M. 1978. Acetylated Histone H4 is Preferentially Associated with Template Active Chromatin. *Proc. Natl. Acad. Sci. U.S.,* **75**:3574.

Ehrlich, M. and Wang, R.Y. 1981. 5-methylcytosine in Eukaryotic DNA. *Science* **212**:1350.

Federoff, N. V. and Brown, D.D. 1978. The Nucleotide Sequence of Oocyte 5S DNA in *X. laevis.* I. The AT Rich Spacer. *Cell* **13**:701.

Felsenfeld, G. 1996. Chromatin Unfolds. *Cell* **86** : 13–19.

Grunstein, M. 1992. Histones as Regulators of Genes. *Sci. Amer*. **267** (4) : 68–74B.

Gurdon, J.B. 1968. Transplanted Nuclei and Cell Differentiation. *Sc. Amer.* **219**:24.

Hartwell, L. H. 1970. Biochemical Genetics of Yeast. *Annu*. *Rev. Genetics* **4**:373.

Judd, B.H. 1979. Mapping the Functional Organisation of Eukaryotic Chromosomes. *In* Cell Biology, A Comprehensive Treatise. Eds. Prescott. D. M. and Goldstein Vol. 2. Academic Press, New York. P. 223.

Kano, Y. *et al.,* 1981. *In Vivo* Correlation between DNA Supercoiling and Transcription. *Gene* **13**:173.

Lewin, B. 1980. Gene Expression. Vol. 2. Wiley Interscience, New York.

Manes, C. and Menzel, P. 1981. Demethylation of C_pG Sites in DNA of Early Rabbit Trophoblast. *Nature* **293**:589.

Maniatis, T., Goodbourn, S. and Fischer, J.A. 1987. Regulation of Inducible and Tissue-specific Gene Expression. *Science* **236** : 1237–1244.

Paulson, J.R. and Laemmli, U.K. 1977. The Structure of Histone-depleted Metaphase Chromosomes. *Cell* **12**:817.

Ptashne, M. and Gann, A. 1997. Transcriptional Activation by Recruitment. *Nature* **386** : 569–577.

17

Genetics of Bacteria

The bacterial genome is devoid of a nuclear membrane. The region of a bacterial cell that contains DNA is called *nucleoid.* A single DNA duplex occurs in a convoluted form in the nucleoid, and represents the bacterial 'chromosome'. The bacterial cell is therefore a haploid organism. As the nucleoid is not bound by a membrane, it may lie in direct contact with cell ribosomes. In this way therefore, it differs from the nucleus of higher organisms.

The most widely studied bacterium, *Escherichia coli* (named after its discoverer Theodore Escherich, the specific name is derived from the colon where it resides), has a single large duplex, circular chromosome. The length of its DNA molecule is 1100 micrometer and its molecular weight about 2.6×10^9 daltons. Besides the single circular DNA molecule, *E. coli* also possesses one or more smaller chromosomes each called a *plasmid,* which may constitute 0.5 to 2 per cent of the cell DNA.

Bacteria reproduce by simple fission. They show a variety of phenotypes expressed mainly through mutations in the wild type such as resistance to drugs, antibiotics and bacteriophages, or the inability to utilise a given sugar or synthesise an essential growth substance. The nutritional mutants are referred to as *auxotrophs* and the wild type as *prototrophs*. Bacterial mutants are very useful genetic materials and easy to work with.

It was found by Joshua Lederberg in 1947 that the wild type *E. coli* can be cultured on a minimal medium containing glucose as a carbon source, KH_2PO_4, K_2HPO_4, sodium citrate, $MgSO_4$ and $(NH_4)_2 SO_4$. The normal *E. coli* can utilise these simple inorganic substances for synthesising all the compounds necessary for its growth. If a mutation in *E. coli* impairs its ability to synthesise a particular amino acid or a vitamin, the mutant strain will not be able to grow on the minimal medium unless the specific metabolite or growth factor is supplemented in the medium. For example, if a strain has a mutation due to which arginine biosynthesis cannot occur, this mutant strain will grow on a medium to which arginine is added. Mutant strains in *E. coli* are useful for studying bacterial recombination. Recombinants can be isolated from a bacterial culture and grown on a selective medium to form colonies.

Genetic exchange is achieved in bacteria by a one way transfer of DNA from a donor cell to a recipient cell. There are three ways of achieving transfer of DNA namely transformation, conjugation and transduction. In *transformation* free DNA is extracted from donor cells and added to living recipient cells which have the ability to take up DNA from the solution. As DNA molecules become fragmented during the process of extraction, the recipient cell generally receives only a small fragment of the donor genome.

In *conjugation* there is transfer of the extrachromosomal genetic material in a plasmid. In *E. coli* such a conjugative plasmid is called the *F* factor. Transfer takes place by direct contact between donor and recipient cells.

In *transduction* fragments of donor DNA are accidentally packaged within the protein coats of the bacteriophage. They are then transferred into recipient cells through infection by phage particles. Under certain conditions, segments of donor DNA are covalently joined to phage DNA which are then replicated along with phage DNA (specialised transduction).

TRANSFORMATION

The phenomenon was first discovered in 1928 from Griffiths experiment with *Diplococcus pneumoniae.* As this historical experiment has been instrumental for identification of DNA as a hereditary material, it is detailed in Chapter 13. Briefly, when a suspension containing a mixture of heat-killed virulent, encapsulated cells and *live,* non-virulent, nonencapsulated cells was injected into a mouse, a small fraction of the live bacteria became transformed into the virulent encapsulated type. The transforming ability was inherited by the descendants of the newly transformed live strain. It was inferred that when cells of the virulent strain are killed by heat, their chromosomal material which is somehow liberated from heat-killed cells can pass through the cell wall of the living cells and become incorporated in the host chromosome. Although this experiment involved genes that control presence/absence of the capsule, later on genes controlling other characters could also be transformed by addition of chromosome fragments.

Transformation has proved useful in locating genes in bacteria like *Bacillus subtilis, E. coli, Hemophilus influenzae, Rhizobium, Neisseria* and others. Attempts have been made to find out if transformation occurs in higher organisms including mammals and man. All results have been so far negative, except for some special cases where viral chromosome, for example of SV40 (simian virus), which can transform normal human cells in culture into cancerous ones. The mechanism of transformation is not fully known.

The Process of Transformation

When recipient cells are grown in presence of killed donor cells, transformation is observed. The DNA of donor cells is transferred to recipient cells where it undergoes genetic exchange with recipient chromosomes to produce recombinant progeny. Analysis of the process indicates that successful transformation depends upon several factors: size of donor DNA fragments which varies in different species of bacteria; molecular configuration of donor DNA which must be double stranded; physiologically competent state (described below) of recipient cells which occurs over a limited period in the growth of a culture and the ability to achieve this state is an inherited character; the amount of DNA added per recipient cell, *i.e.,* the frequency of transformation increases with the concentration of DNA up to the point where 10 molecules of DNA per cell are present. Further increase in concentration of DNA seems to have no effect.

Competence

When a recipient cell is able to absorb donor DNA and become transformed, it is said to be a *competent cell.* The development of the competent state appears to be related with cell density. Thus most cells growing in culture become competent when a critical number of cells is attained. Competence therefore, represents a transient phase in the life of a population. Its time of occurrence and duration are characteristic for a bacterial genus. As competence is acquired by cells in culture, a protein called competence factor is produced which confers competence on other cells. This

factor seems to act by changing the cell surface properties either by formation of receptor sites, or increased permeability to donor molecules. Cyclic AMP is also found to play a role in the development of competence. When added to the medium, this compound greatly increases the level of competence among the cells.

Uptake of DNA

The double stranded donor DNA molecules bind to the receptor sites on the recipient cell surface. Both homologous DNA and DNA from an unrelated species will be taken up by *Pneumococcus* whereas *Hemophilus* will take up only homologous DNA. The donor fragments are cleaved by endonuclease on the surface of the recipient cell to a size which varies in different bacterial species.

After attachment to the recipient cell wall, the donor DNA is actively transported inside the cell. Soon after uptake, one strand of the donor DNA fragment is degraded so that it becomes *single stranded* (Fig. 17.1). Immediately there is no transforming activity (eclipse period). Eventually the fragment pairs with that region of the recipient cell chromosome with which it is homologous. Genetic exchange takes place and a single strand of donor DNA carrying one or more genes from the donor cell becomes integrated in the homologous portion (having corresponding sequence) of recipient DNA. The single stranded segment which breaks of from the recipient DNA is degraded in the cell and lost.

That transformation is a reversible process can be demonstrated experimentally. If donor DNA fragment contains a hypothetical gene t^- and the recipient t^+, the transformed bacteria are found to contain t^-. When these t^- bacteria are used as recipients for donor t^+ DNA, the resulting bacteria become t^+ again.

Transfection

Studies with *Bacillus subtilis* have shown that when DNA from an animal virus or bacteriophage is used in transformation, intact virus particles are formed inside the recipient bacterial cell. The process is called *transfection*. In this case there is no need for donor DNA to become integrated into the host chromosome. When such a bacterium comes in contact with the animal host which the virus is able to infect, it releases the contained virus particles causing infection of the host animal. Experimentally transfection can be assayed by formation of plaques when infected bacteria burst to release the virus progeny.

Linkage and Gene Mapping by Transformation

Fragments of donor DNA which are involved in transformation can be used for detecting linkage and gene order in bacteria. The method consists of counting the number of double transformants (that is cells transformed for two genes) as well as single transformants produced by a single gene.

Suppose two genes *E* and *F* are placed distantly apart on the bacterial chromosome. The probability of both occurring together in the same fragment and producing double transformants is quite low. But a cell can become doubly transformed if it receives two separate donor fragments, one carrying *E*, the other *F*. The probability for such an event would equal the product of their separate probabilities in producing single transformants for *E* and for *F* and would be lower than the single events. But if *E* and *F* genes are closely linked, the probability that both are present on the same fragment and produce double transformants is high. When the experiment is performed, the number of single and double transformants will also depend upon the concentration of DNA containing donor fragments that are given to the recipient cells.

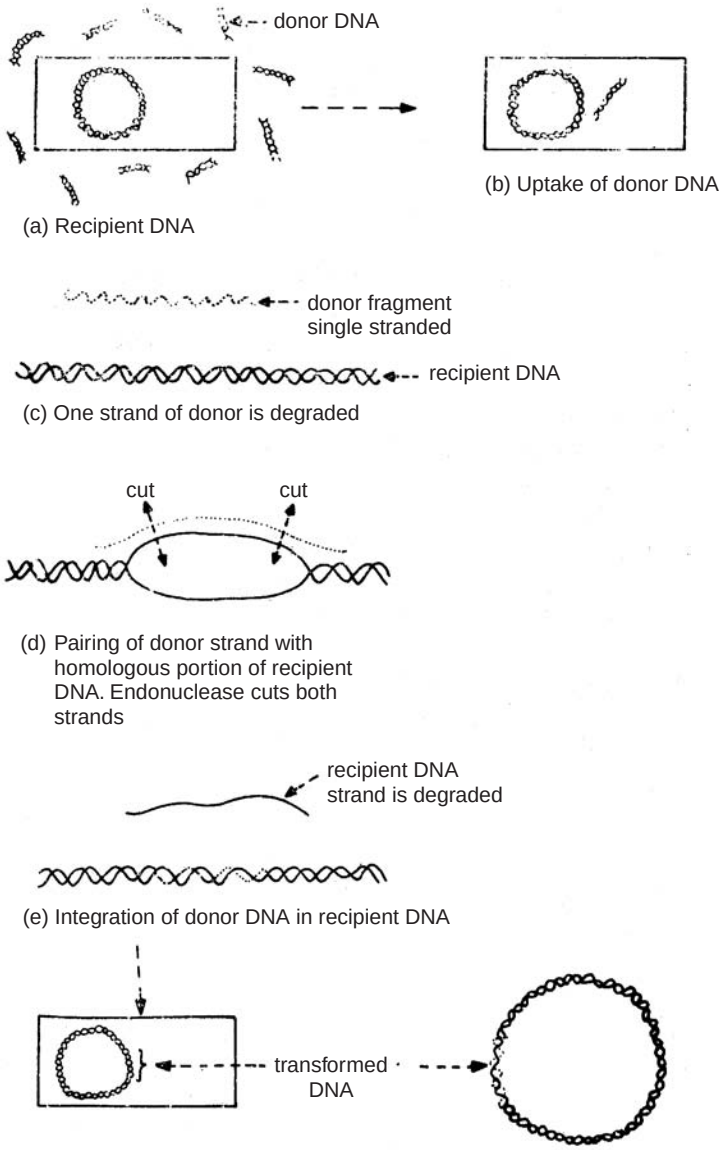


Fig. 17.1 The mechanism of transformation in *E. coli*.

A graph can be plotted to illustrate the curves for single and double transformants with decreasing concentrations of transforming DNA. If genes *E* and *F* are linked, the curve for double transformants for *E* and *F* must be similar to the curve for single transformants for *E* and for *F* (*see* Hayes, W. The genetics of bacteria and their viruses, 1968). By using larger fragments of donor DNA it is possible to map gene loci in the vicinity of *E* and *F*, and also other genes in the genome. Three gene mapping, similar to the three point cross in higher organisms is also done for mapping genes in bacteria.

CONJUGATION

The occurrence of recombination by sexual union was first shown experimentally in 1946 by Joshua Lederberg and E.L. Tatum in *E. coli*. They took two auxotrophic strains neither of which could grow on minimal medium due to mutation in genes controlling synthesis of vitamins thiamine and biotin, and amino acids methonine, threonine and leucine. For simplicity strain I can be designated as $a^- b^- c^- d^+ e^+$ and strain II as $a^+ b^+ c^+ d^- e^-$. A mixture of the two auxotrophic strains was cultured together on a complete medium and samples taken and plated on minimal medium. Surprisingly prototrophic colonies in a frequency of 1 per 10^6 or 10^7 cells plated were found growing on the minimal medium.

Lederberg argued that the prototrophs would have a genotype of the wild type $a^+ b^+ c^+ d^+ e^+$. The question arose on the origin of the prototroph colonies whether from mutation, transformation or recombination by some form of sexual union. Mutation was ruled out because it is improbable for so many gene loci in each strain to undergo mutation simultaneously. Transformation was negated experimentally when broken DNA fragments from either strain failed to produce recombinants. It appeared therefore that some form of sexual union between living cells had produced the wild type genotype.

Further experiment by Hayes, a British geneticist, and by Jacob and Wollman, two French geneticists gave better insight of this process when they found that genetic recombination takes place in bacteria as a one-way transfer of genetic material from a male type donor to a female type receptor and the process was termed *conjugation*.

The F Factor in Conjugation

As mentioned earlier, bacterial cells may carry besides the main chromosome, one or more small DNA molecules in the cytoplasm called plasmids. Of the various kinds of plasmids, a few are involved in conjugation and are called conjugative plasmids. One such conjugative plasmid is the sex element or fertility or *F* factor. The presence or *F* factor in different strains has given rise to two mating types in bacteria namely, the donor which possesses the fertility factor and referred to as F^+ strain, the second which lacks *F* factor is the F^- strain. The *F* factor is itself the genetic element which is passed from donor to recipient cells during conjugation. There is no conjugation between two F^+ strains or between two F^- strains.

The F element contains about 2 per cent of the cell's total DNA. It is capable of autonomous replication. It is made up of a circular, double stranded DNA molecule of molecular weight approximately 35×10^6. It contains about 15 genes, 8 of which control the formation of *F*-pili or sex pili which are hair-like appendages extending from the surface of F^+ cells. *F* pili function in conjugation.

Structure of Pili

The *F* pili originate from cell membrane and project outward beyond the cell wall (Fig. 17.2). The width of pili in different bacteria varies between 4 and 35 nm. The pilus is made up of a phosphate-carbohydrate-protein complex with a single polypeptide subunit called pilin of 11,000 to 12,000 daltons. Each pilin subunit has 2 residues of phosphate and one of glucose. The pili have been isolated and analysed by electron microscopy and X-ray diffraction techniques. The *F* pilus consists of a hollow cylinder 80 Å in diameter. The central hollow core is about 20 Å. The pilin subunits are arranged in the form of four helical chains.

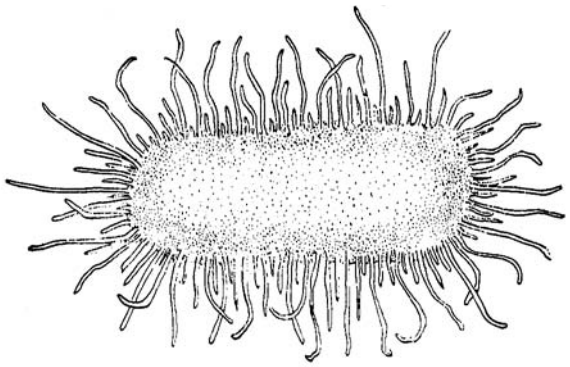


Fig. 17.2 Diagram of an *E. coli* cell showing pili on the surface.

The Mating Process

In a mixture of F^+ and F^- cells an F^+ donor cell establishes contact with an F^- recipient cell by the F pilus. The pilus is essential for recognition of recipient cell with which mating would take place.

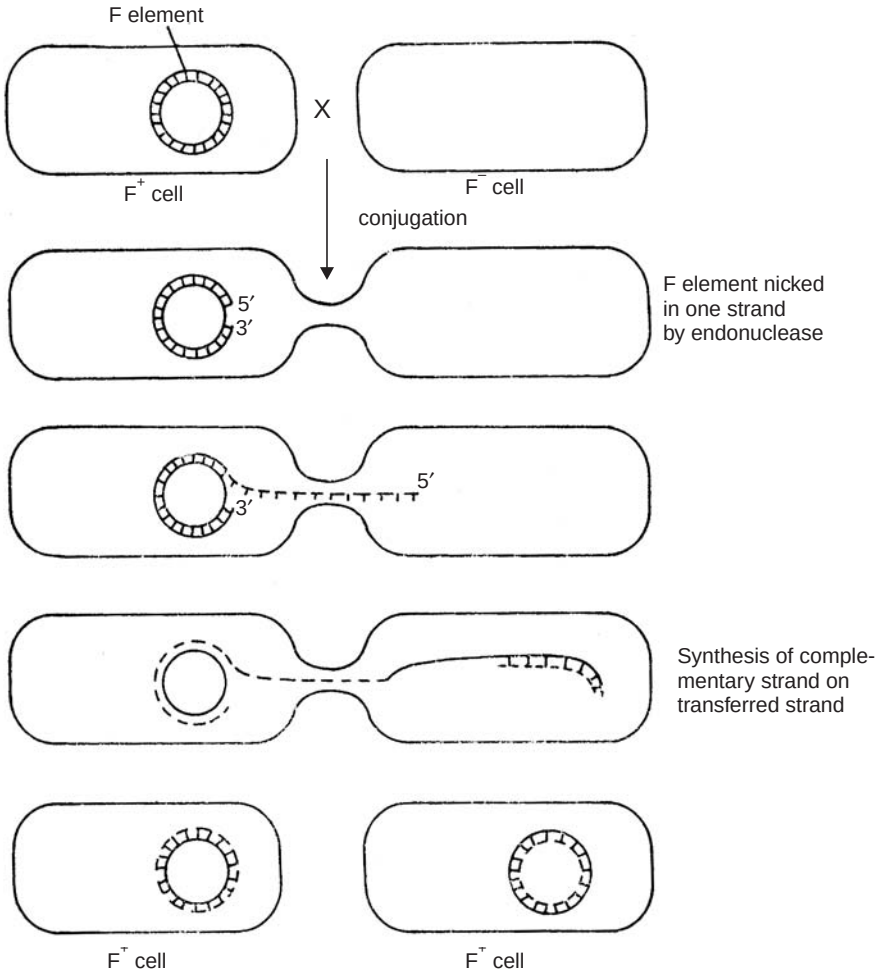


Fig. 17.3 A. Diagram showing transfer of F element from F^+ to F^- cell during conjugation.

After initial contact between the pilus and recipient cell is established, the pilus serves as a protoplasmic connection between the two cells and is called the *conjugation* tube. A donor cell devoid of pili cannot conjugate. The sex plasmid passes from the donor F^+ to the recipient F^- cell through the conjugation tube. Transfer takes place when DNA replicates by the rolling circle method (Fig. 17.3A).

A nick produced by an endonuclease in *one* strand of the plasmid DNA duplex produces a free 5′ and a 3′ end. The strand moves across the cytoplasmic bridge with the 5′ end first, into the F^- cell. The second inner strand of the plasmid DNA duplex is retained in the F^+ cell and synthesises its complementary strand. The two cells separate after mating and are known as exconjugants. Thus the originally mixed population of F^+ and F^- cells comes to have all F^+ cells only.

The transfer of sex element from F^+ to F^- cell has one more important feature. Not only can the plasmid exist in the cytoplasm as an autonomous entity, but it can also become incorporated in the main bacterial chromosome in a frequency of about 1 in 10,000 F^+ cells. Integration takes place at a specific site in the host chromosome which has homologous sequences. Such an integrated plasmid is known as *episome* and promotes the transfer of the main bacterial genophore from donor F^+ to recipient F^- cells during conjugation, an event followed by recombination.

High Frequency Recombination

Sometime after the discovery of F^+ strains, a special kind of strain was noticed which was several hundred times more fertile in crosses with F^- than any known F^+ strain. This strain was isolated by Lederberg *et al.* (1952) and was called Hfr or *high frequency recombination*. The Hfr strain produces about 1,000 times more prototrophs than in the $F^+ \times F^-$ cross. In the mating system of Hfr strains the main bacterial chromosome containing an integrated *F* factor is transferred to F^- cells. The Hfr bacteria arise spontaneously from F^+ cells at a low frequency by integration of *F* factor in the main chromosome. When Hfr cells are mixed with F^- cells there is conjugation and a high frequency of transfer of only portions of the main bacterial chromosome (some selected markers) from donor to F^- recipient cells. The recipient cell remains F^-. An F^+ cell is converted to Hfr when *F* integrates into the main chromosome by reciprocal recombination. The process is reversible so that an Hfr cell becomes F^+ when another recombinational event causes detachment of the *F* factor.

Hayes, Wollman and Jacob conducted experiments which demonstrated recombination during mating between Hfr and F^-. They took wild type Hfr capable of synthesising all its organic requirements, which could also utilise the sugars galactose and lactose, and was susceptible to being killed by streptomycin. The second strain they took was of the F^- type which could not synthesise some amino acids (leucine and threonine), nor utilise galactose and lactose, and was resistant to streptomycin. The Hfr and mutant F^- strains were mixed and grown together. For analysing the progeny cells, samples of the cell mixture were grown on minimal medium containing streptomycin. Recombinants had appeared in the progeny.

Linear Chromosome Transfer by Hfr Strains

Wollman and Jacob (1956) studied kinetics of genetic transfer by the *interrupted mating technique*. After mixing up two parental populations, cell samples are withdrawn at different intervals and agitated in a blender so that the mating pairs become separated and conjugation comes to an end. The mixture of cells is diluted and plated on selective medium and the number of recombinants formed during that interval are determined. The appearance of recombinants indicates formation

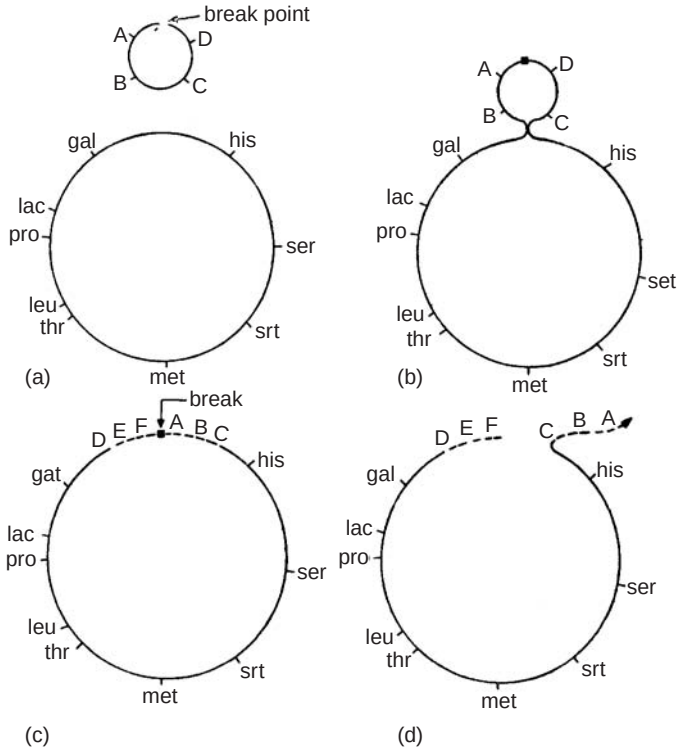


Fig. 17.3 B. Integration of F element into main chromosome of F^+ to form Hfr. d shows break at specific site and transfer from donor Hfr cell.

of zygotes. The different genetic markers are found to appear in the progeny of interrupted matings after different time periods have been allowed before mating is disrupted by agitation. Closely linked markers appear at the same time, whereas distantly placed markers appear at different times. The markers threonine and leucine appear after about 8 minutes whereas *gal* appears after 26 minutes. The entire chromosome containing about 5×10^6 base pairs is transferred in 90 minutes. In this way it is possible to map locations of markers on the donor chromosome.

Further investigations showed that the Hfr donor cells transfer only a part of their genome to F^- cells. Moreover, there are different Hfr strains which are distinct from each other in transferring a different part of the genome to F^- cells. The *E. coli* genome is a closed circular loop. In the Hfr donor cell the loop is broken at a point characteristic for that strain. The break occurs within the *F* element so that a part of it is at the leading end and is transferred to the F^- cell, the other part of *F* is at the extreme distal end which trails behind. Transfer takes place by injection of the linear structure into the recipient cell. The foremost or leading end carries with it gene loci nearest to it (Fig. 17.3B) until conjugation is interrupted. The transfer of DNA may be broken off at any time due to spontaneous rupture of the connection between conjugating cells.

After transfer to the F^- cell the donor DNA fragment becomes incorporated into a homologous region of the host cell chromosome. The corresponding segment of theF^- cell DNA is lost. Crossovers occur between the donor Hfr fragment and the F^- host cell chromosome. This integration is essential before donor genetic markers can express themselves.

The presence of *F* factor in Hfr and F^+ cells endows specific surface properties through formation of *F* pilus, due to which these cells can act as donors. In 1960 Loeb found out that

certain bacteriophages could lyse only donor *E. coli* cells but not the recipient cells. The phages R17 and M12 adsorb to pili present on donors but not on recipient cells and are referred to as *male-specific phages*. Further, the male-specific RNA phages are observed to adsorb along the length of the sex pilus, whereas male-specific DNA phages adsorb to the tip of the pilus.

The surface of the recipient cell also appears to play an important role in mating. When an *F* factor is present in a cell, it prevents the cell from acting as a recipient, so that superinfection does not occur. As this effect is due to a surface component which depends on the *F* factor, the phenomenon is known as *surface exclusion*. Some of the surface proteins coded by the main bacterial chromosome genes are also involved in mating. The *con*$^-$ mutants of *E. coli* are not able to function as recipients and form mating pairs in conjugation. These mutants are found to be deficient in two of the surface proteins.

Although mating usually takes place between pairs of bacterial cells, Achtman (1975) found the presence of cell aggregates of 2–20 cells (called *mating aggregates)* which were involved in transferring DNA from donor to recipient cells.

The F′ Factor

The *F* element can also become integrated into the main bacterial chromosome. Rarely an integrated *F* can undergo excision and become detached, carrying with it some bacterial genes that remain attached to it. Such an *F* element is called an *F′* factor. It behaves like the *F* factor of F^+ cells and can be transferred to F^- cells. Because *F′* carries bacterial genes, it is able to pair with the corresponding region in the bacterial chromosome (Fig. 17.4). A bacterium receiving an *F′* factor becomes a partial diploid for the bacterial genes carried by *F′*.

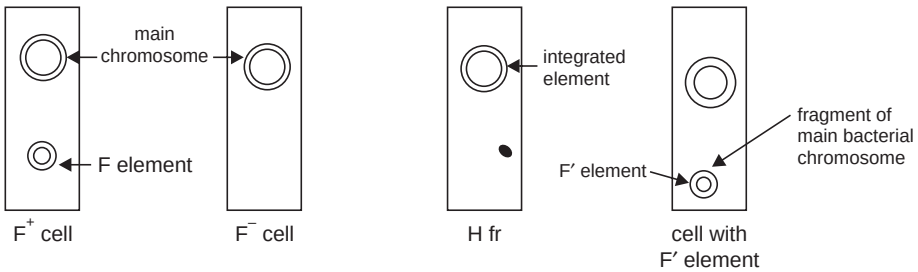


Fig. 17.4 The genomes of *E. coli* cell.

The Transfer Genes

There are certain mutant strains in which transfer of *F* factor cannot take place. The transfer deficient mutants have been useful for identifying the presence of transfer (*tra*) genes in the *F* factor. Transfer genes are found to be necessary for conjugation. About 19 *tra* genes have been identified so far and these are classified into 4 groups. Those in the first group control pilus formation and recognition of recipient cell. The second group genes are involved in stabilisation of mating pairs. Genes of the third group are required for some metabolic changes in DNA necessary for conjugation. There is only one gene in the fourth group (*tra J*) which controls the function of all the other *tra* genes.

TRANSDUCTION

The transfer of a small part of a bacterial genome from a donor to recipient bacterium through the agency of a bacteriophage is called *transduction*. It may be general or specialised. In generalised

transduction almost any part of the donor's chromosome may be transferred. But when it is specialised, as in the case of lambda prophage in strain K12 of *E. coli*, only specific loci (*gal* and *bio* genes) are transduced.

Zinder and Lederberg (1952) showed for the first time that in *Salmonella typhimurium* a bacteriophage of the type P22 could transfer a portion of the bacterial chromosome from one bacterial strain to another. The phages involved in transduction are all *temperate*, that is, those which either lyse the cell at once or remain in the host cell without killing it. They can multiply inside the bacterial cell and finally cause *lysis* of the host cell. A bacterium which harbours temperate phages is said to be *lysogenic*.

Generalised Transduction

Let us consider the transfer of a tryptophan synthesising gene from a wild type donor to a mutant strain of *E. coli* lacking this gene. For transduction the donor bacterial strain is first injected with the suitable phage. Within a short time some cells lyse on their own. The remaining bacteria are killed by addition of chloroform. The phage particles are separated from the bacterial cell remains by centrifugation and used to infect the recipient bacterial cells which are not able to synthesise tryptophan. After the phage is taken up by the mutant bacterial cells, the cells are spread on an agar medium which does not support the growth of the mutant strain. The transduced cells are able to grow on this medium due to the acquisition of the tryptophan synthesising gene through the transducing phage.

Usually only one marker from the bacterial genome is carried by the phage particle. It is possible for two markers to be transduced at a time when the two are closely linked. The process is known as co-transduction and is utilised for mapping genes in the bacterial genome.

Specialised Transduction

This is exemplified by the temperate phage lambda (λ) which exists in a latent form in the *E. coli* strain K12. It is capable of transferring only *gal* (galactose synthesis) and *bio* genes (cluster of genes controlling synthesis of the vitamin biotin). It was discovered by Appleyard (later confirmed by Jacob and Wollman) that the lambda genome becomes *integrated* into the bacterial chromosome through the process of recombination. It is excised from the bacterial chromosome along with *gal* and *bio* genes both of which lie on adjacent sites at its two attachment ends (Fig. 18.3). These bacterial genes will be sent into the next bacterial host resulting in specialized or restricted transduction of the *gal* and *bio* genes.

The temperate phage lambda (λ) attacks *E. coli* cells that are not lysogenic for lambda. After entering the host *E. coli* cell it may either become lytic and produce more than a hundred progeny phage particles which are released when the host cell becomes lysed. Or, alternatively, the single stranded open ends of lambda DNA can become ligated to produce a double-stranded ring (Fig. 18.3). This closed ring becomes integrated into a special attachment site in the host cell chromosome by crossing over between the phage DNA and the bacterial chromosome. In the integrated state the phage chromosome is called the *prophage;* it replicates along with the host cell chromosome and some of its genes are expressed. One of the prophage genes produces a repressor which inhibits its own excision from the chromosome and free multiplication, thus conferring immunity to lysis. The repressor also prevents superinfection from another phage. If the concentration of the repressor in the cytoplasm falls below a critical level, it leads to excision

of the prophage and complete expression of the viral genome. Treatment with UV light and some drugs can also lead to excision of the lambda prophage from the host chromosome. When the excised prophage leaves the host chromosome it carries with it the *gal* genes or the host cell attached on one side and the *bio* genes on the other. This lambda is said to be defective as it is not able to complete lytic development due to defect in its own genome. It is called λ *dgal* (defective λ prophage carrying *gal* genes) or λ *dbio* when it carries the *bio* genes of the host. The defective lambda prophage then attacks a sensitive *E. coli* host and the λ *dgal* chromosome becomes integrated by reciprocal exchange between the *gal* regions of the prophage and host DNA (Fig. 18.3). The recipient bacterial cell thus becomes partially diploid for some genes that are transduced.

THE GENETIC MAP OF *E. coli*

Several methods have revealed that the genetic map of the main chromosome of *E. coli* is an unbranched circle with a circumference of 100 minutes (Fig. 17.5) which corresponds to the transfer time during conjugation. About 650 genes have been mapped. Some important features of the genetic map have been noted. In many cases functionally related genes occur together and form clusters. For example the genes involved in the catabolism of lactose and synthesis of amino acids tryptophan and histidine are all clustered together. On the contrary, some functionally related genes are placed far apart. For example, the genes related with catabolism of arabinose are present at three sites, and those concerned with biosynthesis of leucine occur at several different sites. The orientation of all genes is not the same. Genes are arranged in both clockwise and anticlockwise manner.

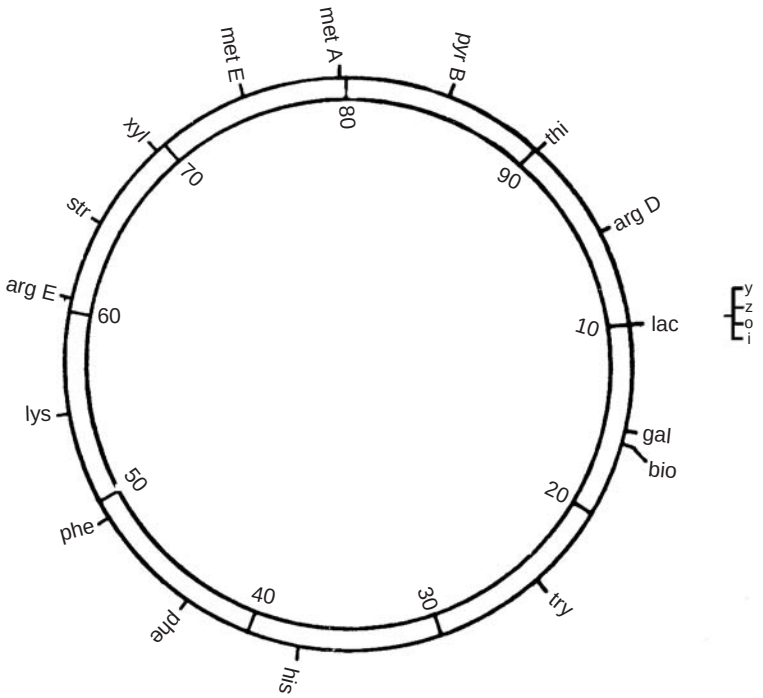


Fig. 17.5 The genetic map of *E. coli*.

PLASMIDS

They are dispensable, extra pieces of DNA found mostly in bacteria, that exist separately from the main bacterial chromosome. They are simpler in organisation than viruses, have no protein coat, and cannot exist outside a cell. They replicate independently within the host cell and are inherited in a regular manner when the cells multiply. First discovered about 30 years ago, they were later found to be of widespread occurrence in bacteria.

The presence of the plasmid appears to be not essential for any metabolic requirement of the host cell. As such, some cells of a plasmid-carrying strain may be devoid of a plasmid. The genes contained in plasmid DNA however, do provide accessory genetic information. The *F* factor for instance, is a plasmid that carries genetic information for conjugation or mating between bacteria. Some plasmids cause diseases in plants and animals; others confer to their host cells, resistance to antibiotics, toxic compounds and UV radiation. They carry genes for traveller's diarrhoea, staphylococcal impetigo, for fermentation of milk into cheese by lactic acid bacteria, for production of restriction and modification enzymes and for some metabolic functions of pseudomonads and soil bacteria.

One of the most important features of plasmids is that they replicate independently of the bacterial genome and segregate to the progeny when a bacterial cell divides. In this way they can be maintained indefinitely in a bacterial lineage. Most plasmids are circular DNA molecules, but some are linear. The number of copies of a plasmid in a cell is variable, depending on the mechanism by which replication is regulated. As many as 50 copies of a plasmid could be present in a host bacterial cell (high copy number), or just 1 or 2 plasmids in a cell (low copy number). Plasmids vary in size from a few kilobases to several hundred kilobases per cell. An *E. coli* cell could have small-sized plasmids up to 10 kb long, or large plasmids greater than 40 kb in length. In general, a typical *E. coli* cell contains 3 different small plasmids, each present in multiple copies per cell, and one large plasmid present in a single copy per cell. Plasmids can be visualised in electron microscopic images, or by performing gel electrophoresis on DNA samples. Plasmids can also be detected through a phenotypic characteristic they confer on the host cell, such as antibiotic resistance. For example, if a plasmid is carrying a gene for resistance to tetracycline (*tet-r* gene), the host bacterial cell will be able to form colonies on a medium containing tetracycline.

Although plasmids replicate independently of host cell DNA, they depend on the DNA replication enzyme of the host cell for their reproduction. But the initiation of replication is controlled by plasmid genes. In bacterial cells having high copy number plasmids, replication of plasmids is initiated multiple times during replication of host genome. But in low copy number plasmids, replication is initiated only once per round of replication of host genome. During division of host bacterial cell which takes place by fission, the segregation of plasmids to both daughter cells is promoted by sequences contained in the plasmid.

Preparation of Plasmids

Restriction enzymes are valuable genetic tools in that they can recognise a specific short sequence of bases and cleave DNA at a particular site within that sequence; wherever that sequence occurs in the DNA molecule it will be cut. In a pure sample of plasmid DNA therefore, a particular endonuclease will cut each DNA molecule at precisely the same site so that a large number of identical molecular fragments will be obtained. These fragments can be separated from each

other on the basis of their length by gel electrophoresis. For this the mixture of fragments is placed in a rectangular slot at one end of gel. When an electric current is passed, the DNA molecules move through the gel toward the positive electrode. Their rate of movement is inversely proportional to the logarithm of their molecular length; consequently, fragments of identical length form a narrow horizontal band. The positions of the bands are determined by first staining the gel with a fluorescent dye (ethidium bromide which binds to DNA) and photographing it under UV radiation.

The Major Groups of Plasmids

R Factors: In the 1960s Watanabe in Japan discovered that patients with bacterial dysentery did not respond to treatment with certain antibiotics. The bacterium causing this dysentery *Shigella dysenteriae* was found to contain genes that made it resistant to several antibiotics. Moreover, the resistance genes were transferred to other intestinal bacteria in the same way as the *F* factor. They were called '*R*' or resistance factors. The *R* factors were later found to be present in *Staphylococcus* and some enteric bacteria. Roth and Novick have found that the *Staphylococcus* plasmids carried genes for resistance to drugs like penicillin, chloramphenicol, tetracycline and to toxic metallic compounds of mercury and cadmium. The *R* factors consist of two parts, the basic resistance transfer factor (RTF), and a variable genetic determinant for antibiotic resistance (*r* determinant), which contains the genes for drug resistance (*R* genes). The genetic determinants (*r*) cannot be transferred unless they fuse with the RTF.

Col Factors: Similar to the *R* factors are the bacteriocins which are toxic proteins produced by certain bacteria that can kill closely related strains of bacteria. Bacteriocins which are produced by *E. coli* or related enteric bacteria are known as *colicins*. The genes for production of colicins are present in plasmids called Col (Colicinogenic factors). Depending upon the colicins they produce, Col factors have been designated Col *B*, Col E1, Col E2, Col I and Col V. Two of the Col factors produce sex pili and promote plasmid transfer. Colicins act by absorbing to specific receptors on the surfaces of sensitive bacteria causing modification or loss of the receptor sites.

The Degradative or Metabolic Plasmids: Some species of *Pseudomonas* inhabiting the soil have a unique characteristic. They are capable of breaking down organic compounds like hexane, phenol, xylene and camphor by special metabolic pathways. It was found out by I. M. Chakrabarty and his colleagues that the enzymes required for each metabolic pathway are under the control of plasmid genes (More details in Chapter 23). These plasmids may be transferred by conjugation, transduction or together with the *F* factor, and the recipient cell acquires the ability to control an entire metabolic pathway.

In the last few years, plasmids have been more extensively studied and manipulated with the idea of introducing non-bacterial genes into bacteria by the techniques of molecular cloning. This technique has revolutionised genetic analysis and given birth to the much publicised and equally criticised science of genetic engineering (Chapter 23).

MOVEABLE GENES

There are genes or segments of DNA that can become incorporated and function at a number of locations on the genome. The *F* factor can integrate at specific sites on the *E. coli* genome. Molecular studies have shown that the *F* element consists of three different functional blocks of genes. One region contains genes necessary for transfer of *F* element through conjugation from one bacterium to another. A second region controls autonomous replication of *F*. The third region contains a

number of different insertion sequences (IS). The integration of the *F* factor is brought about by a recombination between one of the IS sequences on the F factor and an IS sequence on the host chromosome.

The IS sequences are themselves moveable DNA segments which can be inserted at a large number of sites in different chromosomes. There are a number of IS sequences known of which 3, IS1, IS2 and IS3 have been studied in considerable detail. These range in size from 700 to 1400 base pairs. When an IS sequence is inserted into a gene, it breaks the continuity of the gene sequence, and may or may not inhibit expression of the gene. IS sequences also exert some effect on adjacent genes. More often the adjacent genes are inactivated; sometimes however, a previously silent gene could become activated. IS sequences appear also to be hot spots (vulnerable locations) for deletions; they are also involved in a number of recombination phenomena.

Another group of moveable DNA segments, many of which carry drug resistance genes are called *transposons.* They were probably first discovered in plants as controlling elements in maize, but were clearly demonstrated first in bacteria. In 1974 Hedges and Jacob found that when a gene for resistance to antibiotics like penicillin and ampicillin was transferred from one plasmid to another, it resulted in an increase in the size of the recipient plasmid.

The length of a typical transposon is several kilo bases, a few are much longer (details in Chapter 22). Much of the widespread antibiotic resistance among bacteria is due to the spread of transposons that contain one or more antibiotic resistance genes. When a transposon mobilises and inserts into a conjugative plasmid, it can be widely disseminated among different bacterial hosts by means of conjugation. Some transposons have composite structures with antibiotic resistance sandwiched between insertion sequences. Transposons are usually designated by the abbreviation Tn followed by an italicised number, for example Tn*5*. When it is necessary to include the name of the gene it carries in its designation, it becomes Tn*5* (neo-r, str-r) to reflect presence of genes for resistance to neomycin and streptomycin.

The Nonconjugative Plasmids

Besides the conjugative plasmid (*F* factor), there are nonconjugative plasmids. The conjugative and **nonconjugative plasmids** could coexist in the same cell along with host genomic DNA, and when a transposable element is mobilised, all the DNA molecules present are potential targets for insertion. Over a period of time, many of the plasmids in a bacterial lineage can acquire copies of transposable elements present in the host DNA, and the host DNA can acquire copies of transposable elements present in the plasmids. In this manner, the transposable elements become distributed among independently replicating DNA molecules. This results in most bacteria containing multiple copies of different types of transposable elements, some in host genome, some in plasmids, and some in both. For example, natural isolates of *E.coli* contain on an average one to six copies of each of six naturally occurring IS type transposable elements (Chapter 22).

Thus, many **nonconjugative** and **conjugative plasmids** present in the bacterial cell come to have one or more copies of the same transposable element. Because these copies are homologous DNA sequences, they can serve as substrates for recombination. When two plasmids undergo recombination in a region of homology, the recombination produces a composite plasmid called a **cointegrate**. If one of the participating plasmids is nonconjugative and the other is conjugative, then the cointegrate is also a conjugative plasmid and therefore, can be transferred in conjugation. After conjugation the nonconjugative plasmid can become free of the cointegrate by recombination between the same sequences that created it. Thus, the mechanism of cointegrate formation allows nonconjugative plasmids to temporarily ride along with conjugative plasmids and be transferred from cell to cell.

Site-specific Recombinases in Bacteria

During evolution of multiple antibiotic resistance, bacteria seem to have made use of enzymes known as site-specific recombinases. Each type of **site-specific recombinase** binds with a specific nucleotide sequence in double stranded DNA. When the site is present in each of the two duplex molecules, the recombinase brings the sites together and catalyses a reciprocal exchange between the two duplex molecules.

Site-specific recombinases are involved in the assembly of multiple antibiotic resistance units called integrons. An **integron** is a DNA element that encodes a site-specific recombinase and also has a recognition region that allows other sequences with similar recognition regions to be incorporated into the integron by recombination. The elements that integrons acquire are known as *cassettes*. With respect to integrons, a cassette is a circular antibiotic-resistance-coding region flanked by a recognition region for an integron. Because the site-specific recombinase integrates cassettes, the integron recombinase is usually called an **integrase**.

In general, antibiotic-resistance cassettes contain protein-coding regions but do not have the promoter sequences that are required for initiation of transcription. They can be transcribed only by read-through transcription from an adjacent promoter. The integron provides the required promoter called *Pant*, so that when a cassette is captured, the coding sequence can be expressed. More than 40 different promoterless cassettes have been described that encode proteins for resistance to antibiotics. Integrons by themselves are not capable of mobility. They are present in transposons, conjugative plasmids, nonconjugative plasmids as well as bacterial chromosomes. Integrons that are present in mobile DNA elements are particularly important in the evolution of antibiotic resistance because they can capture antibiotic-resistance cassettes, and thereby make possible transcription of the antibiotic resistance coding sequences as well as their mobilisation.

Bacterial Resistance to Antibiotics and Public Health

Over a period of time, a conjugative plasmid can accumulate different transposons containing multiple independent antibiotic resistance genes. The plasmid could also contain transposons having integrons that have acquired multiple antibiotic resistance cassettes. Such a plasmid is capable of conferring resistance to a large number of unrelated antibiotics. Such multiple resistance plasmids are called **R plasmids**. The development and evolution of R plasmids is associated with the use of antibiotics. Evolution selects for resistant cells because in the presence of antibiotics, growth of resistant cells has an advantage over that of cells that are sensitive. In the presence of multiple antibiotics, there is selection in favour of multiple drug resistance. When plasmids resistant to multiple drugs are transferred to pathogenic bacteria that cause human diseases, it results in serious clinical complications. Infections caused by pathogens that contain R factors are difficult to treat owing to pathogen's resistance to available antibiotics.

GENOMES IN MITOCHONDRIA AND CHLOROPLASTS

In Chapter 9 we had described the phenomenon of extranuclear inheritance based on transmission of visible phenotypes through mitochondria and chloroplasts. Studies in the 70s revealed presence of DNA in these organelles. Both mitochondria and chloroplasts are present only in cells of lower and higher *eukaryotic* organisms. Detailed studies established that DNA in these organelles is similar to the DNA in *prokaryotic* bacteria.

The genomes of both mitochondria and chloroplasts code for all of their RNA species and some proteins that are involved in the function of the organelles. The DNA is in the form of a

circular duplex molecule, except in some lower eukaryotes in which mitochondrial DNA is linear. Each organelle contains several copies of the genome, and because there are multiple organelles per cell, organelle DNA constitutes a repetitive sequence. **Mitochondrial DNA** (*mt*DNA) varies enormously in size, whereas **chloroplast DNA** (ctDNA) ranges in size between 120 and 200 kb.

Chloroplast DNA: Chloroplasts are present in green plants and photosynthetic protists. *ct*DNA sequence studied in a number of plants indicates uniformity in size and organisation. The differences in size are due mainly to the differences in lengths of introns and intergenic regions as well as the number of genes. All *cp*DNAs contain a significant proportion of noncoding DNA sequences. The *ct*DNA is double stranded circular, and devoid of histones and other proteins. In many cases, the GC content of *cp*DNA differs from that of nuclear DNA and mitochondrial DNA. Complete *cp*DNA sequences have been determined in tobacco (155,844 bp) and rice (135,42 bp). Multiple copies of *cp*DNA are present in the nucleoid region of each chloroplast. In the green alga *Chlamydomonas*, one chloroplast contains 500 to 1500 *cp*DNA molecules. Chloroplasts divide by growing and then dividing into two daughter chloroplasts.

The proportion of introns in chloroplast DNA could be high, 38% in *Euglena*. Among the expressed genes in chloroplast genome, 70 to 90% of the genes encode proteins including those involved in photosynthesis, *four* genes code for rRNAs (one each for 16S, 23S, 4.5S and 5S), and about 30 genes encode *t*RNAs. Chloroplast genome also contains genes for some of the proteins required for transcription and translation of the encoded genes, and most importantly, genes for photosynthesis. Most of the proteins in chloroplasts are encoded by the nuclear genes. The *m*RNA transcripts of the chloroplast genes are translated according to the standard genetic code. However, the primary structures of several RNA transcripts are found to go through editing consisting of C to U transitions, that cause *m*RNA sequence to deviate from the sequence in the corresponding gene. Editing makes it difficult to convert chloroplast nucleotide sequences into amino acid sequences of the corresponding protein.

Most of the cpDNAs studied share a common feature, that is, a 10 to 24 kb segment present in *two identical* copies as an inverted repeat. The *cp*DNA also contains *two copies* of each of the *r*RNA genes which are located in these two identical repeat sequences in an inverted orientation. Other genes that are found in the repeated sequence are therefore, also duplicated in the chloroplast genome. The location of these repeats defines a short single copy (SSC) region and a long single copy (LSC) region in chloroplast genome.

Chloroplast protein synthesis uses organelle-specific 70S ribosomes consisting of 50S and 30S subunits. The 50S subunit contains one copy each of 23S, 5S and 4.5S *r*RNAs, while the 30S subunit contains one copy of a 16S *r*RNA. Among the ribosomal proteins, some are encoded by the nuclear DNA, some by the chloroplast genome. About 100 open reading frames (ORFs), putative protein coding genes, have been identified by computer analysis. Protein synthesis is similar to that in prokaryotes.

Mitochondrial DNA: Each human cell contains hundreds of mitochondria each containing multiple copies of **mitochondrial DNAs** (*mt*DNA). Mitochondria generate cellular energy through the process of oxidative phosphorylation. As a by-product they produce most of the endogenous toxic reactive oxygen species Mitochondria are also the central regulators of apoptosis or programmed cell death (details in Chapter 21). These interrelated functional systems involve activities of about 1000 genes distributed in the nuclear genome and the mitochondrial genome. Due to their dependence on the nuclear genome, mitochondria are considered as semi-autonomous. This has been shown by experiments in which mitochondria and *mt*DNA could be transferred from one cell to another. The donor cell was enucleated and its mitochondria-containing cytoplast fused with a recipient cell (technique of cybrid transfer).

The genomes of mitochondria show wide variation particularly among plants and protists. Most mitochondrial DNAs (*mt*DNA) consist of a closed circular double stranded supercoiled DNA molecules located in multiple nucleoid regions (similar to those in bacterial cells); some protists however, have varying lengths or multiple circular molecules of DNA as in the trypanosomes. mtDNA in the protist *Amoebidium parasiticum* consists of several distinct types of linear molecules with terminal and subterminal repeats. Although most *mt*DNAs are in the size range of 15 to 60 kb, *mt*DNA in malarial parasite (*Plasmodium* spp) is only 6 kb long, while that of rice (*Oryza sativa*) is 490 kb, and cucurbits 2 Mb. There are about 40 to 50 coding genes in mitochondrial DNA, *Plasmodium* being an exception with 5 coding genes. The large size of mitochondrial genomes in plants are due to noncoding intergenic regions and their content of tandem repeats. Introns are present in many *mt*DNAs, and in some unusual cases, the genes are split into as many as 8 regions that are dispersed in the genome, and located on both strands of the DNA. Transcription takes place separately in portions of the split genes producing discrete pieces of RNA that are held together by base pairing of complementary sequences.

The *mt*DNA contains information for a number of mitochondrial compounds such as *t*RNAs, rRNA, and some of the polypeptide subunits of the proteins cytochrome oxidase, NADH-dehydrogenase and ATPase. Most of the other proteins found in mitochondria are encoded by the nuclear genome and transported into mitochondria. These include DNA polymerase and other proteins for *mt*DNA replication, RNA polymerase and other proteins for transcription, ribosomal proteins for ribosome assembly, protein factors for translation, and the aminoacyl-*t*RNA synthetases. The mitochondrial oxidative phosphorylation complexes are composed of multiple polypeptides, mostly encoded by the nuclear DNA (*n*DNA). However, 13 polypeptides are encoded by *mt*DNA. The *mt*DNA also codes for 12S and 16S rRNAs and 22 tRNAs required for mitochondrial protein synthesis. The *mt*DNA also contains a control region consisting of approximately 1000 base pairs constituting the promoter region and the origin of replication.

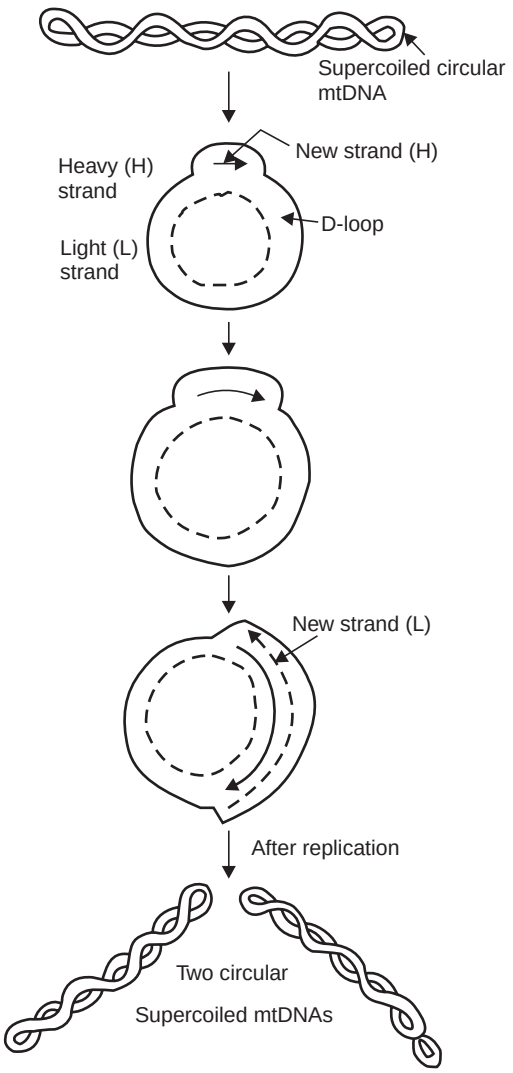
The mRNAs synthesised within the mitochondria remain *in the organelle* and are translated by mitochondrial ribosomes that are assembled within mitochondria. Mitochondrial ribosomes have two subunits. Mitochondria in human cells have 60S ribosomes consisting of a 45S and a 35S subunit. There are only two *r*RNAs in mitochondrial ribosomes of most organisms, that is, 16S *r*RNA in large subunit and 12S *r*RNA in small subunit of most animal ribosomes. There is usually one gene for each *r*RNA in a mitochondrial genome. The proteins in mitochondrial ribosomes are encoded by the nuclear genome and transported into mitochondria from the cytoplasm.

Mitochondrial ribosomes are sensitive to most of the inhibitors of bacterial ribosome function such as streptomycin, neomycin and chloramphenicol. For protein synthesis, mitochondria of most organisms use a genetic code that shows differences from the universal genetic code. Only plant mitochondria use the universal nuclear genetic code.

Transcription of mammalian *mt*DNA is unusual in that each strand is transcribed into a single RNA molecule that is then cut into smaller pieces. In the large RNA transcripts that are produced, most of the genes encoding the *r*RNAs and the *m*RNAs are separated by *t*RNA gene. The *t*RNAs in the transcript are recognised by specific enzymes and are cut out, leaving only the *m*RNAs and the *r*RNAs. A poly (A) tail is then added to the 3′end of each *m*RNA and CCA is added to the 3′end of each tRNA. There are no 5′ caps in mitochondrial mRNAs.

Mitochondrial DNA replication is semi-conservative and uses DNA polymerases that are specific to the mitochondria. The *mt*DNA replicates throughout the cell cycle, independently of nuclear DNA synthesis which takes place in S phase of cell cycle. Observations on *mt*DNA replication in animal mitochondria *in vivo* have resulted in a model referred to as the *displacement loop* (D loop) model as follows (Fig. 17.6). The two strands of *mt*DNA in most animals have

different densities because the bases are not equally distributed on both strands, called H (heavy) and L (light) strands. The synthesis of a new H strand starts at the replication origin for the H strand and forms a D-loop structure (Fig. 17.6). As the new H strand extends to about halfway around the molecule, initiation of synthesis of a new L strand takes place at a second replication origin. Synthesis continues until both strands are completed. Finally, each circular DNA assumes a supercoiled form.



Replication of mtDNA; D-loop model

Fig. 17.6 Model for mitochondrial DNA replication by formation of a D-loop structure.

The *mt*DNA is maternally inherited and has a very high mutation rate. When a new *mt*DNA mutation occurs in a cell, a mixed intracellular population of *mt*DNAs is generated, known as **heteroplasmy**. During replication in a heteroplasmic cell, the mutant and normal molecules are randomly distributed into daughter cells. When the percentage of mutant *mt*DNAs increases, the mitochondrial energy producing capacity declines, production of toxic reactive oxygen

species increases, and cells become more prone for apoptosis. The result is mitochondrial dysfunction. Tissues most sensitive to mitochondrial dysfunction are brain, heart, kidney and skeletal muscle.

The *mt*DNA mutations are associated with a variety of *neuromuscular disease* symptoms, including various ophthalmological symptoms, muscle degeneration, cardiovascular diseases, diabetes mellitus, renal function and dementias. The *mt*DNA diseases can be caused either by *base substitutions* or *rearrangement mutation*. Base substitution mutations can either alter protein (**missense mutation**) or *r*RNAs and *t*RNAs (**protein synthesis mutations**). Rearrangement mutations generally delete at least one *t*RNA and thus cause protein synthesis defects.

Missense mutations are associated with myopathy, optic atrophy, dystonia and Leigh's syndrome. Base substitution mutations in protein synthesising genes have been associated with a wide spectrum of neuromuscular diseases, and the more severe typically include **mitochondrial myopathy**.

Mitochondrial diseases are also associated with a number of different *nuclear* DNA mutations. Mutations in the RNA component of the mitochondrial RNAse have been implicated in metaphyseal chondrodysplasia or cartilage hair hypoplasia which is an autosomal recessive disorder resulting from mutation in nuclear chromosome 9 short arm position (9p13).

QUESTIONS

1. When an *F* type plasmid is transferred to F^- cell, the recipient F^- cell becomes F^+. But when a portion of the main chromosome containing an integrated *F* factor is transferred to an F^- cell, the recipient cell remains F^-. Explain.
2. Under what circumstances can (*a*) phage genes be present in *E. coli* genome; (*b*) *E. coli* genes be present in phage genome; (*c*) *E. coli* genes be present at different sites within the *E. coli* genome?
3. Give one example of dispensable *E. coli* genes controlling each of the following: (*a*) drug resistance; (*b*) structural modifications on *E. coli* cell surface; (*c*) disease to higher organisms.
4. Distinguish between (*a*) temperate and virulent phage; (*b*) plasmid and episome; (*c*) behaviour of *E. coli* cells of F^+ strain and Hfr strain.
5. Comment on (*a*) lysogenic bacteria, (*b*) specialised transduction; (*c*) insertion sequences.

SELECTED READINGS

Adelberg, E. A. 1966. Papers on Bacterial Genetics. Little Brown & Company, Boston.

Arber, W. *et al.,* 1957. The Defectiveness of Lambda Transducing Phage. (In French) *Schweiz. Zeitschr. Allgemeine Path. Bact.* **20:** 659.

Avery, O. T. *et al.,* 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. *J. Exp. Med.* **79:** 137.

Baltimore, D. 1971. Expression of Animal Virus Genomes. *Bact. Rev.* **35:** 235.

Borek, E. and Ryan, A. 1973. Lysogenic Induction. *Prog. Nucl. Acids Res. Mol. Biol.* **13:** 249.

Braun, W. 1965. Bacterial Genetics. Saunders. Philadelphia.

Burger, G., Gray, M.W. and Lang, B.F. 2003. Mitochondrial Genomes : Anything Goes. *Trends Genetics* **19:** 709–716.

Calvo, J. M. and Fink, G. R. 1971. Regulation of Biosynthetic Pathways in Bacteria and Fungi, *Annu. Rev. Biochem.* **40:** 943.

Chen, X. and Schnell, D.J. 1999. Protein Import into Chloroplasts. *Trends Cell Biol*. **9:** 222–227.

Copeland,W.C. 2002. Mitochondrial DNA. Humana Press, New Jersey.

Engler, J. A. and Van Bree, M. P. 1981. The Nucleotide Sequence and Protein Coding Capability of the Transposable Element IS5. *Gene* **14:** 155.

Ferri, K. F. and Kroemer, G. 2001. Mitochondria – The Suicide Organelles. *Bioessays* **23:** 111–115.

Fox, M. S. and Allen, M. K. 1964. On the Mechanism of Deoxyribonucleate Integration in Pneumococcal Transformation. *Proc. Natl. Acad. Sci. US.* **52:** 412.

Fredericq, P. 1963. On the Nature of Colicinogenic Factors: A Review. *J. Theoret. Biol.* **4:** 159.

Gross, J. D. 1964. Conjugation in Bacteria. *In* The Bacteria. Eds. Gunsalus, I.C. and Stanier, R. Y. Vol. **5.** Academic Press, New York.

Hayes, W. 1968. The Genetics of Bacteria and their Viruses. Wiley, New York.

Low, K.B. and Porter, D. D. 1978. Modes of Gene Transfer and Recombination in Bacteria. *Annu. Rev. Genetics* **12:** 241.

Kaiser, A. D. and Hogness, D. J. 1960. The Transformation of *Escherichia coli* with Deoxyribonucleic Acid Isolated from Bacteriophage *dg. J. Mol. Biol.* **9:** 392.

Kleckner, N. 1977. Translocatable Elements in Prokaryotes. *Cell* **11:**11.

Morse, M. L. 1954. Transduction of Certain Loci in *Escherichia coli* K–12. *Genetics* **39:** 984.

Radding, C. M. 1978. Genetic Recombination, Strand Transfer and Mismatch Repair. *Annu. Rev. Biochem.* **47:** 847.

Schoner, Brigitte and Khan, M. 1981. The Nucleotide Sequence of IS5 from *Escherichia coli. Gene* **14:** 165.

Siddiqui, O. and Fox, M. S. 1973. Integration of Donor DNA in Bacterial Conjugation *J. Mol. Biol.* **77:**101.

Sterlinger, P. 1979. Transposon and Insertion Sequences. *In* Recombinant DNA and Genetic Experimentation. Eds. Morgan, J. and Whelan, W. J. Pergamon Press, New York.

Sugiura, M. 1992. The Chloroplast Genome. *Plant Mol. Biol.* **19:** 149–168.

Willet, N. S. *et al.,* 1981. The Insertion Sequence IS21 of R68.45 and the Molecular Basis for Mobilisation of the Bacterial Chromosome. *Plasmid* **6:** 30.

18

Genetics of Viruses

Viruses were first discovered in 1899 when M. W. Beijerinck noticed the existence of micro-organisms invisible in the microscope, that could pass through filters that stopped bacteria. In 1917 Felix d'Herelle gave the name *bacteriophage* (meaning eater of bacteria) to one such microbe that was parasitic on, and capable of killing a certain rod-shaped bacterium. Work on phage genetics was initiated in 1930s independently by Max Delbrück, Martin Schlesinger and F. M. Burnet. Delbrück, who was educated as a physicist, is credited with the most important contributions in phage genetics. In the 1940s he, along with Salvatore Luria and A. D. Hershey discovered genetic recombination in phages. Thereafter, phages have been extensively used as tools for the study of gene structure and function.

Viruses are obligate intracellular parasites designated in relation to the host cell they parasitise. Thus we have bacteriophages parasitic on bacteria, fungal viruses on fungus and cyanophages on blue green algae. Structurally viruses are the smallest organisms (exceptions being the viroids described later) consisting of a nucleic acid (either DNA or RNA) and a protein coat or *capsid.* Inside the host cell they exist in the form of replicating nucleic acid molecules devoid of protein coats. The nucleic acid is double-stranded DNA or single-stranded RNA in most viruses. The phage ϕX174 and some coliphages have single-stranded DNA; the reoviruses and some plant viruses have double stranded RNA as their genetic material. All the DNA viruses and most RNA viruses have a *single* molecule of nucleic acid of constant length in circular or linear form. In different viruses the length varies from a few thousand nucleotides to over 250,000 nucleotides.

After penetration of host cells there is synthesis of enzymes for viral nucleic acid replication and for production of virus-specific proteins. Viruses also regulate the growth and differentiation of host cells by taking control of their metabolic machinery. In the following account the most studied bacteriophages and cancer viruses are described.

Phages consist of protein and nucleic acid. The much investigated T-phages infecting *E. coli* are named as T1, T2, T3, T4 ..... T7 (T stands for "type"). The T-even phages namely T2, T4 and T6 are closely related as they share some common features such as presence of 5-hydroxymethylcytosine instead of cytosine. The T-phages consist of a hexagonal *head* of protein containing a double helical DNA molecule. In the phage T4 the tail consists of two hollow tubes,

the outer *sheath* and the inner hollow *core* or *needle.* There is a *base plate* below the tail with *tail fibres* and *tail pins* (Fig. 18.1). The phage ϕX 174 is without tail and has a single stranded DNA molecule; lambda has a tail but no tail fibres or pins.

LIFE CYCLE OF VIRULENT BACTERIOPHAGES

Phages which lead to destruction or *lysis* of host cell are called *virulent.* The T-phages become attached (adsorb) to the host *E. coli* cell. The polysaccharide cell wall of the bacterium recognises specific phage types for attachment. The process of attachment involves uncoiling of tail fibers along the base plate. The tail sheaths contract and the needle penetrates the bacterial cell wall and DNA is injected.

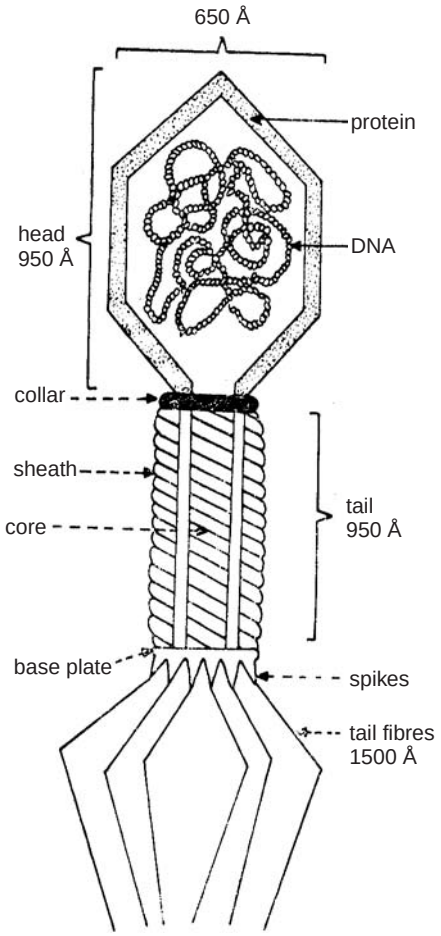


Fig. 18.1 The T-even phage particle.

Once inside the host cell, T4 phage DNA takes control of the host cell machinery. Normal activities of bacterial DNA stop, and synthesis of phage DNA and coat protein components begins. Host cell DNA disintegrates and disappears. These activities are regulated by three sets of phage

genes expressed in a sequence. First, the *early gene* produces viral messenger RNA. This is translated on host ribosomes to form protein products that inhibit *E. coli* RNA synthesis. The enzymes required for replication of phage DNA are also synthesised by the early genes. The second set of gene products release nucleases that digest the *E. coli* chromosome to provide free nucleotides for phage DNA synthesis. Third, the *late genes* control synthesis of phage coat proteins and tail components. The DNA is incorporated into the protein coats by a packaging process involving condensation of DNA (Fig. 18.2). In T-even phages replication of phage DNA produces

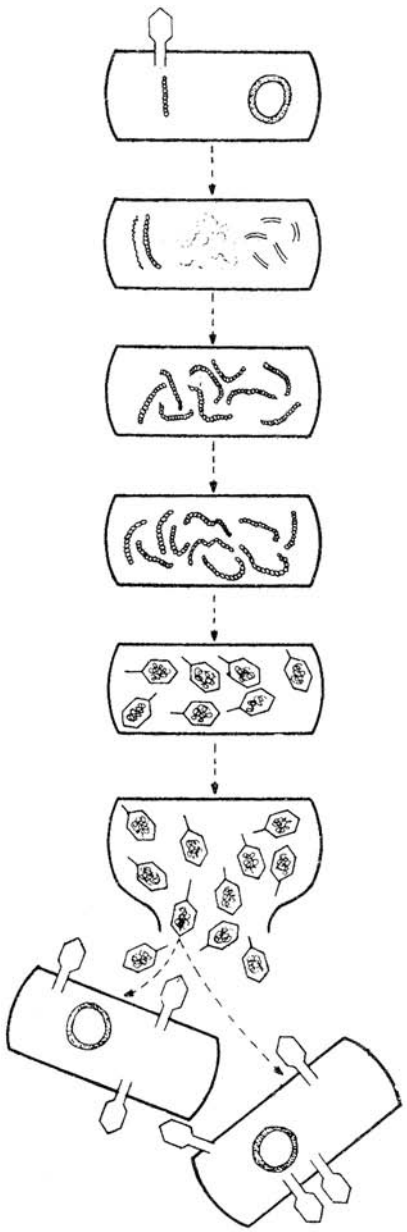


Fig. 18.2 The life cycle of T2 phage.

DNA molecules many times longer than the phage chromosome called *concatamers.* These are cleaved into smaller segments of required length so that each maturing phage head receives a headful of DNA. In all phages new phage particles are released by cell lysis, aided by the enzyme lysozyme which is synthesised under the direction of phage DNA. E. L. Ellis and Max Delbrück in 1939 showed that about 100 new phage particles are produced from a single virus within 30 minutes.

TEMPERATE PHAGES AND PROPHAGE

All phages are not virulent. Many phages infect and live inside the host cell without causing lysis or cell destruction. Such a nonvirulent virus is called symbiotic or *temperate phage.* Phage lambda (λ) is one studied in detail. It has a linear, double helical DNA molecule with a unique feature—the 3′ end of each strand is longer by 12 nucleotides than the 5′ end of the opposite strand. The nucleotide sequences of the single stranded extended regions are complementary to each other.

When phage lambda infects *E. coli* cell it can either cause host cell lysis (lytic response), or it may become integrated into the *E. coli* chromosome and enter the state of lysogeny (Fig. 18.3 A). The integrated viral DNA called *prophage* is replicated along with the host chromosome and is transmitted to the cell progeny. In case the prophage is excised from the host chromosome, it replicates and produces progeny virus particles which are liberated through cell lysis. That is why the term *lysogenic* is used for bacteria which carry such a virus. Alternatively, the integrated lambda genome can remain in the host cell as prophage for many generations.

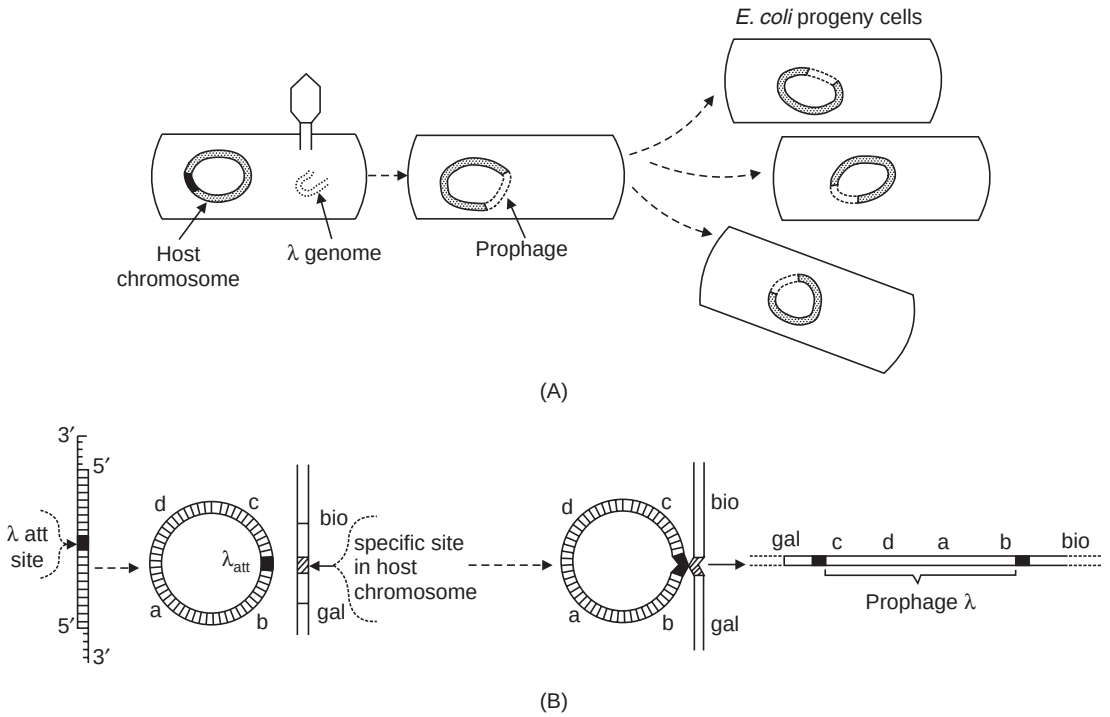


Fig. 18.3 A. The lysogenic life of λ. B. Integration of phage chromosome into *E. coli* chromosome by site-specific recombination to form prophage.

When the linear molecule of lambda DNA enters the cell, it becomes circular by forming H bonds between the base sequences of the extended region at the 3′ end followed by joining by ligase enzyme. There is a mechanism of phage repression which inhibits the lytic cycle and allows for the lysogenic state. The DNA of phage λ has a gene (the C1 gene) which produces a repressor protein. The λ repressor binds to a specific receptor site in the lambda DNA. The repressor is involved in the integration of the λ genome into the bacterial chromosome at the specific λ attachment site to form the prophage (Fig. 18.3 B). When the receptor site in λ DNA is not occupied by the repressor, the DNA is replicated and all the genes are expressed to form the viral proteins. But when the repressor molecule is bound to its receptor site, replication of DNA and expression of all genes (except Cl gene) are inhibited. The Cl gene continues to produce repressor protein when all the other λ genes are not functioning.

GENETIC RECOMBINATION IN PHAGES

Recombination in phages may be vegetative or *site specific*. Vegetative recombination occurs during the lytic response when the host cell contains a large number of phage DNA molecules. As described later, there are several rounds of replication of viral DNA and recombination occurs due to coming together of homologous sequences in the DNA molecules followed by breakage and exchange of sequences. In *site specific* recombination there is reciprocal exchange by breakage and reunion between a specific site on the phage DNA (*att* in lambda) and a site on the bacterial chromosome.

As in Mendelian inheritance, the study of genetic recombination in phages requires identification of phenotypes. Structural details of head and tail are phenotypes visible only in the electron microscope. Other characters like plaque size and shape and *host range* are more convenient for identifying normal and mutant phenotypes.

Plaque formation can be demonstrated by adding a solution of phage particles on the surface of nutrient agar on which sensitive *E. coli* cells are growing to form a colony or *lawn* of cells. Even if one virus particle infects a bacterial cell, it will multiply to form a hundred progeny particles which are released by cell lysis in 15–60 mins. The progeny particles attach to new bacteria and multiply. After several such cycles of attachment, multiplication and release, all the bacteria lying near the *original* virus particle are killed. This results in clear, circular zones called plaques in a lawn of healthy bacteria (Fig. 18.4). The morphology of the plaque is used to identify viral strains and mutants. For example the *r* (rapid lysis) mutant of T-phage produces a large-sized plaque with smooth edge. The normal, wild type phage (*r*+) produces small plaques with fuzzy outline.

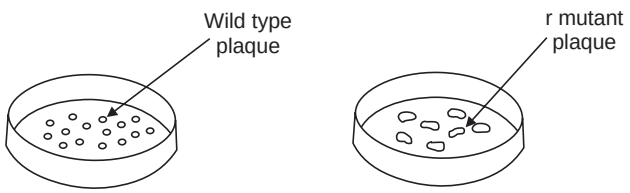


Fig. 18.4 Plaque types formed after infection of *E. coli*.

The *host range* mutants of phage are able to infect certain additional bacterial strains. Cells of *E. coli B* strain can become infected with phage *T*2. A mutant strain *E. coli B*/2 is resistant to infection by phage *T*2, because the *B*/2 mutation changes the cell surface thus preventing attachment of *T*2. But if a mutation occurs in *T*2, the mutant phage *T*2/*h* (*h* stands for

host range) can adsorb to and infect *E. coli B*/2 because it possesses altered tail fibers. Now *E. coli B*/2 can again become resistant to phage *T*2*h* if the cells undergo a mutation to become *E. coli B*/2/*h*. *T*2*h* phage can once again acquire ability to infect *E. coli B*/2/*h* after another mutation which changes it to *T*2*h* form.

The host range mutants (*h*) can be identified by adding the phage to a mixture of *B* and *B*/2 cells. The *h* mutants will infect both *B* and *B*/2 cells and form plaques distinct from those produced by wild type phage.

It is possible to have mixed infection with phages carrying two different mutations so that both types of phages enter the host bacterium. Due to the mutation, neither of the two phages can multiply inside the host cell to form plaques. But if the two mutations are for different functions, one will complement the other and progeny virus particles will form. This is known as *complementation test.*

Hershey and Rotman (1948, 1949) carried out the first detailed analysis of genetic recombination in phages by *single burst experiments.* By this method it is possible to analyse the progeny (recombinants and parentals) released from a single infected bacterial cell. In such an experiment bacteria growing in culture are exposed to mixed infection of two parental phage types. The infected bacteria are diluted to the extent that a single infected bacterium is present in a culture tube. Hershey and Rotman made crosses between wild type (*r* +, *h* +) and the *r* and *h* mutants of phage *T*2, and determined the frequencies of parental and recombinant types in the progeny. The cross consisted of parental phages with genotypes *hr* + and *h* + *r* used to produce mixed infection in *E. coli* cells. After cell lysis, progeny virus particles were added to a mixture of *B* and *B*/2 cells, and the plaques formed were identified to be of four types (Fig. 18.5). It was found that parental combinations appeared with higher frequency than recombinant phenotypes.

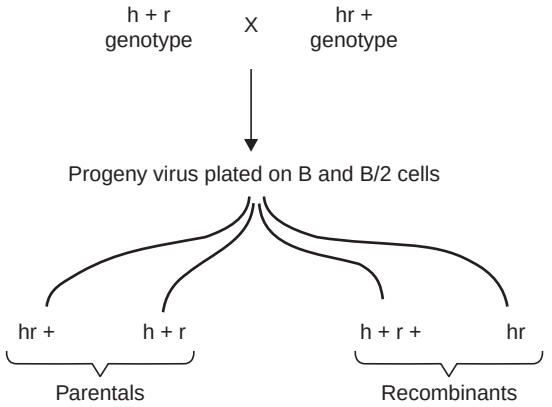


Fig. 18.5 Results of a cross between two mutant phages by mixed infection.

The experiments of Hershey and Rotman demonstrated genetic recombination in T2 phages. A striking feature was that in an individual cross (a single burst) the recombinant phenotypes did not appear with equal frequency. But when the average from many experiments was considered, the proportions of the two recombinant classes were roughly equal. The occurrence of one recombinant type more frequently than the other in the progeny of an individual cross indicates that there is little, if any, *reciprocal* exchange of genes. The percentage frequency of recombination between genes could be calculated as follows: no. of recombinant plaques/total plaques scored *x* 100. If we assume that the frequency of recombination between two genes is related to the distance separating them, then it is possible to map positions of genes on the phage chromosome. The data on recombination frequencies was utilised by Hershey and Rotman to assign *r* mutations to 3 different loci in relation to *h.*

The pattern of recombination in phage is significantly different from that in eukaryotes. In a Mendelian cross, reciprocal genetic exchange is restricted to two of the four nonsister chromatids, and occurs only once during meiotic prophase I. The two classes of recombinant progeny occur with equal frequency, and maximum frequency of crossing over between two loci never exceeds 50 per cent.

A distinctive feature of phage recombination is that pairing and exchange may occur repeatedly throughout the period when free chromosomes are present within the host cell; there are many cycles of phage DNA replication. Maximum amount of recombination between two genes is 30 per cent. When all the progeny arising from a doubly infected cell is analysed the reciprocal recombinants are not present in equal proportions.

That phage chromosomes undergo several successive rounds of exchanges can be shown by infecting a host bacterial cell with *three* genetically different phages such as *xyz* which is possible only if each of the three phages has contributed one gene. Figure 18.6 shows how separate exchange events can occur involving breakage and fusion in all the three phage types during the latent period of phage in the host cell.

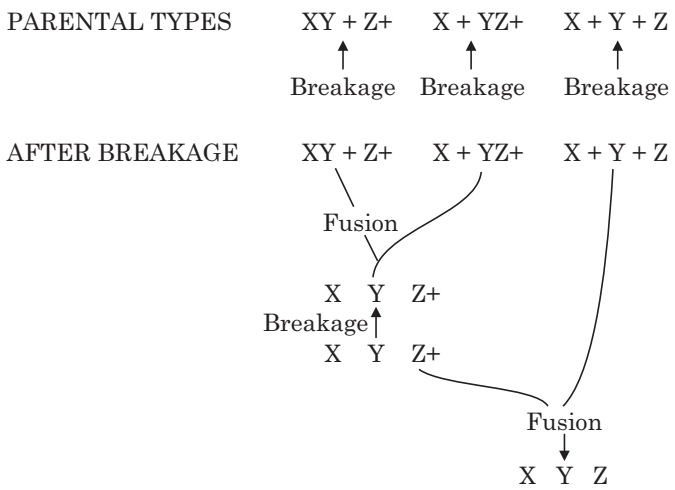


Fig. 18.6 Successive exchanges in phage DNA molecules after triple infection of a host cell.

It is due to multiple rounds of exchanges that recombinants recovered from any one infected cell do not occur with equal frequency. The frequency of recombination between two widely spaced genes is rarely more than 30 per cent can be explained as due to multiple exchanges as follows. Consider a cross between *xy* and *x* + *y* + where the two genotypes are present in equal frequency. There is an equal probability of *xy* exchanging with *xy i.e.* $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. The recombinants (*x* + *y* and *xy* +) can arise if both *xy* and *x* + *y* + have exchanges with each other, that is a probability of $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. If each such heterozygous exchange were to produce a recombinant, the maximum of 50 per cent recombinations would be obtained. However, when *x* and *y* are far apart they also tend to assort independently to produce *xy*, *x* + *y* +, *x* + *y* and *xy*+ in the ratio 1 : 1 : 1 : 1. Thus, only half of all the heterozygous matings produce recombinants, *i.e.*, half of 50 per cent = 25 per cent. In this way for genes that are widely spaced recombination frequency is always less than 50 per cent.

Recombination frequency in phages is influenced by factors like number of DNA molecules present in an infected cell, replication rates of different genotypes, and by proportions of the parental genotypes. By performing three-point crosses in phages, Streisinger and Berne (1960) demonstrated that *T*-phages had one linkage group of genes. This was later confirmed along with the circular nature of the chromosome.

Negative Interference in Phage Crosses

In eukaryotes a crossover in one region suppresses chiasma formation in an adjacent region. Double crossovers therefore, are fewer than expected and the phenomenon is known as *positive interference*. It is measured in terms of the ratio of double crossovers observed to those expected *i.e*., coincidence. When coincidence is less than 1.0, positive interference has occurred. In phage *negative interference* occurs so that double crossovers are observed *more* often than expected and coincidence is greater than 1.0. This is partly due to the successive rounds of exchanges between phage DNA molecules in the host cell.

Negative interference is of two types. One called *high negative interference* is localised within those regions of the chromosomes which have a high concentration of enzymes involved in recombination. This type of interference is observed in eukaryotes also within genes separated by very small gap distances. The *low negative interference* is found only in viruses, it occurs between genes widely separated from each other, and is probably due to multiple rounds of exchanges in DNA molecules.

Heterozygosity in Phages

Phages are haploid organisms with each gene represented once in the single DNA molecule. The experiments of Hershey and Chase in 1951 indicated that the phage DNA could be partially 'diploid' for some genes. When *E. coli* is doubly infected with *r*+ and *r* phages and the infected bacteria are plated on agar *before* bursting, some of the plaques formed are "mottled" in appearance (Fig. 18.7). Such plaques contain a mixture of *r*+ and *r* phages. That a *single* phage has given rise to both *r* and *r*+ progeny can be explained by assuming the formation of a *heteroduplex,* a hybrid DNA molecule with one strand containing *r* gene, the other *r*+ gene. Such a heterozygous condition leads to the occurrence of different base sequences at a specific site *(r* locus) in the two DNA strands. Such mismatched sites in DNA are acted upon by certain correction systems which restore homozygosity by removing bases in one strand and substituting bases complementary to those in the other strand.

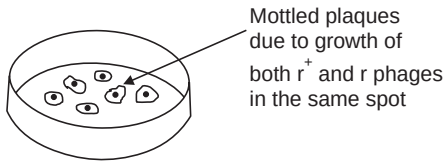


Fig. 18.7 The appearance of mottled plaques.

MAPPING GENES IN PHAGE LAMBDA

Various methods of gene mapping have been applied to the lambda chromosome. In one method, the sites of mutations are physically located by breaking up the chromosome into fragments of specific lengths. Isolated DNA helices from lambda are stirred in the blender at a certain speed

which breaks the helix into two half lengths. Increasing the speed of the blender produces quarter fragments, and so on. Fragments of different lengths are separated from each other by density gradient centrifugation in caesium chloride (CsCl). The various fragments of lambda have different AT : GC ratios. Therefore fragments of higher GC content which are denser, separate from the lighter AT rich fragments. Moreover, one half of the lambda genome is much richer in GC than the other half. The purified halves are used in DNA transformation experiments to find out which genes are located in each fragment.

A physical method of mapping the phage genome is as follows. The DNA of lambda is denatured by heating to 100°C, the high temperature breaking all the H-bonds to yield single strands (denaturation). If the separated strands are *gradually* cooled, the H-bonds reform to yield double helical molecules of DNA (renaturation). If DNA from two different viral genomes is denatured, then allowed to reanneal, heteroduplex molecules are formed of which each strand comes from a different viral genome.

The renatured DNA is observed in the electron microscope. Wherever a deletion or a mutated region is present in a strand, the corresponding segment of the other strand will not be able to pair with it, but will extend outwards in the form of a loop. This is a convenient method of detecting physical locations of deletions, mutations and mismatched bases. The data can be compared with the genetic map prepared from recombination frequencies.

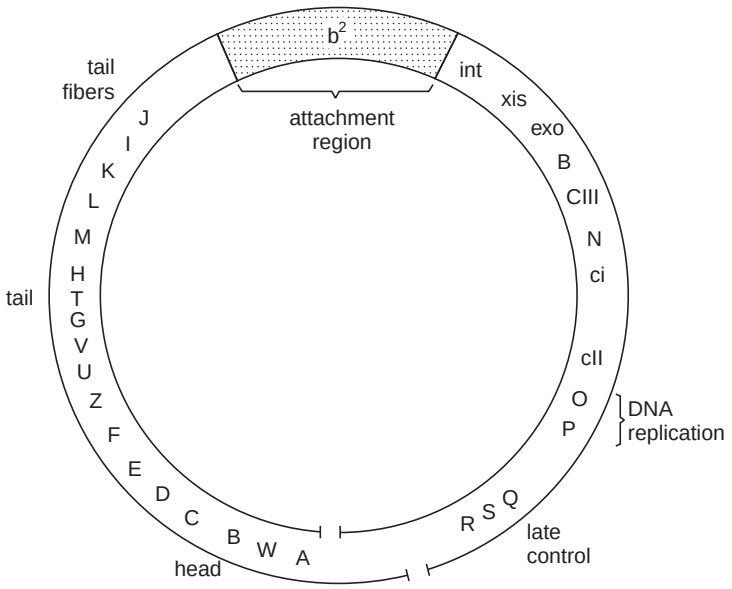


Fig. 18.8 The genetic map of phage lambda.

The genetic method of constructing a gene map of lambda consists in performing two-point and three-point crosses. The technique based on linkage studies in higher organisms (Chapter 8), was applied by Kaiser (1955) for mapping genes in lambda.

The genome of lambda has been worked out to be a linear DNA molecule about 3.8×10^6 daltons, corresponding to about 465 kilobase (one kilobase equals a length of 1000 bases). There are about 50 genes arranged in functional blocks (Fig. 18.8). The proteins involved in assembly of the phage particles are coded by the left half of the DNA molecule. The head protein genes (*A–F*) are all located to the left of the tail protein genes *(Z–J)*. The DNA segment in the *b2* region

between the *J* gene and the attachment site *(att)* does not perform any function necessary to phage growth. The point at which the phage integrates into the host chromosome (lysogeny) is denoted by *att*. The region from the attachment site onwards to *Y* controls genes which control the various phage mediated recombination events, that is, site-specific recombination (integration) and general recombination. The region from *Y* to *c*II is the control region that regulates genetic expression of the various operons in the genome. The *OP* region controls replication of phage DNA, and *Q* region controls the functioning of the late genes of lambda. The genes *S* and *R* at the right end of the genome code for proteins needed to lyse the cell.

The sequence of genes on lambda is similar to that in other lambdoid phages. In fact many regions in their genomes are homologous when heteroduplex molecules are observed.

THE RNA PHAGES

Some bacteriophages contain a single RNA molecule as the genetic material. They are classified into groups: phages *f*2, R17, MS2 and a few more are placed together in one group, *Q*β belongs to another group. RNA phages have an exceedingly small amount of genetic material, there being 3,600 nucleotides in *f*2 and over 4,000 in *Q*β. The single-stranded RNA is extensively folded into hairpin loops, giving rise to a characteristic flower arrangement (Fig. 18.9). Besides the hairpin loops, the secondary structure of RNA phages is also due to some base pairing between complementary base sequences located apart from each other.

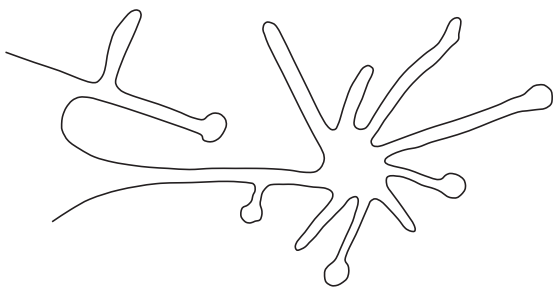


Fig. 18.9 The flower arrangement of a segment of an RNA phage genome.

The entire genome of *Q*β has only 3 genes controlling the synthesis of the following proteins: "maturation" or A protein, coat protein, and one subunit (*ii*) of the enzyme *Q*β replicase (Fig. 18.10). A fourth protein is also synthesised, starting from the coat protein gene by a readthrough of the codons beyond the termination site.

Infection involves attachment of specific RNA phages to the sex pilus and injection of RNA into the host cell. The infecting RNA (+strand) serves as a template for synthesis of the complementary minus (–) strand by the enzyme *Q*β-*replicase*. A base sequence near the 3′ end of the RNA molecule, called *Q*β *replicase recognition gene* regulates the start of replication by *Q*β replicase. The – strand is synthesised in the 5′ to 3′ direction and serves as the template for the synthesis of the + strand. The – strand requires the replicase and a protein factor for its synthesis, the plus strand only the replicase. The minus strand is used repeatedly for synthesis of + strand, so that the number of + strands far exceeds that of the – strands.

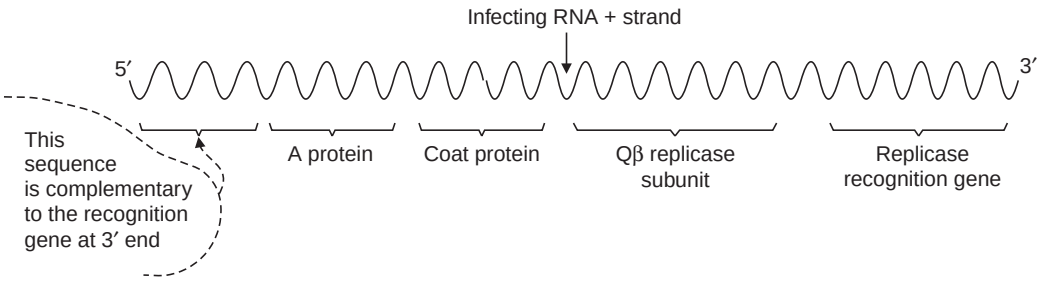


Fig. 8.10 Genome of the RNA phage Qβ.

The enzyme *Q*β replicase consists of four subunits (Fig. 18.11). Only one polypeptide chain (subunit II) is coded for by the viral genome. The remaining three subunits represent host cell proteins.

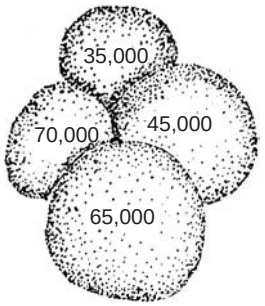


Fig. 8.11 The four subunits of the enzyme Qβ replicase.

TUMOUR VIRUSES AND CANCER

Besides mutations, some RNA and DNA viruses are also considered to be causal agents of cancer. They possess the ability to transform normal cells into cancerous ones. A cancerous cell is one which has uncontrolled growth and division and some altered surface characteristics. Among DNA tumour viruses are SV40 (a simian virus), mouse polyoma, adenoviruses which produce symptoms like common cold and tumours, and herpes group of viruses which produce mononucleosis, Burkitt's lymphoma and some other cancers. The RNA viruses include Rous Sarcoma virus (RSV) which causes solid tumours in chickens, and some viruses causing leukemias and sarcomas. The ability of cancer cells to multiply indefinitely is transmitted from parent to progeny cells. Tumour cells do not show virus particles in the EM. That is because in most cases part or all of the genetic material of a virus becomes inserted into the host chromosome (called *provirus).*

The DNA Tumour Viruses

SV40 is a spherical virus about 450 Å in diameter that infects cells of the monkey. When SV40 enters a monkey cell, it may be acted upon by the host cell's defense mechanisms and may disappear. Sometimes it multiplies to produce about a million progeny viruses present in the *nucleus* of the host cell. The cell nucleus no longer functions normally and the cell dies (*lytic* response). But when SV40 enters a cell other than that of monkey (mouse, hamster or human cell in culture), cell lysis does not occur. Instead, the cell becomes *transformed* into a cancer cell.

SV40 is one of the smallest DNA viruses (mol. wt. about 28 million). The spherical particle has a protein shell made up of 72 identical capsomeres (Fig. 18.12). Each capsomere is made up of 5 or 6 smaller protein molecules. There are three classes of polypeptide chains (VP-1, VP-2, VP-3) making up the shell protein. Within the shell is a single circular double helical DNA molecule. Four of the host cell histones H2A, H2B, H3 and H4 are bound to the SV40 DNA (minichromosome). Histone H1 is absent. Recent evidence indicates that the SV40 genome contains as few as 3 genes.

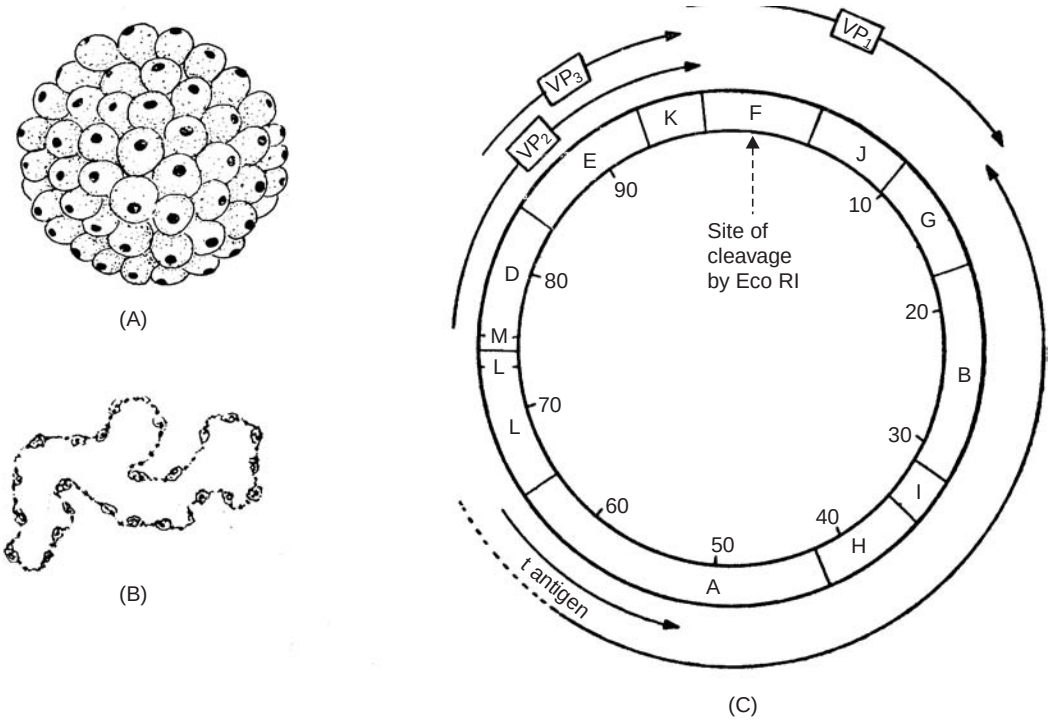


Fig. 18.12 A. Arrangement of capsomeres of SV40. B. Genome of SV40 as seen in the electron micrograph; nucleosome-like bodies are seen. C. Genetic map of the circular chromosome of SV40; arrows indicate direction of transcription and protein products of the gene.

The life cycle of SV40 is controlled by *early* and *late* genes. In the early stage there is synthesis of SV40 DNA and mRNA that codes for a protein called the T-antigen. The late genes code for the structural proteins VP1, VP2 and VP3 (see Chapter 22).

Studies with DNA-RNA hybridisation have shown the presence of SV40 DNA sequences integrated into DNA of transformed host cells. Each transformed cell contains about 1 to 20 SV40 genomes. When SV40 DNA is integrated into a human cell, it becomes inserted into chromosome No. 7.

RNA Tumour Viruses and Oncogenes

The best known RNA tumour virus is Rous sarcoma virus (RSV) and was first discovered by Peyton Rous in 1910 in chickens. It is spherical with a membrane-like lipoprotein coat surrounding an RNA core. The coat contains two different viral specific glycoproteins which project as knobs from the membrane surface.

The life cycle of RSV is similar to that of other RNA tumour viruses. After adsorption to the surface of the host cell, the intact virus particle comes into the cytoplasm. The membrane is shed

and the naked RNA is transcribed into a complementary DNA strand by the enzyme reverse transcriptase coded for by the viral genome. The discovery of reverse transcriptase enzyme by Temin, Mitzutani and Baltimore (rewarded with a Nobel Prize) has been crucial in understanding the genetics of RNA viruses, many of which belong to a family of viruses called retroviruses. The single-stranded DNA becomes converted into a circular double-stranded DNA molecule called the *provirus.* The provirus integrates into the host chromosomal DNA by a recombination process. The proviral DNA makes RNA transcripts from which virus-specific proteins are synthesised. The RNA cores become enclosed in the newly synthesised capsomere proteins and move to the cell surface. There they become enveloped by portions of the cell membrane which contain viral-specific membrane glycoproteins. The multiplication of RNA tumour virus does not interfere with normal cellular functions. Although thousands of progeny viruses are released from the cell surface, it does not lead to cell death as in the case of DNA tumour viruses. However, if the viral genome is carrying a special segment or gene known as *oncogene* (cancer gene), the activity of this gene can transform a normal cell into a cancerous one. Most tumour viruses can produce transformation also in cells growing in artificial culture media.

How exactly the RSV causes cancer is explained by the study of a very important group of mutants—the *deletion mutants* which are not able to transform cells into cancerous ones. They have a deletion of the small segment of the genome called *oncogene.* The deleted segment is the only feature distinguishing transforming and non-transforming viral strains. It is believed (this view has been strengthened by recent experiments described later) that the segment contains cancer-specific genes. Experiments conducted by Huebner and Todaro led them to suggest that oncogene-like DNA sequences are not present exclusively in viruses, they are not even peculiar to cancer cells, but appear to be present in the genomes of all cells. In normal cells oncogenes are considered to be repressed, never transcribed and harmless. In transformed cells they are stimulated to become active and convert cells into cancerous growth. Some viruses have a single oncogene others have a few of them. It is also likely that RNA tumour viruses become oncogenic (that is they develop the ability to induce tumours) after acquiring the oncogene from the host cell.

It must be noted however, that some tumour viruses are oncogenic only in animals which they do not infect in nature, while other tumour viruses are oncogenic in their natural host. It is fairly well established now that in oncogenesis the viral genome or portions of it become integrated into the host cell DNA. If the tumour virus does not carry an oncogene, in that case integration of the viral genome produces mutations in the host DNA. When mutations occur at certain sites they bring about cancerous changes in the cell.

The oncogene of RSV is called *src* (for sarcoma) and was first described by Martin in temperature sensitive conditional mutants of the virus. Weissmann and his colleagues at Zurich conducted a detailed study of *src* gene in deletion mutants of RSV mentioned earlier. They could identify the *src* gene as one segment of RNA near one end of the genome. By applying recombinant DNA techniques (see Chapter 23) RSV was shown to carry a single oncogene *src* coding for a single protein product and capable of transforming normal cells. The protein coded by *src* has been designated pp60v-*src* and has a molecular weight of about 60,000 daltons (Bishop, 1982). This protein is a kinase enzyme and acts by phosphorylating the amino acid tyrosine. In other words, the enzyme attaches phosphate ions to tyrosine in protein chains. Enzymes encoded by oncogenes have also been analysed in polyoma and a few other viruses, and phosphorylation of tyrosine seems to be a common function of oncogene-encoded enzymes. In the case of pp60v-*src* transformed cells have about 10 times as much phosphorylated tyrosine as compared to a normal cell. It has also been seen that pp60v-*src* phosphorylates tyrosine in a membrane protein called vinculin which in transformed cells decreases the ability of cells to adhere to solid surfaces. Decrease in cell adhesion is a common property of cancer cells.

If oncogenes are present within the genetic material of vertebrate animals and humans, it should be possible to locate them. Bishop *et al.,* (1982) prepared radioactive DNA probes complementary to the *src* gene in RSV and used them for molecular hybridisation with DNA of higher animals and humans. A sequence related to *src* was indeed found in humans and fishes and designated *c-src* (c for cellular). Interestingly *c-src* has been found to be transcribed and translated in normal cells. Its protein product is similar to pp60v-*src* in structure and function, that is, it phosphorylates tyrosine. So far 16 retrovirus oncogenes have been found to be present in normal vertebrate cells as cancer genes (Bishop, 1982).

It may be concluded that cancer genes are normal components of the genetic material of higher animals including humans. Very recent evidence for existence of cancer genes has come from the independent works of Weinberg and of Wigler who could transform normal human cells (NIH3T3 cells) in culture by means of DNA fragments cloned from human bladder carcinoma cell lines (Goldfarb *et al.*, 1982).

VIROIDS

Viroids are the smallest organisms causing infectious disease. All they have is a naked short strand of RNA. The protein coat present in viruses is lacking in viroids. So far, viroids have been identified only in diseases of higher plants, although their association with animal diseases is also suspected.

The best studied viroid is the one causing potato spindle tuber disease (PSTV). The infectious agent is a single stranded linear or circular RNA molecule about 50 nm long having molecular weight of 130,000; Diener (1960) called this agent viroid. The single stranded molecule is folded due to intrachain base pairing. The complete nucleotide sequence of PSTV was worked out by Gross *et al.,* (1978). The viroid has 359 nucleotides comprised of 73 adenines, 77 uracils, 10 guanines and 108 cytosines. A unique structural feature of PSTV is that it is a closed circular RNA molecule which seems to give the impression of a long double stranded molecule due to intrachain base pairing. The unpaired regions project outwards as loops. Viroids can replicate in the host cell, but do not seem to be translated.

In a few other plant diseases, viroids have been established as the causal organism. Among these are Cadang-cadang, a fatal disease that nearly wiped out coconut-trees in Philippines. The cucumber pale-fruit disease, the stunt disease of hops, and two diseases of chrysanthemums, namely, chlorotic mottle and stunt are all caused by viroids.

QUESTIONS

1. (*a*) State which nucleic acid and how many strands thereof are present in genomes of a *T*. phage, $\phi \times 174$, and a viroid.
   (*b*) Which component of the SV40 genome is not present in other viruses?
2. When Hershey and Rotman used two mutant phages for mixed infection of *E. coli,* which among the following results did they actually find:
   (*a*) No recombinants in the progeny.
   (*b*) Recombinants and parentals appeared in equal proportions.

(*c*) The two recombinant types did not occur in equal proportions.

(*d*) Parentals occurred more frequently than recombinants.

3. In higher organisms double crossovers occur less frequently than expected, in viruses more frequently than expected. Explain.
4. How does the DNA in a *T*4-phage head differ from a phage concatemer?
5. Comment on: (*a*) host range mutants; (*b*) minichromosome; (*c*) reverse transcription.
6. State three differences between the process of recombination as it occurs in phage and in higher organisms.

SELECTED READINGS

Akusjarvi, G. and Patterson, U. 1979. Sequence Analysis of Adenovirus DNA. IV. The Genomic Sequences Encoding the Common Tripartite Leader of Late Adenovirus mRNA. *J. Virol.* **134:** 143.

Alberts, B. and Frey, L. 1970. T4 Bacteriophage Gene 32: A Structural Protein in the Replication and Recombination of DNA. *Nature* **227:** 1313.

Berstein, H. and Berstein, C. 1973. Circular and Branched Concatenates as Possible Intermediates in Bacteriophage T4 DNA Replication. *J. Mol. Biol.* **77:** 355.

Bishop, J. M. 1981. Enemies Within: The Genesis of Retrovirus Oncogenes. *Cell* **23:** 647.

Bishop, J.M. 1982. Oncogenes. *Sc. Amer.* **246:** 68.

Delbrück, M. and Bailey, W. T. Jr. 1946. Induced Mutations in Bacterial Viruses. *Cold Spring Harb. Symp. Quant. Biol* **11:** 33.

Diener, T. O. 1981. The Viroids. *Sc. Amer.* **244:** 58.

Doermann, A. H. 1973. T4 and the Rolling Circle Model of Replication. *Annu. Rev. Genet.* **7:** 325.
Eigen, M. *et al.,* 1981. The Origin of Genetic Information. *Sc. Amer.* **244:** 78.

Epp, C. *et al.,* 1981. Downstream Regulation of *int* Gene Expression by the *b*2 Region in Phage λ. *Gene* **13:** 327.

Fiers, W. *et al,* 1978. Complete Nucleotide Sequence of SV40 DNA. *Nature* **273:** 113.

Goldfarb, M. *et al.,* 1982. Isolation and Preliminary Characterisation of a Human Transforming Gene from T24 Bladder Carcinoma Cells. *Nature* **296:** 404.

Gross, H.J. *et al.,* 1978. Nucleotide Sequence and Secondary Structure of Potato Spindle Tuber Viroid. *Nature* **273:** 203.

Hunter, T. 1980. Proteins Phosphorylated by the RSV Transforming Function. *Cell* **22:** 647.

Lewin, B. 1976. DNA Sequences Coding for More than one Protein. *Nature* **264:** 11.

Mansfield, *et al.,* 1980. Recognition Sequence of Bacteriophage ϕX174 Gene A Protein, An Initiator of DNA Replication. *Nature* **288:** 561.

Reddy, V.B. *et al.,* 1978. The Genome of Simian Virus 40. *Science* **200:** 494.

Streisinger, G. *et al.,* 1964. Chromosome Structure in Phage T4. I. Circularity of the Linkage Map. *Proc. Natl. Acad. Sci. U.S.* **51:** 775.

Wyke, J. 1981. Strategies of Viral Oncogenesis. *Nature* **290:** 629.

19

The Eukaryotic Chromosome

Genes are present in chromosomes. Questions regarding the chemical nature and functioning of genes made scientists probe the detailed structure of chromosomes. Conventionally, chromosomes are highly condensed, distinct rod-like structures visible in the light microscope at metaphase—anaphase stages of eukaryotic cell division. At interphase the material comprising chromosomes is in a different form, consisting of diffuse, unorganised threads with dark and light staining regions within the nuclear membrane. Cytologically chromosomes and chromatin are interchangeable forms of the genetic material at different stages of the cell cycle.

GROSS STRUCTURE OF CHROMOSOMES

The chromosome consists of two rod-like structures called *chromatids* joined with each other at the *primary constriction* or *centromere*. The portions of the chromatid on either side of the centromere are the arms of the chromosome. The ends of the chromosome are called telomeres. If the centromere lies midway along the length, the chromosome is said to be metacentric; if the centromere is slightly off-centre, then submetacentric; if the centromere is very near one end so that one arm is exceedingly short, the other equally long, the chromosome is acrocentric. There has been some dispute about the existence of telocentric (terminal centromere) chromosomes. Studies on the fine structure of the centromere indicate that telocentric chromosomes do exist.

Sometimes a secondary constriction is present (in addition to primary constriction) at specific places in specific chromosomes. In human beings chromosome numbers 1, 9 and 16 have it. The nucleolus organisers of many species are located at the secondary constriction. Some secondary constrictions have large amounts of satellite DNA.

In chromosomes fixed in OsO_4 the centromeric region of each chromatid appears to have a small intensely staining granule, called the spindle spherule. The spindle fibers become attached to this spherule rather than to the whole region of the primary constriction. The term *kinetochore* is applied to this granule. The terms centromere and kinetochore are often used synonymously which is incorrect. The term centromere includes the entire area of the primary constriction at which the chromatids are held together; it is visible at metaphase stage in the light microscope. The kinetochore is a specialised structure within the primary constriction to which the spindle fibers become attached; it is seen only in the electron microscope.

Whole mount preparations of chromosomes in EM show the centromere to be a region with lesser number of fibres, and is also the region where fibres of the two chromatids intermingle.

Kinetochore: Its structure is not the same in all organisms. There are two main types; (1) the trilaminar or stratified structure as in many animals and lower plants; (2) ball and cup structure as in higher plants. The trilaminar type consists of an outer dense layer 30–40 nm thick, a middle layer of low density 15–60 nm thick, and an inner dense layer, 15–40 nm thick, which is granular like the chromatin, and is dense and compact (Fig. 19.1). The middle layer is structureless, and has a clear area called the *corona*. The kinetochore usually takes after the shape of the centromere in which it lies. In some elongated centromeres the kinetochore may be 1.4 nm long and only 0.4 nm wide. Its size is increased by spindle poisons.

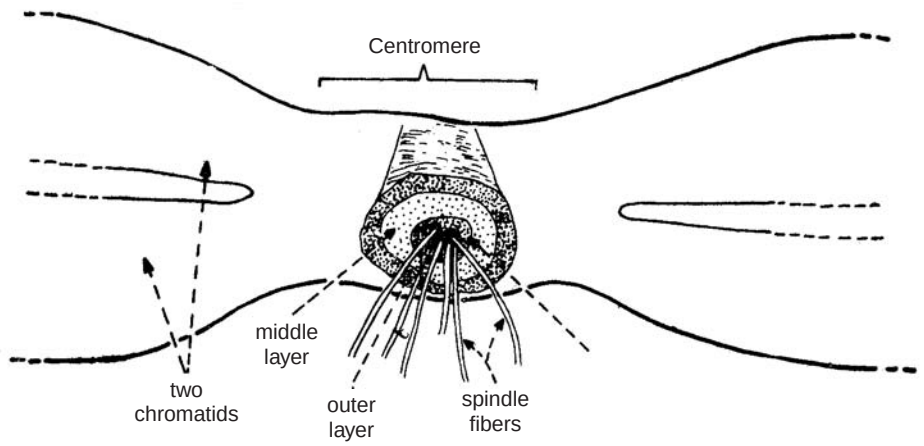


Fig. 19.1 Diagrammatic representation of the trilaminar structure in longitudinal section of the kinetochore.

The ball and cup type of kinetochore has a depression (the cup) about 1.5 μm across, in the surface of the chromosome on the side facing the spindle pole. In the middle of the depression is an amorphous mass, the ball, about 0.8 μm across. The ball is attached to the bottom of the cup, and the spindle microtubules appear to be attached all round the sides of the ball.

Diffuse centromere: Sometimes there is no localised centromere in a chromosome. Such *holocentric* chromosomes are present in certain plants like *Luzula*, some members of Cyperaceae, algae, protozoa and insects like *Steatococcus* and *Tamalia*. The kinetochore is said to be diffuse throughout the length of the chromosome; the spindle fibres also attach along the entire length. When such a chromosome is broken into small fragments by radiation, all the fragments move independently to the poles.

Heterochromatin and Euchromatin

Depending upon the degree of condensation, Heitz (1928, 1933) distinguished two types of chromosome material—*heterochromatin* which remains highly condensed throughout interphase, and *euchromatin* which unravels at the end of mitosis and stains weakly in interphase nucleus. Brown (1966) suggested two distinct types of heterochromatin, *facultative* and *constitutive*.

Facultative heterochromatin is present in one or the other of a pair of homologous chromosomes, not both. For example, the inactivation of one chromosome in mammalian females during early stages of development. In adults it is visible as the Barr Body in cells of the buccal

mucosa (Fig. 19.2). It is absent from male cells. Genetic experiments have indicated that although the DNA of facultative heterochromatin contains structural gene sequences, these are permanently repressed. The sequences on the homologous chromosome are euchromatic and expressed.

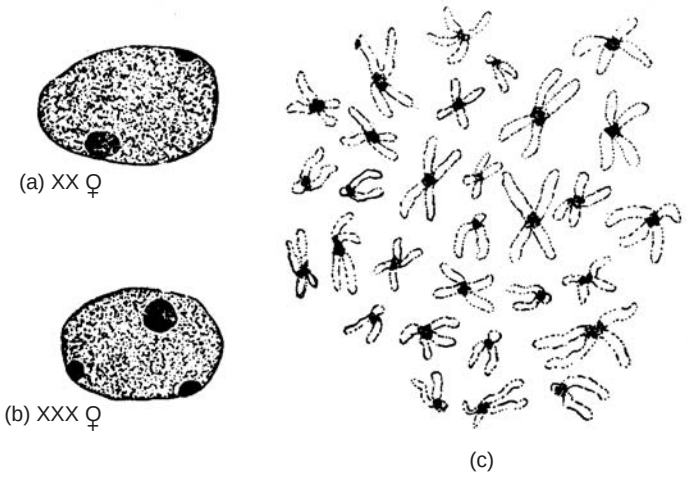


Fig. 19.2 (*a*) Single Barr body in normal female. (*b*) Two Barr bodies in an abnormal triple X female. (*c*) Centromeric heterochromatin in mitotic chromosomes.

The mealy bug (*Planococcus citri*) is an interesting example of facultative heterochromatin. In males the entire set of chromosomes inherited from the male parent becomes heterochromatic, genetically inactive, and eventually eliminated during spermatogenesis. These same chromosomes were euchromatic in the father. The set of chromosomes from the maternal parent remains euchromatic and contains active genes. The maternal set is passed on to the next generation. In males of the subsequent generation, this paternal set becomes inactivated (Nur, 1967), but in female offsprings it remains euchromatic.

Constitutive heterochromatin is permanently condensed and is found in the *same* locations in *both* homologous chromosomes. It is often present in specific regions of chromosomes such as the centromeres (Fig. 19.2*c*), telomere, nucleolus organising regions and other secondary constrictions. Biochemical experiments suggest that coding sequences in constitutive heterochromatin are inactive.

Centromere Function

The cellular DNA is replicated during S (synthetic) phase of interphase resulting in formation of two copies of DNA prior to mitosis. During cell division metaphase chromosomes reveal two identical sister chromatids held together at the centromere. Spindle fibres consisting of bundles of microtubules bind to the centromere, and pull the sister chromatids apart to the two poles. Centromeres consist of specific DNA sequences to which centromere-associated proteins bind. The DNA-protein complex is the **kinetochore**. These proteins act as molecular motors that drive the movement of chromatids to the two poles. Centromeric DNA sequences have been studied in yeast (*Saccharomyces cerevisiae*) by following the segregation of plasmids at mitosis. Plasmids that contain functional centromeres segregate like chromosomes and are equally distributed to the daughter cells formed after mitosis. In the absence of a functional centromere, the plasmid is not able to segregate properly, with the result that many daughter cells fail to receive plasmid DNA. These studies led to the determination of sequences required for centromere function. In

yeast, the centromere sequences are contained in approximately 125 base pairs consisting of three sequence elements, that is, two short sequences of 8 and 25 base pairs separated by 78 to 86 base pairs of very AT-rich DNA. The centromere sequences of another species of yeast, *S. pombe* are much larger, spanning 40 to 100 kb of DNA, making them about 1000 times larger than those of *S. cerevisiae*. Moreover, the centromere in *S. pombe* consists of a central core of 4 to 7 kb of single copy DNA flanked by repetitive sequences, making it more complex than that of *S. cerevisiae*.

The centromere in *Drosophila* spans 420 kb, out of which about 85% consists of two highly repeated satellite DNAs having sequences AATAT and AAGAG. The remainder of the centromere consists of interspersed transposable elements, that may also be present at other sites in the genome, and a nonrepetitive region of AT-rich DNA. Deletion of the satellite sequences, transposable elements and the nonrepetitive DNA reduced functional activity of the centromere. This implies that both repetitive and nonrepetitive sequences are required for kinetochore formation and centromere function.

Centromeres in humans and mammals have been studied through their centromere-associated proteins. Mammalian centromeres include extensive heterochromatin regions that contain highly repeated satellite DNA sequences. The precise function of centromere components in mammalian cells is not known.

Telomeres

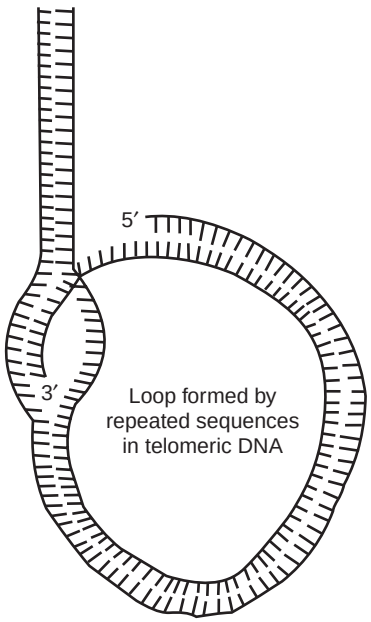
The **telomere** is a specialised structure at the extremity of a linear chromosome that is essential for the maintenance of chromosome stability. Classical studies recognised telomeres as structures whose absence produced sticky ends and unstable chromosomes. **Telomeres** contain *sequences* that play important roles in chromosome replication. Experiments in which telomeres from the protozoan *Tetrahymena* were added to the ends of linear molecules of yeast plasmid DNA allowed plasmids to replicate as linear chromosome-like molecules in yeast. It became clear that telomeres are required for replication of linear DNA molecules.

DNA sequences in telomeres are similar across a wide range of lower and higher eukaryotes, the same type of sequence is present in plants and humans. Each telomere consists of tandem arrays (arranged end-to-end or head to tail) of highly repeated sequences of DNA containing clusters of G-C residues on one strand. Thus, the sequence of telomere repeats in humans and various mammals is 5′-TTAGGG-3′ and in *Tetrahymena* it is 5′-GGGGTT-3′. These sequences span up to several kilobases, being repeated hundreds or thousands of times. One unusual feature of the telomere sequence is extension of the G-C rich strand, by about 14 to 16 bases, as an overhanging tail of single-stranded DNA. DNA replication cannot begin precisely at the 3′ end of a template strand, therefore, the 3′ end of the replicated duplex DNA must terminate in a short stretch in which the DNA is single-stranded.

The single-stranded overhanging DNA is subject to degradation by the action of nucleases during replication. If there was no mechanism to restore the end digested by nucleases, the DNA molecule in a chromosome would become slightly shorter with each round of replication. Such a mechanism exists, and studies of mutant cells in which this mechanism is defective indicate that each chromosome end does become shorter in each replication due to degradation.

The mechanism for restoring ends of a DNA molecule in a chromosome involves the enzyme **telomerase**. The enzyme first discovered in the ciliated protozoan *Tetrahymena*, functions by adding tandem repeats of a simple sequence—TTGGGG-3′ to the 3′ end of a RNA strand; in humans this enzyme adds the sequence –TTAGGG-3′. The tandem repeats of these sequences constitute the telomere. As the repeating telomere sequence is being elongated, DNA replication takes place resulting in synthesis of a partner strand.

A few copies of the telomere repeat seem to be necessary to prime the telomerase to add additional copies and form a telomere. The telomerase enzyme is remarkable because it incorporates an essential RNA molecule referred to as a **guide RNA**, that contains sequences complementary to the telomere repeat. The guide RNA serves as a template for telomere synthesis and elongation. The guide RNA undergoes base-pairing with the telomere repeat and serves as a template for telomere elongation by the addition of more repeat units. Notably, the complementary DNA strand of the telomere is synthesised by the *cellular DNA synthesising enzymes*. Recent studies provide evidence that the repeated sequences of telomeric DNA form loops at the ends of the chromosomes, perhaps to protect the chromosome terminus from degradation (Figure below).



The Telomere Loop

The telomeric DNA is anchored to the nuclear matrix by proteins called **Ku proteins**. Ku is an abundant nuclear protein. It exists as a heterodimer of two subunits of 70 kDa and 80 kDa. Ku functions in capping the telomeres, preventing chromosome end fusions, and in telomere length control. Ku binds DNA ends in a sequence-independent manner.

Structure of Human *Y* Chromosome

Out of the 46 human chromosomes, the sex chromosomes *X* and *Y* stand apart from the 44 autosomal chromosomes because of their distinctive role in mammalian sex determination. A human with *XY* in chromosome complement is male, while *XX* is female. Evidence suggests that *X* and *Y* originated a few hundred million years ago from a single ancestral **autosome**. The two then diverged in sequence in such a manner that today we find that only relatively short regions at either end of the *Y* chromosome are homologous to the corresponding region of the *X* chromosome. The remaining 95% of the present day *Y* chromosome is male specific referred to as **MSY** (**male-specific region** of *Y*). Recent studies of Skaletsky (2003) describe the MSY region as a mosaic of discrete sequence classes, namely heterochromatic (genetically inert) sequences as well as three classes of euchromatic (transcriptionally active) sequences designated **X-transposed**, **X-degenerated** and **ampliconic**. About 15% of MSY consists of *X*-transposed sequences, and as expected, they are even today 99% identical to their corresponding sequences in *X* chromosome.

These sequences consist predominantly of a high proportion of dispersed repetitive sequences and contain *only two* genes. About 20% of the MSY is comprised of *X*-degenerate sequences that are most distantly related to the *X* chromosome and have a *higher* gene content. The remaining portion of MSY consists of a web of *Y*-specific repetitive sequences called amplicons, that constitute a series of **palindromes**. These palindromes display a range of sizes, up to 3 Mb in length. The ampliconic DNA has the *highest gene content* and also a very high *pseudogene* content compared with the rest of the MSY.

Chromatin Organisation in Nucleus

Components within the nucleus are organised in relation to their functions. The nucleolus, chromatin, chromosomes, RNAs and nuclear proteins are localised to discrete sites. Consequently, their functions such as DNA replication, transcription of *r*RNA genes and assembly of ribosomal subunits in nucleolus, and processing of pre-*m*RNA become localised to distinct nuclear domains.

DNA in the nondividing nucleus is organised as chromatin fibres about 30 nm in thickness. The 30 nm fibre is organised into higher order structures. Folding of the chromatin fibres produces chromatin loops which have a DNA content of about 100 kb each. The loop-domain organisation is most clearly visualised in the highly transcribed chromosomes (**lampbrush chromosomes**) of amphibian oocytes. The fibres and loops are further organised into chromatin domains with DNA content of about 1 Mega base each. When a cell becomes committed to divide, the chromatin undergoes condensation and organises distinct linear chromosomes that can be visualised during the stages of cell division. That chromosomes are distributed in a nonrandom manner was first suggested by C. Rabl, way back in 1885, and proved correct a hundred years later in 1984 by studies on polytene chromosomes in salivary glands of *Drosophila*. Each chromosome occupies a discrete region of the nucleus with centromeres and telomeres clustered at opposite poles.

Further studies have shown that individual chromosomes occupy discrete territories within the nucleus. **Chromosome territories** may differ in position in different cell types at different times in development. Chromosome territories have been found to correlate with gene densities and gene expression. The territories of chromatin domains containing relatively few genes tend to be located near the periphery of the nucleus or near the nucleolus, whereas the territories of domains that are relatively gene rich tend to be located toward the interior of the nucleus. For example, human chromosome 18 that is about 85 Mega base in size, is relatively gene poor while chromosome 19, about 67 Mega base in size is relatively gene rich. In the nucleus, chromosome 18 territories tend to be at the nuclear periphery, whereas those of chromosome 19 tend to be in the interior.

Experiments reveal that the position of a gene in chromosomal DNA affects the level at which the gene is expressed. For example, in the case of genes introduced into transgenic mice, the transcriptional activity of these genes depends on their sites of integration in the mouse genome. This effect of chromosomal position on gene expression may be overcome by sequences referred to as **locus control regions**, which result in a high level of expression of the introduced genes, regardless of their site of integration. Locus control regions are found to stimulate only transfected genes that have been integrated into chromosomal DNA; they do not influence the expression of unintegrated plasmid DNAs. The mechanism of action of locus control regions suggests that they are not involved with individual promoters, but seem to activate large chromosomal domains, perhaps by inducing long-range alterations in chromatin structure.

The chromosomal domains appear to be separated from each other by boundary sequences called **insulator elements**, which prevent the chromatin structure of one domain from spreading to its neighbours. Insulators also prevent enhancers in one domain from acting on promoters located in an adjacent domain. Further, the spaces between chromatin domains are seen to form

a network of channels like the holes in a sponge. The channels are large enough to permit passage of molecules for the nuclear processes such as replication, transcription and RNA processing. Evidence suggests that these molecules move to their site of action in chromatin by passive diffusion.

DNA replication in mammalian cells appears to take place in discrete clustered sites resulting in multiple DNA molecules. Experiments by labelling cells with bromodeoxyuridine, an analogue of thymidine, have indicated that DNA replication appears to take place in large structures that contain multiple replication complexes organised into distinct functional domains. Transcriptionally active genes seem to be distributed throughout the nucleus. But components of the splicing process are located in discrete structural domains within the nucleus. Fluorescence microscopy of nuclei stained with immunofluorescent techniques using antibodies against small nuclear RNPs and splicing factors have demonstrated that components of the splicing apparatus are localised in 20 to 50 discrete structures referred to as nuclear speckles. Thus, splicing components seem to be recruited from their storage site, speckles, to the transcribed genes for pre-*m*RNA processing.

THE NUCLEOLUS

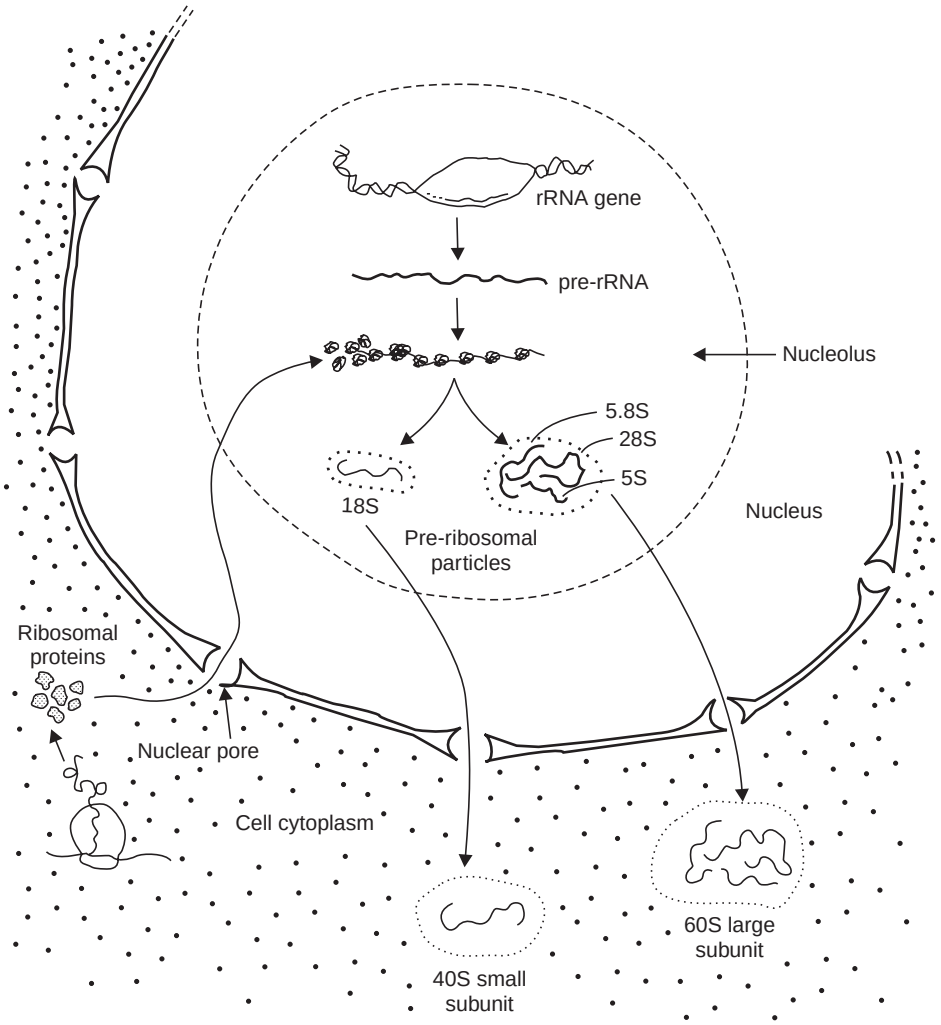
The nucleolus is the site for *r*RNA transcription and processing as well as for assembly of ribosomal subunits. The nucleolus is organised at the **nucleolus organising regions** (**NORs**) of some chromosomes. These chromosomal regions contain the genes for 5.8S, 18S and 28S *r*RNAs. There are four types of RNAs present in eukaryotic ribosomes, namely, 5S, 5.8S, 18S and 28S *r*RNAs. Among these, three, that is, the 5.8S, 18S and 28S are transcribed as a single unit within the nucleolus by RNA polymerase I producing a 45S ribosomal precursor RNA. The 45S pre-*r*RNA undergoes processing to yield 18S *r*RNA for the small ribosomal subunit as well as 5.8S and 28S *r*RNAs for the large ribosomal subunit. The 5S *r*RNA which is also present in the large subunit of ribosomes is transcribed outside the nucleolus by RNA polymerase III.

Ribosomes are central to protein synthesis and present in very large numbers, about 5 to 10 million in actively growing mammalian cells. The synthesis of ribosomes requires transcription of very large numbers of *r*RNA molecules. Cells meet these requirements by the presence of multiple copies of *r*RNA genes. The human genome contains about 200 copies of the gene that encodes the 5.8S, 18S and 28S *r*RNAs and about 2000 copies of the gene that encodes 5S *r*RNA. The genes for 5.8S, 18S and 28S *r*RNA are clustered in tandem arrays on 5 different human chromosomes, namely chromosomes 13, 14, 15, 21 and 22, while the 5S *r*RNA genes occur in a single tandem array on chromosome 1. Electron microscopically three regions can be distinguished in a nucleolus, the fibrillar centre, dense fibrillar component and the granular component. The rRNA genes are located in the fibrillar centre. Transcription occurs at the boundary of the fibrillar centre and the dense fibrillar component. Processing of the pre-*r*RNA initiates in the dense fibrillar component and proceeds into the granular component. In the granular component, *r*RNA is associated with ribosomal proteins to form pre-ribosomal particles that are exported into the cytoplasm.

Assembly of Ribosomes

The formation of ribosomes is completed by assembly of ribosomal precursor RNA with both ribosomal proteins as well as 5S *r*RNA. The genes that encode ribosomal proteins are transcribed outside the nucleolus by RNA polymerase II to produce mRNAs that are translated on cytoplasmic

ribosomes. Ribosomal proteins thus formed, are transported from the cytoplasm to the nucleolus where they associate with *r*RNAs to form preribosomal particles. Although the genes for 5S *r*RNA are transcribed outside the nucleolus by RNA polymerase III, the assembly of 5S *r*RNA into preribosomal particles takes place in the nucleolus. Ribosomal proteins associate with rRNAs while pre-*r*RNA is still being synthesised. The smaller ribosomal subunit which contains only 18S *r*RNA matures earlier than the large subunit which contains three types of *r*RNAs, 5S, 5.8S, and 28S *r*RNAs. The pre-ribosomal particles are then transported out into the cytoplasm where active 40S and 60S subunits are formed (Figure below).



Role of nucleolus in production of ribosomes

Sequence Complexity of Chromosomal DNA

Eukaryotic genomes contain large amount of repetitive sequences, sometimes present in hundreds or thousands of copies per genome. The understanding of repetitive sequences is based on studies conducted on **denaturation** (separation of DNA double helix into its two component strands) and **renaturation** (reassociation of the single strands into stable double-stranded DNA molecules) of DNA.

The two strands of a DNA molecule are held together by weak noncovalent bonds. When DNA is warmed in saline solution, a temperature is reached when two strands begin to separate, leading to single-stranded molecules in solution. This is called **thermal denaturation** or **DNA melting**. The progression of thermal denaturation can be followed by observing increase in absorbance of the dissolved DNA. The nitrogenous bases of DNA absorb ultraviolet radiation with an absorbance maximum near 260 nm. In single stranded DNA, the hydrophobic interactions caused by base stacking are increased which increases the ability of the bases to absorb ultraviolet radiation. The temperature at which the shift in absorbance is *half completed* is called the **melting temperature** (*Tm*) of DNA. The higher the GC content of the DNA, the higher the *Tm*. The reason being that there are 3 hydrogen bonds between G and C which confer stability on GC pairs, in comparison with AT pairs that are joined by two hydrogen bonds. Thus AT rich sections of DNA melt before the GC rich.

When denatured DNA is cooled *slowly*, the single strands reassociate to form double-stranded molecules, and properties of double helical DNA are restored, that is, it absorbs less ultraviolet light. This is called **renaturation** or **reannealing**. As described later, the property of reannealing has led to the development of methodology called nucleic acid hybridisation.

Britten and Kohne (1967) studied renaturation kinetics of DNA and discovered repeated sequences. Walker (1969) distinguished 3 kinetic classes of DNA: fast reannealing fraction or **highly repetitious DNA**, intermediate reannealing fraction or **moderately repetitious DNA**, and the slow annealing **unique** or **single copy** fraction.

1. **Highly repeated DNA sequences**: Also called reiterated or redundant DNA. Consists of sequences present in at least a million copies per genome, constitutes about 10% of the total DNA in vertebrates. Such sequences are usually short, about a few hundred nucleotides long, and present in clusters in which the given sequence is repeated over and over again without *interruption* in tandem arrays (end-to-end manner). Highly repeated sequences include the satellite DNAs, minisatellite DNAs and the microsatellite DNAs.

Satellite DNA: Consists of short sequences about 5 to 100 bp in length. During density gradient centrifugation, satellite DNA separates into a distinct band, because the base composition of satellite DNA is different from that of bulk DNA. A species may have more than one satellite sequence as in *Drosophila virilis* which has 3 satellite sequences, each 7 nucleotides long. Satellite DNA is present around centromeres in centromeric heterochromatin. In humans, 3 blocks of satellite DNA are present in the secondary constrictions of chromosomes 1, 9 and 16. A fourth block is present at the distal portion of the long arm of the *Y* chromosome.

Minisatellite DNA: These usually occur in clusters with about 3000 repeats, their size ranging from 12 to 100 bp in length. Minisatellite sequences occupy shorter stretches of the genome than the satellite sequences. Minisatellites are often unstable and the number of copies of minisatellites can increase or decrease from one generation to the next. The length of the minisatellite locus could vary within the same family, and in the population (*polymorphism*). Changes in minisatellite sequences can affect expression of nearby genes.

Microsatellite DNA: These include the shortest sequences one to five base pairs long, present in clusters of about 50 to 100 base pairs in length. They are dispersed evenly throughout

the DNA. The human genome contains about 30,000 different microsatellite loci. Changes in the number of copies of certain microsatellite sequences are responsible for some inherited diseases.

2. **Moderately Repeated DNA Sequences**: These are partially redundant. The sequences are highly similar but may not be identical. This fraction includes sequences that are repeated within the genome from a few times to tens of thousands of times. The genes for RNAs and histones are of this type. They constitute 15% of the DNA in mouse, 45% in *Xenopus*, and 80% in wheat, onion and salmon.

3. **Unique or Single-copy Sequences**: These sequences are present only once in the genome, or at the most, in few copies. They have a slow rate of reassociation. Most of the structural genes are found among the unique sequences. Mouse contains 70% and **Xenopus** about 55% of single copy sequences.

Dispersed Repeated Sequences

Unlike repeated DNA described above in which repeated sequences are clustered in a tandem manner, there are some repeat sequences that are scattered throughout the genome, referred to as **dispersed** or **interspersed DNA**, instead of being clustered as tandem repeats. Dispersed repeated sequences have been studied in many organisms. These are families of repeated sequences interspersed throughout the genome with unique sequence DNA. Often, small numbers of families have very high copy numbers and make up most of the dispersed repeated DNA in genome. In general, two interspersion patterns are encountered which allow these sequences to be classified as **SINEs** (short interspersed elements) or **LINEs** (long interspersed elements). Families of SINEs have sequences about 100 to 400 bp long, whereas LINEs have about 1000 to 7000 bp . All eukaryotic organisms have LINEs and SINEs, although their relative proportions vary widely. *Drosophila* and birds have mostly LINEs, humans and frogs have mostly SINEs. LINEs and SINEs represent a significant proportion of all the moderately repetitive DNA in the genome.

Mammalian diploid genomes have about 500,000 copies of the LINE-1 (L1) family of repeated sequences representing about 15% of the genome. Other LINE families are much less abundant than LINE-1. Full length LINE-1 family members are 6 to 7 kilo bases long. The full length LINE-1 elements are transposons, that is, they encode enzymes for movement of these elements in the genome (details in Chapter 22).

A good example of SINEs are the *Alu* sequences in mammalian genomes, so called because they contain a single site for the restriction endonuclease *Alu*I. Alu sequences are about 300 base pairs long, and about a million such sequences are dispersed throughout the genome, accounting for nearly 10% of the total cellular DNA. *Alu* sequences are transcribed into RNA, but they do not encode proteins, and their function is not known. Significantly, like the LINE-1 sequences, *Alu* sequences are also transposable elements, and capable of moving to different sites in genomic DNA if enzymes required for movement are supplied by active LINE elements.

In Situ Localisation of Satellite DNA

The precise locations of repeated DNA sequences on eukaryotic chromosomes have been determined by the technique of *in situ* hybridisation, first developed by Pardue and Gall (1970). The method is based on the fact that only those single strands of DNA/DNA or DNA/RNA hybridise which have complementary base sequences.

Cytological preparations of chromosome spreads are treated with NaOH which dissociates DNA. The preparations are incubated in a solution containing single-stranded nucleic acid molecules

(either DNA or transcribed RNA), which are labelled with tritium. The regions of the chromosomes that contain complementary base sequences hybridise with the corresponding sequences in single-stranded molecules. Their locations are determined by autoradiography.

Using labelled mouse satellite DNA, Pardue and Gall (1970) could determine the location of satellite sequences in the constitutive heterochromatin adjacent to the centromeres of mitotic chromosomes (Fig. 19.2*c*). Except for *Y*, all the remaining mouse chromosomes have satellite DNA at the centomeres. Later on many materials have shown satellite DNA in constitutive heterochromatin, that which forms C-bands with Giemsa (described later).

Sometimes there may be more than one type of satellite DNA in a genome. Human chromosomes have 4 satellite subfractions present in chromosomes 1, 9, 16 and *Y*. All the 4 satellite subfractions hybridise with chromosome 9. It is also interesting from the evolutionary standpoint that all the human satellite subfractions hybridise with monkey and chimpanzee DNA.

SOME SPECIFIC GENE SEQUENCES IN CHROMOSOMES

The way genes for ribosomal RNA, 5S RNA and histones are organised in chromosomes has been analysed. These are described below.

Ribosomal RNA Genes: A cluster of genes associated with the nucleolus represent ribosomal RNA genes. They have been studied most in the toad *Xenopus laevis.* The cluster consists of 450 tandemly repeated (arranged one behind the other) units. Each unit is made up of two regions, one which is transcribed to produce the *r*RNA molecule, and a spacer region which is not transcribed. The transcribed region is richer in *A* + *T* nucleotides. The spacer region is heterogenous in nucleotide composition and also varies in length within the same species.

Genes for 5S *RNA:* These have been sequenced completely in a number of organisms. They have been however, analysed best in two species of *Xenopus, X. laevis* and *X. mulleri.* The genes are repeated 10,000 to 25,000 times in *X. laevis* and are located in the telomere regions of 15 out of the 18 chromosomes. There is a transcribed region of 120 nucleotide pairs, and a nontranscribed spacer region about 600 nucleotide pairs long in *X. laevis.*

Genes for Histones: The development stages of sea urchin embryos have been extensively analysed for histones genes. This is a favourable material because it provides histone *m*RNAs in large quantities; the *m*RNAs are used as probes for locating their complementary sequences (histone genes) in chromosomal DNA.

The histone genes are repeated 300–1,000 times in *Xenopus,* but only 10–20 times in man. By use of restriction endonucleases, the genes for the 5 types of histones have been determined. The genes are clustered together and the repeat unit is about 6,000 to 7,000 nucleotides long (*i.e.,* 6–7 kilobase). Out of the 6–7 kilobase repeat unit only 2 kb are required to code for the five histones. It was on the basis of this finding that spacers were discovered in histone genes. The transcribed region is rich in *G* + *C* nucleotides and the spacer is *A* + *T* rich.

Inverted Sequences

A small proportion of DNA in higher organisms is present as *inverted repeated* sequences, also known as *palindromes*. If a nucleotide sequence is repeated and reversed on the same linear molecule, each of the repeated single strands will contain two sequences complementary to each

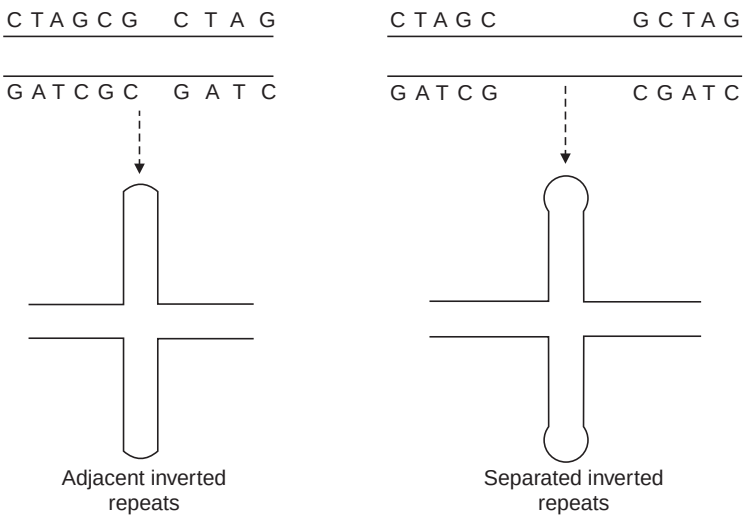


Fig. 19.3 Occurrence of inverted repeats in DNA and formation of hair-pin loops.

other. Each single strand folds back on itself to form a hairpin duplex structure. Such structures are visible in the electron microscope. They are also known as foldback or snapback DNA. The inverted repeats may be adjacent to each other, or separated by a number of nucleotides (Fig. 19.3). Inverted repeats are widely distributed throughout the genome. Their function is not known clearly, although some of them appear to be transcribed.

CHROMOSOME BANDING

Until about the 70s the only visible differentiation in mitotic chromosomes was the centromere and arm length. Around 1971 Caspersson and his colleagues in Sweden initiated staining of chromosomes with quinacrine mustard. A new substructure in the form of horizontal bands was revealed in the chromosomes. Later on a number of dyes were found to produce bands which were arranged in patterns characteristic for specific chromosomes (Fig. 19.4). A classification has been proposed for the various banding techniques at the Paris Conference (1971) as follows:

1. *Q-banding:* Chromosomes are stained with quinacrine mustard, quinacrine or Hoechst 33258, or some other dyes and observed in fluorescence microscope. The $A + T$ rich regions of chromosomes show intensely fluorescent bands, the $G + C$ rich regions do not (Fig. 19.4*B*).

2. *G-bands:* Chromosome spreads are first incubated in saline then stained with Giemsa (Schnedl, 1971; Drets and Shaw, 1971). The slides are examined in the light microscope. Treatment with urea and detergents has the same effect. G-bands (Fig. 19.4*A*) appear to be related in some way to differences in the state of protein sulfur along the chromosome.

3. *R-bands:* Chromosome preparations are incubated in a buffer at high temperature, (Dutrillaux and Lejeune, 1971) and stained with Giemsa. The banding pattern is reverse of the one observed with G-bands.

4. *C-bands:* Chromosome preparations are treated with a moderately strong alkali followed by warm saline and staining with Giemsa. The satellite DNA present around the centromeres becomes deeply stained to form C-bands (Arrhigi and Hsu, 1971).

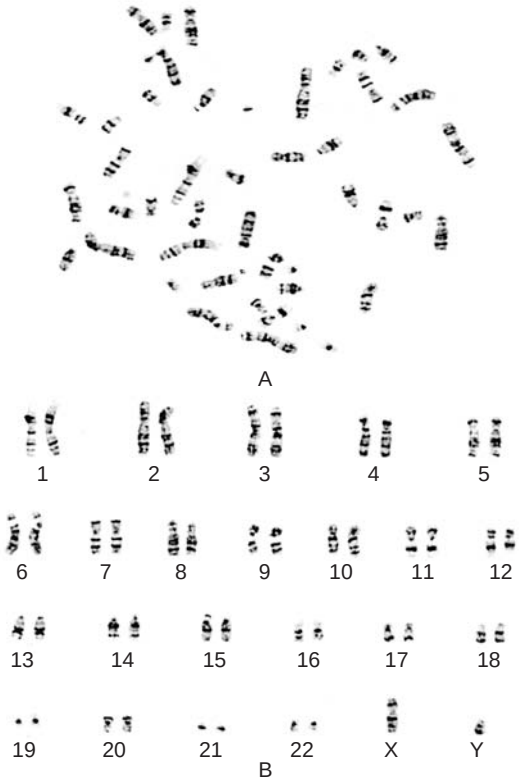


Fig. 19.4 (A) Metaphase chromosome spread stained for G-bands, (B) Karyotype prepared from the metaphase spread (courtesy Dr. Arundhati Sharma).

Importance of Banding

The chromosomes of some materials like mouse and ox are mostly acrocentric and form a continuous gradation in size. Banding pattern is useful for identifying individual autosomes with certainty. The bands can identify each chromosome of a normal complement. They can also identify small parts of chromosomes involved in structural rearrangements such as deletions and translocations. The banding technique has established the Philadelphia (Ph1) chromosome which is associated with chronic myeloid lukemia in humans. This chromosome was earlier believed to be a deleted chromosome 21, but banding pattern showed that it was in fact chromosome 22. Moreover, the positions of bands also indicated that in many cases the deleted portion of chromosome 22 was translocated to the end of the long arm of chromosome 9 (Rowley, 1973). The banding technique is also useful for assigning different linkage groups to specific chromosomes and for accurate gene mapping.

THE SYNAPTONEMAL COMPLEX

During their studies of paired meiotic chromosomes, Moses (1956) and Fawcett (1956) independently observed ribbon-like structures in the electron microscope. These were called *synaptonemal complex* (SC). The complexes were considered to represent the state of chromosome pairing.

The complex consists of two dense parallel lines called *lateral elements* (LE) and a less dense filament in the centre called the central element (CE). Across the space from the CE run *transverse filaments* towards each lateral element. This tripartite structure has chromatin material of paired homologous chromosomes lying on either side of the complex, next to each lateral element (Fig. 19.6). Thus each LE with its associated chromatin comprises one homologue. However, sister chromatids belonging to each homologue are not distinguishable in the complex. The dimensions of the complex are known; the LE is about 40 nm thick, and the CE 60–80 nm thick. Usually SCs are attached at both ends to the nuclear envelope. Digestion studies with trypsin and some other proteolytic enzymes have established that both LE and CE are made up of protein; they are resistant to DNAse digestion.

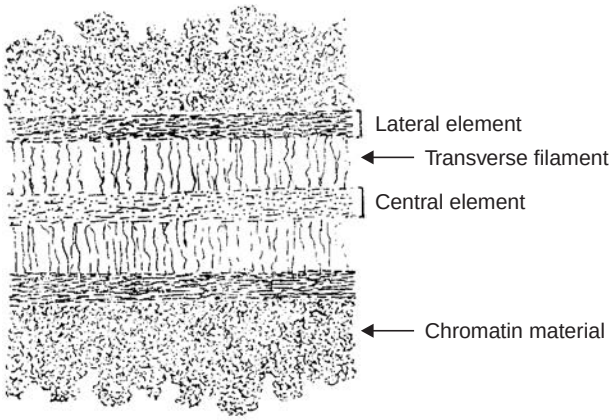


Fig. 19.6 Diagrammatic representation of a portion of a synaptonemal complex.

There may be one or more dense bodies in the central element called *nodes*, first observed in *Neurospora* and yeast, later in other fungi. Nodes are a common feature in SCs of *Drosophila* oocytes where they are about 100 nm in diameter (Carpenter, 1975). They are also called recombination nodules although their involvement in genetic exchange is not clear.

In spermatocytes of the Chinese hamster, Dresser and Moses (1980) found that at early pachytene the LEs are single elements, at mid pachytene they start appearing double; by late pachytene all the LEs are double along their lengths.

Besides the typical tripartite form, SCs with different configurations have been observed. In some insects the complex consists of multiple elements. In the triploid anthers of lily, each of the 3 homologues has its own dense axis at leptotene. Each axis then joins first with one and then with another of the remaining two homologues to form a double synaptonemal complex (Moens, 1969). The haploid spermatids of *Gryllus* contain multiple core complexes resembling stacked SCs somewhat separated from the chromosomes (Sotelo and Trujillo, 1960).

The synaptonemal complex first becomes visible at leptotene when unpaired chromosomes develop single dense axes, each of which will become a lateral element of a future SC. At zygotene the dense axes form parallel pairs similar to SC although CE is not yet clearly visible. As pairing advances, the complete tripartite structure of SC is seen. At pachytene full length SCs are visible, often twisted around their long axes. At diplotene SCs disintegrate and the paired homologues separate. However, the dense single axes remain visible resembling the leptotene configurations. The relationship between SCs and crossing over is not fully known.

STRANDEDNESS OF CHROMOSOMES

The eukaryotic cell contains an enormous amount of DNA. A human diploid cell has DNA which could be stretched to a length of 174 cm. This much DNA is distributed into 46 distinct chromosomes. The largest chromosome contains more than 7.3 cm, and the smallest 1.4 cm. But the highly condensed metaphase chromosomes are many thousand times shorter than the length of the DNA they contain. Obviously the DNA is greatly compressed and packaged inside the chromosome. This raises two questions: first, how many linear, duplex molecules of DNA are present in a chromosome? In other words, is there one DNA duplex in a chromatid (unineme or single strandedness), or more than one (multineme or multistrandedness)? The second question concerns the organisation of the DNA molecule within a chromosome and will be discussed under nucleosome.

The question regarding strandedness of chromosomes has been much debated. It is now generally accepted that the chromosome is *unineme,* containing a single DNA duplex per chromatid. The following evidences point to the unineme structure of chromosomes.

1. Taylor's experiments with autoradiography of root tip chromosomes in *Vicia faba* have shown segregation of newly synthesised DNA to only one chromatid in two generations. This finding is difficult to explain if two or more molecules of DNA are present in a chromatid.

2. Enzyme digestion studies with the giant lampbrush chromosomes in amphibian oocytes have shown that whole chromosomes are broken by the treatment with DNAse, but not by proteolytic enzymes or RNAse. This suggests the absence of nonDNA linkers made of protein or RNA. Gall (1963) studied kinetics of DNAase digestion of lampbrush chromosomes and concluded that each chromatid consists of a single DNA duplex. It was assumed that the number of breaks required at a particular site to produce a visible break could indicate the number of subunits present at the site of the break. It was possible to show that four hits or fractures are required to cause a break in a pair of sister chromatids. Since DNAase I produces single strand breaks, it suggests that a single chromatid consists of one DNA duplex.

3. Petes and Fangman (1972) applied the technique of sedimentation velocity analysis to chromosomal DNA in yeast *(Saccharomyces cerevisiae).* The haploid nucleus of yeast contains about $8.4 – 12.0 \times 10^9$ daltons (1 dalton = mass of 1 hydrogen atom = 3.32×10^{-21} gm). Yeast has 17 chromosomes, therefore the average DNA content per chromosome is from 4.9 to 7.1×10^8 daltons. Petes *et al* (1973) observed linear DNA molecules of yeast in the electron microscope and calculated that the above mentioned range in the amount of DNA is expected if there is a single DNA molecule in each chromosome.

4. Kavenoff and Zimm (1973) applied the viscoelastometric method which is suitable for measurement of long DNA molecules to yeast. In this method a cylinder is rotated inside another fixed cylinder which contains lysed cells. The long molecules which are stretched by rotation return to the relaxed state by forming coils; during this process they rotate the inner cylinder back to its initial position. The rate of the recoil movement is proportional to the molecular *weight* and the *number* of the largest DNA molecules in solution. Although the technique is not free from error, the data support the unineme model.

CHROMOSOMAL PROTEINS

The eukaryotic chromatin contains a wide variety of different proteins of which 2 broad categories can be recognized, the *histones* or basic proteins and the *nonhistones* which include all the

remaining proteins that are less basic (previously called acidic proteins). The two types of proteins can be distinguished by their isoelectric points. The isoelectric point of a protein is the pH at which the average charge of the whole molecule is zero; it depends upon the proportions of the basic (lysine, arginine, histidine) and acidic (asparagine, glutamine) aminoacids. Histones represent the largest proportion of chromosomal proteins and usually exceed DNA by weight. Histones have large amounts of lysine and arginine and isoelectric points all greater than 10. Proteins with isoelectric points below 10 are designated as nonhistones. The nonhistones are lesser in amount than the histones.

Although all histones are rich in basic amino acids, they differ in the relative proportions of lysine and arginine. They are classified into 5 different types designated H1, H2A, H2B, H3 and H4. The data on calf thymus histones has shown the following: H1 has lysine: arginine ratio of 22 and is said to be a lysine-rich histone. H2A and H2B have lysine: arginine ratios of 1.17 and 2.5 respectively; they are called moderately lysine-rich histones. H3 and H4 have lysine: arginine ratio of less than 1 and are called arginine-rich histones. The histones were earlier assigned the role of regulators of gene expression (Stedman and Stedman, 1950). But since histones do not have species or tissue specificity, the idea was ruled out (see Chapter 16) Much is now understood about the interaction of histones with DNA and the organization of the nucleosome.

The nonhistones are a heterogeneous group including proteins with structural, enzymatic and probably gene controlling functions. They have not been located in the nucleosome. They show both species and tissue specificity. The role of nonhistone proteins is not fully established.

THE NUCLEOSOME

When chromatin or whole chromosomes are spread on an air-water interface and examined in the electron microscope, fibers 250 Å in diameter are observed (Comings and Okada, 1971; some others). These are deoxynucleoprotein fibres in which DNA is complexed with protein. Removal of proteins by various methods reveals the 10 nm fibre (1nm = 10 Å), the ultimate subunit of chromatin. Pronase digestion of the 10 nm fibre leaves a DNAse sensitive 2 nm fibre which contains the single DNA double helix.

It is now well known that only a small fraction of eukaryotic chromatin is transcribed *(active* chromatin) in a particular nucleus while most of it is silent *(inactive* chromatin). In 1973 Miller and Beatty devised a technique for observing *active* transcribing chromatin in the electron microscope. The method described in Chapter 15 has been applied to many different organisms and two kinds of active regions have been recognised: the nuclear regions coding for rRNA and the nonribosomal cistrons. The active chromatin shows nascent RNA molecules on a DNA axis (lampbrush or christmas trees) separated by nontranscribed spacers (Chapter 15).

The inactive chromatin shows a beads-on-a-string appearance called the *nucleosome.* In 1974 when Olins and Olins applied Miller's technique to chicken erythrocytes, they found numerous fibres with regular knob-like structures released out of the nuclei. The name nu bodies was given to the beaded structures. Oudet, Gross-Bellard and Chambon (1975) observed similar beads in chromatin prepared from many vertebrate tissues and introduced the term *nucleosome.*

Nucleosomes are present in all the eukaryotic cells studied, in plants and animals and in mitotic and polytenic chromosomes. The bead and the connecting string (seen in Fig. 19.7) constitute one *repeat unit.* Structurally it contains about 200 base pairs (bp) of DNA and all the 5 histones. The bead contains about 140 bp of DNA wound in two complete superhelical turns around a

protein core which consists of 2 molecules of each of the histones H2A, H2B, H3 and H4. About 60 base pairs of DNA are present in the connecting string with which the histone H1 is associated. The term *core particle* has been used for the bead, and *linker* region for the thin connecting filament.

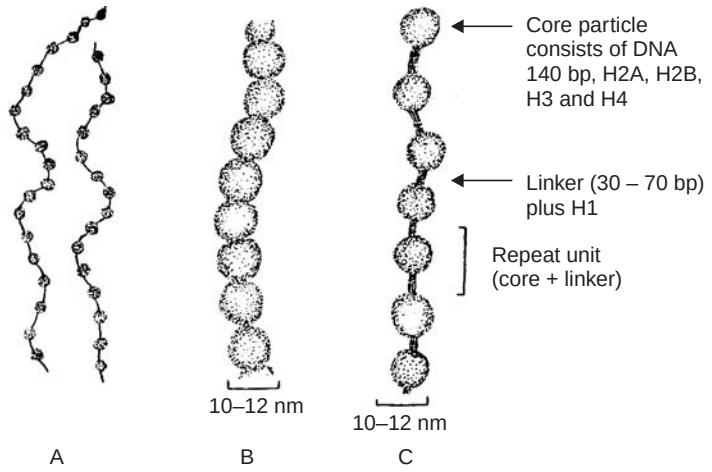


Fig. 19.7 Structure of the nucleosome. *A*. Whole mount of the 10 nm fibre, beads-on-a-string appearance. *B*. Compact nucleosomes; linkers are not visible. *C*. After removal of H1 histone, the linker and core particle become visible.

Thus the DNA in chromatin (Fig. 19.8*a*) is packaged into a series of *repeat units;* each repeat unit consists of two distinct parts, a *core particle* and a *linker region* (Fig. 19.8*b*), Chromatin from widely different species such as protozoa, lower and higher eukaryotes has shown that the core particle has a constant length of DNA of about 140 bp; but the linker length is variable within the same tissue and in different species (Tasanev, 1978). The nucleosome model for chromatin substructure is attractive because it explains the packaging of long strands of DNA into smaller repeat units (nucleosomes).

Biochemical analyses of chromatin have supported the nucleosome concept. When chromatin is digested with micrococcal nuclease, it produces oligomeric DNA fragments of specific lengths which correspond with the repeat unit observed in EM. The core particles are most resistant to degradation by nuclease and the linkers are the most nuclease sensitive regions.

X-ray diffraction and neutron scattering studies of crystallised DNA fragments have also established the existence of nucleosomes. A combined X-ray diffraction and EM study by Finch (1977) has shown that the nucleosome core particle is disc-like with dimensions of 110 Å × 110 Å × 57 Å. It consists of double helical DNA wound around a core of histone proteins with a superhelix pitch of about 28Å. It is calculated that each core particle is associated with 1.75 turns of DNA.

Although it is often stated that nucleosomes are present in inactive, non-transcribing chromatin, there are some conflicting reports on this aspect. EM observations have shown the absence of nuleosomes at the sites of rRNA transcription (Franke *et al.*, 1976; Woodcock *et al.*, 1976; Puvion-Dutilleul, 1977). However, nucleosomes have been observed in chromatin regions involved in nonribosomal RNA synthesis (Laird *et al.*, 1976; Mcknight *et al.*, 1977). Some studies on the structure of chromatin have shown the presence of nucleosomes in both active and inactive chromatin.

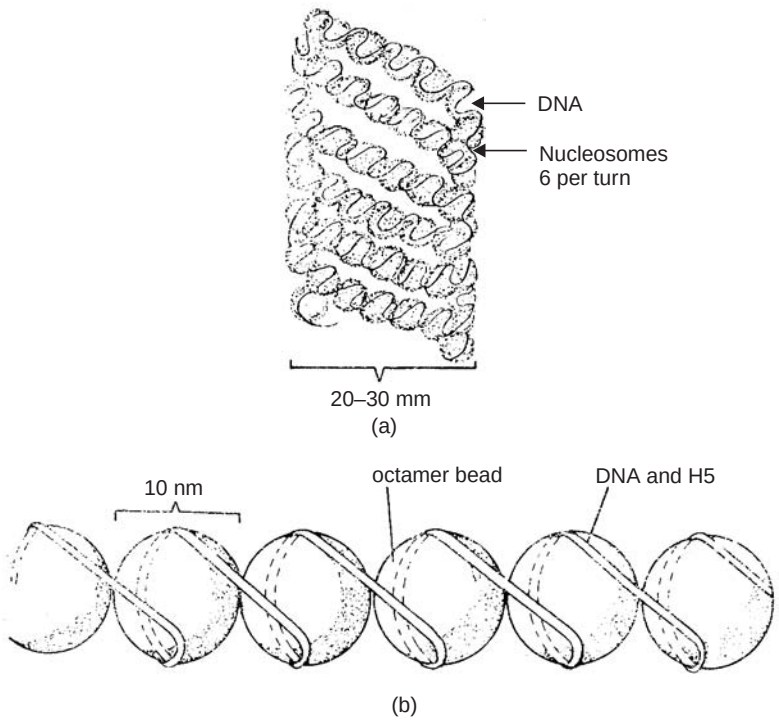


Fig. 19.8 (*a*) Supranucleosomal organisation of chromatin. (*b*) The nucleosome.

THE SUPRANUCLEOSOMAL STRUCTURES

The question now is how is the nucleosome packed into the higher order structures (250 nm fibres) of chromatin? A close packing into a supranucleosomal organisation has been suggested. Chromatin fibres isolated from some materials show globular units or knobs about 200 Å in diameter (Hozier *et al.*, 1977; Scheer *et al.*, 1978). There is evidence that the knob is a supranucleosomal bead containing on the average about 8 folded nucleosomes. Several investigators have suggested a helical arrangement of nucleosomes to form the higher order structure (Finch and Klug 1976; Carpenter, 1976; Alberts, 1977). Finch and Klug have found a close packing of nucleosomes to produce a nucleofilament, a fiber about 100 Å in diameter. The nucleofilament is further coiled up to form the "solenoid" with diameter of 30–35 nm (Fig. 19.8). There are about 6 nucleosomes per turn of the solenoid coils. The involvement of histone H1 in helical coils of the nucleosome has been suggested by some workers. According to Kornberg and Klug (1981), H1 leads to folding of the 10 nm fibre to form the solenoid. In Albert's (1977) model direct interactions among the nucleosomes generate crystal packing forces which give stability to the helix instead of histone H1.

The involvement of nonhistone proteins in nucleosome is not fully established (Colquhoun and Holmes, 1980). Comings (1979) has proposed a model in which the 250 Å fibre containing condensed nucleosomes is arranged into a series of loops or chromomeres by the association of DNA with nonhistone proteins.

THE SPECIALISED CHROMOSOMES

Two types of specialised chromosomes, the polytenic and lampbrush chromosomes exhibit some unique features and provide a system by which gene function mainly transcription can be visualised and studied.

Polytene Chromosomes

When repeated duplications of chromosomal material occur without anaphasic segregation, it results in an interphase nucleus having a number of parallel chromatids organised into a giant-sized chromosome. Called polytenic or salivary gland chromosomes, they were first described in salivary glands of *Chironomus tentans* by Balbiani (1881). Thereafter many species of Dipteran insects including *Drosophila* were found to have these chromosomes in the salivary glands and some other tissues.

Besides the enormous length which could reach up to 2,000 μm, a distinctive feature of the giant polytene chromosomes is the presence of horizontal bands ranging in number between 2,000 and 5,000 per chromosome. The crossbands stain deeply and alternate with lightly stained interbands. The thickness of individual bands in the larvae of various species of *Chironomus, Drosophila* and *Sciara* ranges between 0.05 μm and 0.5 μm; the narrowest bands being visible only under the electron microscope. The inter bands show longitudinal striations, or fibrils which are interpreted as indicating polyteny (meaning multistrandedness). It has been calculated that the thinnest fibre represents an individual chromatid. The number and arrangement of bands along the length of a chromosome is constant for the species. The banding pattern is also identical in all cells of the same tissue and have been used for chromosome mapping. Deletions and inversions in chromosomes are also readily identified from the position of the bands. Bridges (1935) was able to determine that the linear order of bands would correspond with the linear arrangement of genes on the chromosome.

Some of the bands are visible in an expanded form known as *puff.* When the compact DNA in a band becomes active, it unravels to form a swollen puff. When a puff becomes very much enlarged it is called a Balbiani ring. There is a characteristic pattern of puffing in different tissues and at different times during larval development. It was demonstrated that the presence of a specific puff is related with the appearance of a cellular phenotype, that is, a specific protein. For example, the salivary proteins were shown to be associated with a particular puff (Beerman, 1961; Grossbach, 1969).

That puffs represent sites of RNA synthesis (gene transcription) was shown by Pelling (1964) by the autoradiographic technique. Transcription occurs also in the bands but to a very small extent (Bonner and Pardue, 1976). The accumulation of ribonucleoprotein has been demonstrated cytochemically in the region of a puff. Inhibitors of transcription such as actinomycin-D and α-amanitin prevent puff formation and lead to some amount of regression of existing puffs. There is an increase in puffing activity during those stages of larval development at which the moulting hormone ecdysone is released from the prothoracic gland. This has also been shown experimentally by injection of ecdysone into fourth instar larvae which respond by increased formation of puffs.

Lampbrush Chromosomes

These are giant chromosomes present during meiotic prophase in oocytes of amphibians, birds and mammals including man. First described by Ruckert (1882), these chromosomes were so designated because of their resemblance to brushes used for cleaning the chimneys of victorian lamps.

At the diplotene stage of meiosis the oocytes of amphibians contain chromosomes which become modified into extremely long bivalents with lateral loops projecting from the chromomeres. This is considered to be an adaptation to the intense metabolic and synthetic activities of primary oocytes. This modified diplotene stage has been called the dictyate stage. In mammalian females, dictyotene is an arrested state of meiosis in which the oocytes remain for a very long time from birth until shortly before ovulation. It must be noted however, that the giant size of chromosome and clear lateral loops are visible only in amphibian oocytes. In most other species the chromosomes are not enlarged, and the loops are few and indistinct, in some there are hair-like lateral appendages or only fuzziness may be visible.

Lampbrush chromosomes have rarely been recorded at male meiotic prophase from spermatocytes of insects, birds, vertebrates (listed in Bostock and Sumner, 1978) and also in man (Walia and Minocha, 1980, Fig. 15.7). Among plants, they have been reported in tomato, wild onions, and the fungus *Neurospora*. In spermatocytes of two species of *Drosophila* large prominent lateral loops are present on the *Y* chromosome.

Due to their large size, lampbrush chromosomes can be dissected out of the oocytes and analysed chemically, enzymatically, or by autoradiographic techniques. Whole mounts of these chromosomes can be examined in an inverted phase contrast microscope. The lateral loops arise from the chromomeres, and consist of an axis of DNA from which numerous fine fibres project outwards (Fig. 15.6). That fibres consist of ribonucleoprotein (RNP; nascent RNA molecules coated with protein) was shown by Gall and Callan (1962). The loops are said to be asymmetrical as they have very long RNP fibres at one end, gradually shorter ones midway, and smallest fibres near the other end. There are usually no loops in the region of the centromere nor in the nucleolus organising region of specific chromosomes.

Enzyme digestion studies have shown that DNA is continuous throughout a lampbrush chromosome, being present in the interchromomeric regions, the chromomeres and the loop axis. Maximum growth of lampbrush chromosomes occurs at the dictyotene stage; as the cell enters metaphase the loops regress and finally disappear. Interestingly, each pair of loops has its own characteristic morphology. It has been possible to construct cytological maps on the basis of loop morphology. The specific features of a particular loop are transmitted as in Mendelian inheritance. An organism could be homozygous or heterozygous with respect to the morphology of a specific loop pair. The pattern of loops can be altered by mutations and can be observed in the heterozygotes. That loops are the sites of gene transcription has been shown by Gall and Callan (1962) and others. Soon after formation, the nascent RNA molecules become associated with newly synthesised proteins. When RNA synthesis is blocked by actinomycin-D, it leads to complete or partial collapse of loops.

Miller and Beatty (1969) pioneered a new spreading technique for lampbrush chromosomes. It is possible to visualize how transcriptionally active chromatin is organised in lampbrush chromosomes, at the level of the electron microscope (details in Chapter 15). The method has been extensively applied to transcribing chromatin in diverse materials such as embryos of rabbit (Cotton *et al,* 1980) and sea urchins (Busby and Bakken, 1979). HeLa cells (Miller and Bakken, 1972), hen erythrocytes (Scheer, 1978), rat liver (Puvion-Dutilleul, 1977), the green alga *Acetabularia* (Scheer *et al.,* 1976) and some others.

The spreading technique has revealed that the chromomeres are either compact or composed of numerous extended chromatin fibrils. The loop axes are densely covered with lateral RNP fibrils of increasing length, all these together constituting a transcription unit. Each lateral RNP fibril shows a basal granule identified to be an RNA polymerase molecule, at the point where the RNP fibril emerges from the DNA axis. In some cases the loop has a single transcription unit. In

others termination and reinitiation of transcription occur within a loop and produce several transcription units. The loops of lampbrush chromosomes in amphibian oocytes containing repeated genes for ribosomal RNA produce tandemly arranged transcription units separated by silent regions or spacers (christmas trees; Fig. 15.6, 8). One of the highly transcribed genes, the silk fibroin gene of *Bombyx mori* has been identified by the spreading technique (McKnight and Miller, 1976).

Molecular hybridisation techniques have been used to identify specific loops on lampbrush chromosomes. In this method DNA of known sequence labelled with radioactive isotopes is hybridised to nascent lampbrush loop RNA. Pukilla (1975) utilized purified DNA coding for 5S ribosomal RNA for hybridising with nascent RNA transcribed on the loops. The genes for 5S rRNA were thus located near the centromere of chromosomes 1, 2 and 6.

The study of lampbrush and polytenic chromosomes has shown that gene activity occurs on loops formed from *chromomeres* in the case of lampbrush *chromosomes* and in regions of *bands* in the case of polytenic chromosomes. Both chromomeres and bands consist of compact chromatin which appears deeply stained in the light microscope. Compact chromatin has been generally considered to be genetically inert. On the contrary however, neither the interchromomeric regions of lampbrush chromosomes nor interbands of polytene chromosomes have been found to show transcriptional activity.

QUESTIONS

1. State the points of difference between the following:
   (*a*) chromosomes and chromatid; (*b*) kinetochore and primary constriction; (*c*) G-bands and R-bands; (*d*) synaptonemal complex and pachytene chromosomes.
2. Give one example each for eukaryotic genes that are (*a*) permanently repressed, (*b*) present in clusters, (*c*) present in the nucleolus organising region of a chromosome.
3. Mention briefly an unusual feature of chromosome structure exhibited by each of the following:
   (*a*) the plant *Luzula;* (*b*) the bug *Planococcus citri;* (*c*) the mammalian female; (*d*) the crab; (*e*) spermatids of *Gryllus.*
4. Describe one method for studying redundant DNA sequences in each of the following situations:
   (*a*) for determining their presence/absence in a genome;
   (*b*) for a general idea about the relative abundance of AT and GC pairs;
   (*c*) for locating their sites on metaphase chromosomes.
5. Which protein is present in the eukaryotic chromosome that is absent in bacteria?

SELECTED READINGS

Abuelo, J .G. and Moore, D.E. 1969. The Human Chromosome. Electron Microscopic Observations on Chromatin Fibre Organisation. *J. Cell. Biol.* **41:** 73.

Adolph, K.W. *et al.,* 1977. Role of Nonhistone Proteins in Metaphase Chromosome Structure. *Cell* **12:** 805.

Angerer, R.C. *et a.l,* 1976. Single copy DNA and Structural Gene Sequence Relationships Among Four Sea Urchin Species. *Chromosoma* **56:** 213.

Arrighi, F. E. and Hsu, T.C. 1971. Localisation of Heterochromatin in Human Chromosomes. *Cytogenetics* **10:** 81.

Arrighi, F. E. *et al.,* 1970. Localisation of Repetitive DNA in the Chromosomes of *M. agrestis* by Means of *in situ* Hybridisation. *Chromosoma* **32:** 224.

Ashburner, M. 1970. Formation and Structure of Polytene Chromosomes during Insect Development. *Adv. Insect Physiol.* **7:** 1.

Aula, P. and Saksela, E. 1972. Comparison of Areas of Quinacrine Mustard Fluorescence and Modified Giemsa Staining in Human Metaphase Chromosomes. *Exp. Cell Res.* **71:** 161.

Bahr, G.F. and Golomb, H. M. 1971. Karyotyping of Single Human Chromosomes from Dry Mass Determined by Electron Microscopy. *Proc. Natl. Acad. Sci. U.S.* **68:** 726.

Bajer, A. and Mole Bajer, J. 1969. Formation of Spindle Fibres. Kinetochore Orientation and Behaviour of the Nuclear Envelope during Mitosis in Endosperm. *Chromosoma* **27:** 448.

Basu, S. 1979. Evidence for Superstructure of Wet Chromatin. *In* Chromatin Structure and Function. Part B. Ed. Nicoline. Plenum Press, New York.

Bauer, W.R. *et al.,* 1980. Supercoiled DNA. *Sc. Am.* **243:** 100.

Bonner, J. 1979. Properties and Composition of Isolated Chromatin. *In* Chromatin Structure and Function. Part A. Ed. Nicoline. Plenum Press, New York.

Callan, H. G. 1967. The Organization of Genetic Units in Chromosomes: *J. Cell Sci.* **2:** 1.

Carpenter, B.G. *et al.,* 1976. Organization of Subunits in Chromatin. *Nucleic Acids Res.* **3:** 1739.

Clark, R.J. and Felsenfeld, G. 1971. Structure of Chromatin. *Nature* **229:** 101.

Comings, D. E. 1978. Mechanisms of Chromosome Banding. *Annu. Rev., Genetics* 12: 25.

Crick, F.H.C. 1976. Linking Numbers and Nucleosomes. *Proc. Natl. Acad. Sci. U.S.* **73:** 2369.

Crick, F.H.C. and Klug, A. 1975. Kinky Helix. *Nature* **255:** 530.

Eichler, E.E. and Sankoff, D. 2003. Structural Dynamics of Eukaryotic Chromosome Evolution. *Science* **301:** 793–797.

Finch, J. T. and Klug, A. 1976. Solenoidal Model for Superstructure in Chromatin. *Proc. Natl. Acad. Sci. U.S.* **73:** 1897.

Gillies, C. B. 1975. Synaptonemal Complex and Chromosome Structure. *Annu. Rev. Genetics* **9 :** 91.

Greider, C.W. 1996. Telomere Length Regulation. *Annu. Rev. Biochem*. 6**5 :** 337–365.

Griffith, J.D. *et al*., 1999. Mammalian Telomeres End in a Large Duplex Loop. *Cell* **97 :** 503–514.

Grossbach, U. 1973. Chromosome Puffs and Gene Expression in Polytene Cells. *Cold Spring Harb. Symp. Quant. Biol.* **38:** 619.

Hozier, J.C. *et al.,* 1977. The Chromosome Fibre: Evidence for An Ordered Superstructure of Nucleosomes. *Curomosoma* **62:** 301.

Kano. Y. *et al.,* 1981. In Vivo Correlation between DNA Supercoiling and Transcription. *Gene* **13:** 173.

Klug, A. *et al.* 1980. A Low Resolution Structure for the Histone Core of the Nucleosome. *Nature* 287: 509.

Kornberg, R.D. and Klug, A. 1981. The Nucleosome. *Sc. Am.* **244:** 48.

Moens, P. 1978. Ultrastructural Studies of Chiasma Distribution. *Annu. Rev. Genet.* **12:** 433.

Maclean, N. and Hilder, V.A. 1977. Mechanisms of Chromatin Activation and Repression. *Int. Rev. Cytol.* **48:** 1.

Nagl, W. 1977. Nuclear Structures during Cell Cycles. *In* Mechanisms and Control of Cell Division. Eds. Rost and Gifford.

Olins, A.L. and Olins, D.E. 1974. Spheroid Chromatin Units (v bodies). *Science* **183:** 330.

Olins, A. L. *et al.,* 1976. Chromatin nu bodies: isolation, subfractionation and physical characterisation. *Nucl. Acids Res.* **3:** 3271.

Petes, T.D. and Fangman, W.L. 1972. Sedimentation Properties of Yeast Chromosomal DNA. *Proc. Natl. Acad. Sci. U.S.* **69:** 1188.

Portmann, R. 1976. Partial Denaturation Mapping of Cloned Histone DNA from the Sea Urchin *Psammtechinus miliaries. Nature* **264:** 31.

Rattner, J.B. and Hamkalo, B.A. 1978. Higher Order Structure in Metaphase Chromosomes. II.The Relationship between the 250 Å Fibre, Superbeads and Beads on a String. *Chromosoma* **69**: 373.

Ris, H. and Kornberg, Julie 1979. Chromosome Structure and Levels of Chromosome Organisation. In Cell Biology, A Comprehensive Treatise. Vol. 2. Eds. Prescott, D.M. and Goldstein, L. Academic Press, New York.

Schaffner, W. *et al.,* 1978. Genes and Spacers of Cloned Sea Urchin Histone DNA Analysed by sequencing. *Cell* **14:** 155.

Scheer *et al.,* 1979. DNA is assembled into globular supranucleosomal chromatin structures by nuclear contents of amphibian oocytes. *Exp. Cell. Res.* **129:** 115.

Shapiro, J.A. and Van Steinberg, R. 2005. Why Repetitive DNA is Essential to Genome Function. Biological Reviews of the Cambridge Philosophical Society. **80 :** 227–250.

Sharpless, N.E. and DePinho, R.A. 2004. Telomeres, Stem Cells Senescence and Cancer. *J. Clin. Invest*. **113 :** 160–168.

Skaletsky, H. *et al.,* 2003. The Male-specific Region of the Y Chromosome is a Mosaic of Discrete Sequence Classes. *Nature* **423:** 825–837.

Stansel, R.M. and Griffith, J.D. 2003. Telomeric Regulation in Eukaryotic Cells. *In*: Chromosomal Stability and Ageing. Eds. Hisama, F.M., Weismann, S.M., and Marlin, G.H. Marcel Dekker, New York. pp 73–106.

Wettstein, R. and Sotelo, J .R. 1971. The Molecular Architecture of Synaptonemal Complexes. *In* Adv. Cell. Mol. Biol. Ed. DuPraw. B.J. **1:** 109.

20

Mutations

The sudden heritable changes in genes, other than those due to Mendelian segregation and recombination constitute *mutations*. The idea of mutation first originated from observations of a dutch botanist Hugo de Vries (in the 1880's) on variations in plants of *Oenothera lamarckiana* (evening primrose) growing in Holland. This plant had been introduced from America and had grown wild in Europe. De Vries collected seeds from *Oenothera* plants, raised plants from them, and analysed the progeny for transmission of traits showing variation. He found that heritable variations were distinct from environmental variations. He gave the name mutation (latin *mutare* meaning change) to heritable changes and in 1901 published a book entitled "The Mutation Theory". Although De Vries is credited with the discovery of the idea of mutations, it was realised almost half a century later that perhaps none of the plants studied by him actually showed a gene mutation. Instead, the heritable phenotypic variations observed by him were found to be due to rare crossovers between translocated chromosomes.

The early concepts of mutation therefore arose out of genetic studies of visible phenotypes. Later, by the middle of the twentieth century, the molecular basis of heredity began to be investigated. It became established that the transmission of hereditary traits takes place due to an accurate process of self replication of the genetic material which is DNA. Today gross structural changes in the genetic material at the level of chromosomes are classified under chromosomal aberrations. They are treated separately. Only alterations in a very localised region of the chromosome at the molecular level are called mutations. They may involve one or more genes or nucleotides in DNA. Mutations therefore cause the substitution, deletion, addition or alteration of the sugar, base or phosphate of a nucleotide or of one more whole nucleotides. When mutations bring about a change in a single nucleotide they are called *point mutations.* When several nucleotides are altered, it results in *gross mutations*. The product of a mutation is called a *mutant*, and could be a genotype, cell, a polypeptide chain or an individual.

Mutations have been broadly categorised as *somatic* and *germinal* mutations. When a mutation occurs in a somatic cell, it does not change the whole organism, but produces a phenotypic change in the organ to which the mutant cell belongs. The resulting individual is a mosaic for mutant and normal tissues. Navel oranges (so named because when, first discovered in South America, the orange had a shrivelled, indented portion resembling the human navel), golden

delicious apples, emperor seedless grapes, some horticultural varieties of flowering plants, and white sectors in the red eyes of *Drosophila* males are examples of somatic mutations. In plants somatic mutations are transmitted to the progeny by methods of vegetative propagation such as budding and grafting.

Germinal mutations take place in cells of the germ line. A classic example of germinal mutations and perhaps also the one first recorded is that of short-legged sheep. In 1791 a farmer named Seth Wright in New England, U.S.A. noticed a short-legged lamb which could not jump over the fence and run away like all the other sheep in his small flock. Considering this to be an advantage, he started breeding work on the short-legged lamb. He had 15 ewes (females) and one ram (male) out of which he established a line of short-legged sheep. The name *ancon* was given to this breed (Greek meaning elbow), because the crooked looking forelegs showed resemblance to the human elbow.

The first short-legged lamb arose as the result of a recessive mutation in one of its recent ancestors. From the breeding work done in Wright's farm the crosses indicated that this lamb was homozygous for the mutant gene and several of the ewes were heterozygous carriers. The same mutation is said to have occurred in a flock of sheep in Norway in 1925 and another breed of ancon sheep has been produced out of it.

DETECTION OF SPONTANEOUS MUTATIONS

Normally genes are extremely stable structures, yet mutations have been occurring in nature all the time in a spontaneous manner. Although mutations are rare and occur at random, they have an advantage in providing a constant source of new variability which is required by organisms to adapt to the changing environment. There are numerous physical and chemical agents which can increase the frequency of mutations. If a mutation is dominant it is easily detected because it is always expressed. But dominant mutations are exceedingly rare. The mutant gene which causes aniridia (absence of iris in human eye) is dominant over the normal allele. Chondrodystrophic dwarfism is also due to a dominant gene. In this type of dwarfism, the head and trunk are of normal size, but arms and legs are extremely short. Mutations which cause lethal effects in adult individuals produce a clearly visible phenotype and are simple to detect. In *Drosophila* less than 20 % of all mutations are lethal.

Mutations which are most frequent are usually recessive. They are most difficult to detect if the changed locus lies on an autosome. Since a single allele of the recessive mutant gene will not express itself, such mutations perpetuate themselves for generations in heterozygous carriers without being noticed. They are detected when both gametes involved in fertilisation carry the mutation, thus bringing together two alleles of the recessive mutant gene.

Haploid organisms offer an advantage for the study of autosomal recessive mutations owing to the fact that a single allele expresses itself. If a recessive mutation is sex-linked it follows a specific pattern of transmission. Such a mutation is often detected when it expresses itself even in the single dose in hemizygous males.

Various tests have been devised for detecting spontaneous mutations in a population. Some are described below.

Replica Plating

In 1946 Luria and Delbrueck noticed one of the first cases of spontaneous mutation in a bacterial population. When *E. coli* cells were infected with the bacteriophage *T*1 most of the cells lysed releasing progenies of newly synthesised viruses. However, a very small proportion of bacterial cells (about 1 in 10^8) did not lyse but could multiply and form colonies. These bacteria as well as their descendants obtained by subculturing the colonies were found to be resistant to *T*1. It appeared that some of the originally infected cells had become mutant forms resistant to phage. The *replica plating* technique devised by Lederberg in 1952 demonstrated that *E. coli* cells had undergone spontaneous mutation and become resistant to *T*1.

The technique is as follows. *E. coli* cells are grown on a master plate of nutrient agar until they form a continuous lawn of cells. A sample of cells is transferred by pressing on it a sterile velvet surface supported on a circular block (replicator) to a fresh plate of agar medium already coated with T1 phage particles. The method produces an identical pattern of colonies on the second plate. Some mutant phage resistant colonies also grew on this plate.

The question arises about the origin of these mutant colonies. If the mutant colonies had originated from the master plate, then it would be possible to obtain a culture of such cells by inoculating from locations on the master plate corresponding to positions of resistant colonies on the second phage-containing plate. Lederberg therefore repeated the procedure by preparing a second master plate and *replica plating* on to a phage coated plate. This was done 3 to 4 times, each time plating cells thinly. In this way single resistant colonies could be identified (Fig. 20.1) and a true breeding phage resistant strain would be developed. The experiment proves that resistant cells arose by mutation on the first master plate and were not induced by the phage to become resistant.

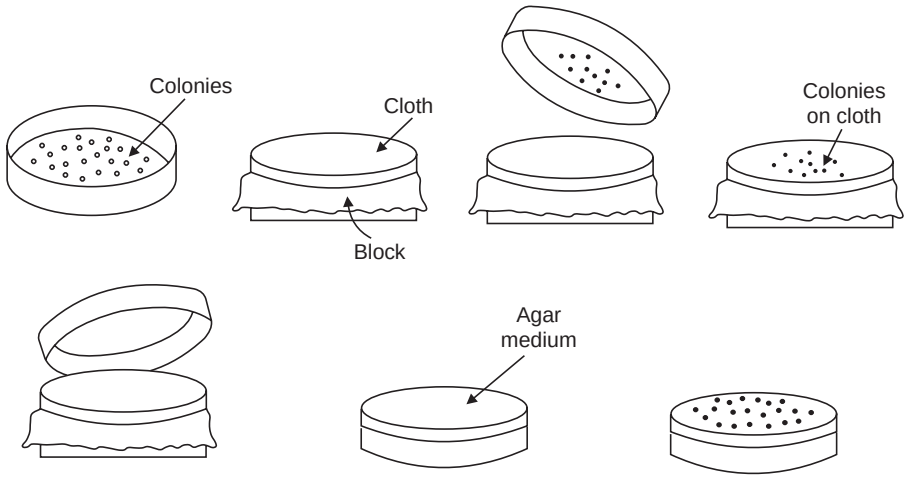


Fig. 20.1 The technique of replica plating.

Mutation Rates in Maize Endosperm and Drosophila

L.J. Stadler studied mutation frequency in maize population affecting endosperm characteristics. It was found that mutation frequency varied with eight different genes. The colour gene *R* mutated about 492 times in a million gametes; the gene for full endosperm *(sh^+)* mutated to *sh* for shrunken endosperm about 3 times in 2.5 million gametes; the gene for normal endosperm did not mutate to the gene for waxy conditions in 1.5 million gametes.

From studies of sex-linked lethals in *Drosophila*, Muller has estimated that one out of 20 flies at some stage of life undergoes spontaneous mutation of some gene. He has also devised a method of estimating mutation rate from matings of attached *X* females (described earlier).

When an attached *X* female is fertilised by a normal male, it produces 4 types of offsprings: *XXX* females and *YY* males-both inviable; the other two types are attached *X* females and *XY* males where the males inherit their *Y* chromosome from the mother and *X* from the father. In such F1 males all the sex-linked recessive traits carried by their father are expressed and easily detected.

Detection of Mutations In Micro-organisms

In 1946 Beadle and Tatum discovered a method of detecting nutritional mutants in the bread mold *Neurospora crassa.* Wild type *Neurospora* grows on a minimal medium containing inorganic acids, sugar, nitrogen, a few salts and a vitamin biotin. Beadle and Tatum found nutritional mutants of *Neurospora* which failed to grow on minimal medium in absence of a particular nutrient.

The exact requirements of nutritional mutants can be determined by adding a different growth substance to each of a series of different cultures of minimal media. For instance, a mutant strain of *Neurospora* deficient in the enzyme tryptophan synthetase would be able to grow on the minimal medium only if tryptophan is added to the medium. The technique has turned out to be applicable to many micro-organisms.

Mutants of the kind described above are known as *auxotrophs*, and their growth requirement may be for a vitamin, an amino acid, a nucleoside or a nucleic acid base. Auxotrophic mutants have lost the ability to synthesise for themselves the substance which they require. The substance has to be supplied to the medium.

In 1949 B. D. Davis found an elegant way of detecting auxotrophs in *E. coli* by a penicillin selection method. Certain antibiotics like penicillin inhibit cell division only in growing, actively dividing bacteria. Thus when penicillin is added to a minimal medium containing bacteria the wild type cells (also called prototrophs) which are growing are killed. The auxotrophs which are not growing are not acted upon by penicillin and survive. After a certain time period the auxotrophs are separated from the medium by centrifugation, washed free of penicillin and planted on fresh minimal medium supplemented by the necessary growth substance. The auxotroph will divide and form a more or less pure colony.

Auxotrophs themselves belong to a class of mutants broadly known as conditional mutants; their growth is dependent on the presence of a particular environmental condition. The *temperature-sensitive* mutants fail to grow normally in a certain temperature range. Usually higher temperatures affect growth because of defects in thermolability of specific proteins. Temperature sensitive mutants in bacteria can be detected by the technique of replica plating described earlier. The replica plate is incubated at or above the maximum temperature which permits growth of the wild type (for *E. coli* 42°C). If the replica colony fails to grow under these conditions it indicates that it is probably a temperature sensitive mutant. In yeast some mutants become manifest at elevated temperatures through deficiencies affecting growth.

CYTOLOGICALLY VISIBLE GENE MUTATIONS

In some higher plants gene mutations produce four types of cytologically visible meiotic abnormalities. These are:

(*a*) Failure to enter meiosis, for example the ameiotic mutant in *Zea mays*.

(*b*) Failure of homologous chromosomes to pair at zygotene and resulting failure of chiasma formation, as in the asynaptic mutant in pea (*Pisum sativum).*

(*c*) Mutants in which synapsis occurs at zygotene but chiasmata are few or none. Found in the desynaptic mutants of many plants including *Zea mays*.

(*d*) Mutants having precocious centromere division as in tomato.

(*e*) Mutants which produce chromosome condensation, preferential segregation, and those which affect the amount of recombination.

Since in all the above cases sterile gametes are produced (pollen and eggs), it is difficult to devise tests for detection of the number of loci involved in each case. However, some estimates made suggest that in pea four, and in tomato six loci determine asynapsis.

MUTABLE GENES

Although each gene is a potential site for a mutation, yet some genes mutate more frequently than other genes. Such *mutable* genes are widespread in plants and animals. Somatic mutations in plants giving rise to variegations in leaves and flowers are due to mutable genes. In *Drosophila*, mutations affecting eye and/body colour and size of wings are due to mutable genes.

In bacteria spontaneously arising gene mutations occur with a frequency of about once in 10^6 gene duplications (*i.e.,* cell cycle). Many times a change in a gene or nucleotide pair does not produce a detectable mutation. Fine structural analysis of the *rII* locus which shows a high rate of detectable mutations has shown that there are 500 mutable sites. Moreover, changes at a number of different mutable sites can lead to the same mutation. This is best demonstrated by the enzyme tryptophan synthetase in *E. coli*. The enzyme consists of two polypeptide chains A and B. The two chains can be separated and have shown that neither chain A nor B alone is active in synthesis of tryptophan. Mutants of *E. coli* which lack ability to synthesise tryptophan have shown that the A chain is not functional. Detailed genetic analysis of the A chain revealed that there are a large number of mutable sites in the A chain which are responsible for the inactivity of the enzyme.

Furthermore, a single mutable site can undergo different mutations. In the enzyme tryptophan synthetase, a change at the same mutable site results in two different kinds of polypeptide products, one in which glycine is replaced by glutamic acid, and a second in which glycine is replaced by valine.

INDUCTION OF MUTATIONS

In 1927 H. J. Muller showed for the first time that mutations could be induced in *Drosophila* by use of external agents or mutagens. He was awarded Nobel Prize in 1946. When flies were irradiated

with X-rays, he found that the offsprings showed new phenotypes which were similar to those produced by spontaneously occurring mutations. He also found that increasing the dose of X-rays results in a linear increase in the frequency of mutations. By 1930s, it became established that physical agents such as X-rays, gamma rays, UV radiation, and some chemical agents are all effective as mutagens. At the same time L. J. Stadler demonstrated that X-rays could produce gene mutations in plants of barley. Mutagenesis by radiation and chemicals are discussed separately below.

Theoretical Background of Radiation

An atom is composed of a positively charged nucleus and negatively charged electrons orbiting around the nucleus. The nucleus contains uncharged neutrons and positively charged protons. The electrons can move from one orbit to another. An electron absorbs energy when it jumps to a high energy level, and releases energy when it moves to a lower energy level. The released energy is in the form of electromagnetic radiation. Visible and UV light, X-rays and gamma rays are all electromagnetic waves.

Radiation breaks chromosomes by a chemical reaction which requires energy. The mutagenic effect of radiation depends upon the amount and type of energy left in the tissue. Visible light is a less energetic radiation as it leaves energy in the form of heat. But the energetic radiations such as ultraviolet (UV) leave energy in the form of heat and activation which leads to chemical change. Activation is the type of energy which makes an electron move from an inner to an outer orbit of the atom. X-rays and gamma rays are high energy radiations that not only heat and activate, but leave energy in the cell in the form of ionisation.

Ionising Radiation

Due to their shorter wavelength, X-rays and gamma rays penetrate tissues deeper than visible and UV light. They can impart enough localised energy to absorbing tissue to ionise atoms and molecules. When a highly energetic wave moving at high speed is stopped, it releases energy. This energy makes an atom lose an electron and become a charged particle or *ion.* The process is called ionisation. The free moving electron causes other atoms to lose electrons and become positively charged ions. The two processes generate pairs of positively and negatively charged ions. A number of ions may be clustered together to form an *ion track*. Ions undergo chemical reactions to neutralise their charge to reach a more stable configuration. While doing so they (ions) produce breaks in chromosomes (DNA) thereby inducing mutations. The free ions moreover, may combine with oxygen and produce highly reactive chemicals which may also react with DNA and cause mutagenesis.

Some ionising radiation is electromagnetic such as X-rays and gamma rays and some consists of subatomic particles such as electrons, protons, neutrons and alpha particles. X-rays and gamma rays have a low rate of linear energy transfer as they produce ions sparsely along the ion track and penetrate deeply into the tissue. Charged particles have a higher linear energy transfer, they do not penetrate deeply and produce more damage than X-rays and gamma rays.

Units of Measurement: Radiation is measured in terms of an ionisation unit called roentgen or *r* unit, one *r* being equal to 1.8×10^9 ion pairs per cubic cm of air. In tissue which is ten times as dense as air, a high energy radiation produces about 1000 times the number of ion pairs per cubic cm as it does in air. Another unit called rad measures the total amount of radiant energy absorbed by the medium. One rad equals 100 ergs per gram of tissue. Another unit called *gray* is equivalent to 100 *rads.* In the case of X-rays about 90 % of the energy left in the tissue is used to produce ions, the rest produces heat and excitation. Ultraviolet (UV) is a nonionising type of radiation and is measured in *rads* instead of *r* units. When ionisation is caused by subatomic

particles, the doses are measured in different units called *rem* and *sievert.* One *rem* is defined as the amount of any radiation that produces a biological effect equivalent to that resulting from one *rad* of gamma rays. A *sievert* is equal to 100 *rems*. For detecting radiation the Geiger-Muller tube is used. The tube contains a gas which is ionised by radiation. The amount of radiation is gauged from suitable amplifiers and counters.

Effects of Ionising Radiation

Zirkle in 1930 showed that in plants the nucleus is more sensitive to ionising radiation than the cytoplasm. It is now known with certainty that many molecules including DNA are affected by ionising radiation. The purines are less sensitive to radiation than pyrimidines. Out of the pyrimidines, thymine is most sensitive. Large doses of ionising radiation destroy thymine, uracil and cytosine in aqueous solutions. By depolymerising DNA, ionising radiations prevent DNA replication and stop cell division.

Several mechanisms have been proposed to explain the effects of X-rays and gamma rays. They can break different kinds of chemical linkages and damage genetic material in a variety of ways. Figure 20.2 shows that the effect may be direct or indirect. When a hydrogen atom consisting of one proton and one electron is ionised, the free electron may directly interact with DNA. Or the electron may interact with a molecule of water to produce OH, a free radical which can cause damage to DNA in the same way as the free electron. The following types of destruction of DNA are possible; hydrogen bonds may break between chains; a base may be changed or deleted; a single or double chain fracture may occur; cross linking might take place within the double helix; a deoxyribose may become oxidised.

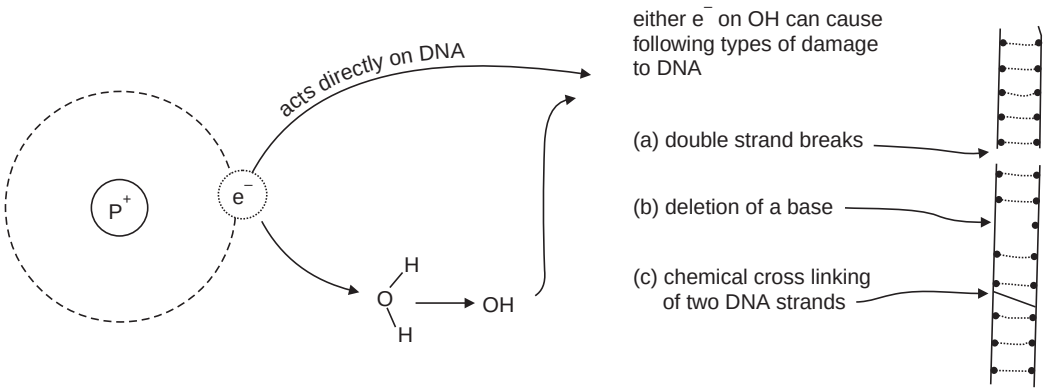


Fig. 20.2 Diagram showing types of damage to DNA by ionising radiation.

If a cell is irradiated in the S phase, DNA replication is inhibited resulting in failure of cell division and cell death. But if the cell is irradiated during mitosis or in *G*1, in that case DNA replicates normally but mitosis is delayed. Ionising radiation causes breakage and rearrangements in chromosomes which may interfere with normal segregation of chromosomes during cell division. When breaks in two different chromosomes in a cell occur close together in time and space they can join to produce chromosomal aberrations such as inversions, translocations and deletions. Micro-organisms are more resistant to ionising radiation than higher organisms. It is found that D_{37} dose, that is the radiation dose to a cell population with 37% survival is about 2000 to 30000 rads in bacteria. In human cells D_{37} is about 120 rads.

Some chemicals have a protective effect on the cell in reducing the effect of a radiation dose. Aminothiols which have an – SH and – NH_2 group separated by two carbon atoms are most powerful in reducing the effect. The protective effect is expressed as dose reduction factor (DRF). DRF is the ratio of $LD_{50(30)}$ for protected animals to $LD_{50(30)}$ for unprotected animals. *LD* is the lethal dose or the amount of radiation that kills all individuals in a large group of organisms. $LD_{50(30)}$ is the dose which kills 50 % of organisms within 30 days of exposure. LD_{50} for dog is estimated to be 350 rads, for mouse 550, goldfish 2300.

Whether the natural background radiation, though small in amount is dangerous for human beings or not has been questioned. The background radiation consists mainly of cosmic rays, emissions from radioactive elements in the earth such as uranium, radium and thorium, as well as emissions from radioactive isotopes (carbon 14, potassium 40) occurring naturally in the body. People living at sea level receive an average dose of about 0.8 millisievert of radiation per year (Upton, 1982). A study of the coastal area of Kerala in South India, a region having high background radiation, has revealed a high incidence of Down's syndrome (condition described in Chapter 21) in the population. Radiation-induced genetic and chromosomal anomalies were also observed (Kochupillai *et al.*, 1976).

The Atomic Bomb Explosions at Hiroshima and Nagasaki

On August 6, 1945 the atomic bomb exploded over Hiroshima killing over 78000 people and leaving many affected survivors who were studied for the effects of radiation on human beings. On August 9, a second atomic bomb was exploded over the city of Nagasaki. The words hibakusha (explosion affected person) and higaisha victim or injured person became understood to people in the world.

A number of individuals who had received extensive radiation had no children for several years. Many of the survivors have shown visible chromosome abnormalities such as breaks and translocations. Persons over 30 years of age who had received more than 200 rads of radiation were found to be more sensitive to radiation than persons of the younger age group. About 2 per cent of survivors developed leukemia within the following decade. The children of survivors have not shown a detectable increase in genetic abnormalities.

UV RADIATION

E. Altenberg was first to show that ultraviolet radiation can induce mutations. UV rays have longer wavelengths than X-rays and gamma rays, hence they cannot penetrate tissues as deeply. Their mutagenic action is limited to bacteria, fungal spores or other free cells whose genetic material lies very near the irradiated surface. L.J. Stadler induced mutations in pollen grains of maize plants. He found that UV rays with a wavelength of about 260 nm (Fig. 20.3) were more effective in mutagenesis. In fact the wavelength most readily absorbed by DNA also happens to be 260 nm, thus proving a direct correlation between UV induced mutations and DNA. The sun is a powerful source of UV. As such the sun's rays would be expected to cause widespread damage through mutations in all living organisms. But fortunately a layer of ozone in the upper atmosphere

absorbs most radiation below 290 nm. The UV rays falling on DNA are absorbed by the pyrimidine bases especially thymine. Cross linking between adjacent pyrimidines takes place to form thymine *dimers*.

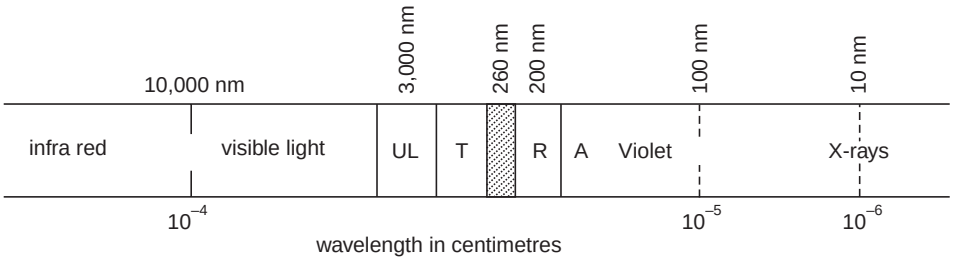


Fig. 20.3 Diagram of wavelengths of different types of rays.

The two thymines join at their 4 and 5 positions to form a dimer. Dimer formation can take place between two thymine residues (TT), cytosine (CC) or uridine (UU) or two different pyrimidines (CT). When cytosine is exposed to UV a molecule of water is added across the double bond between the fourth and fifth carbon atoms (Fig. 20.4).

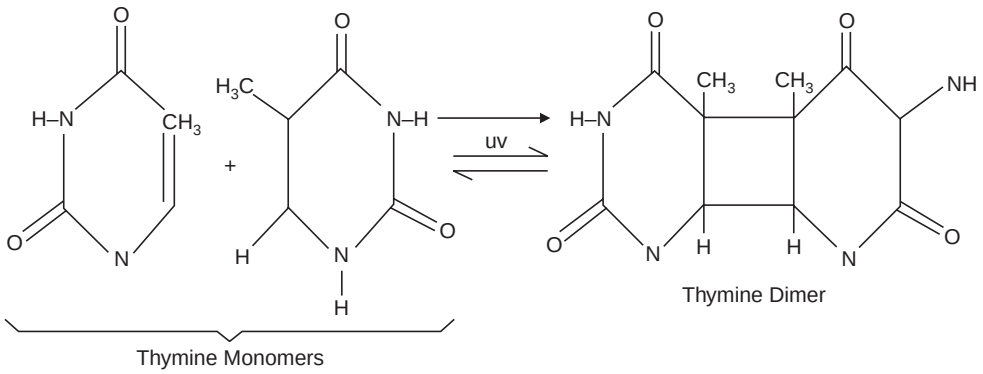


Fig. 20.4 Formation of thymine dimers.

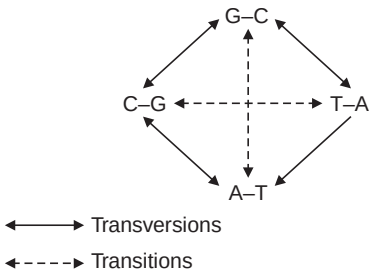
When heated or exposed to acidic conditions, the hydrated photoproduct can revert to the original form. If allowed to remain as such long enough, the hydrogen bonds between the pyrimidine and purine on the complementary strand break, leading to strand separation in that region. Both dimerisation and hydration of double stranded DNA affect DNA replication.

CHEMICALS AS MUTAGENS

Molecular Basis of Point Mutations

Mutations which alter nucleotide sequences within a gene are of two types: *base pair substitutions* and *frame shift mutations*. In base pair substitution one base pair, for example *AT* may be replaced by another such as *CG* or *GC*. These are of two further types, namely *transitions* and *transversions*. Transitions are base changes in which a purine is substituted by another purine

as when $A = T$ pair is replaced by a $G = C$ pair or *vice versa*; or when a pyrimidine is replaced by a pyrimidine such as when $T = A$ is replaced by $C = G$, or *vice versa*. Transversions are alterations in bases in which a purine is substituted by a pyrimidine, that is, when an $A = T$ pair is replaced either by $T = A$ or $C = G$, and *vice versa*.



In frame shift mutations, (so called because they shift the normal reading frame of base triplets in *m*RNA) single base pairs are *deleted* from or *added* to DNA in interstitial position. The genetic code requires reading of consecutive base triplets from a fixed starting point. If a single nucleotide is inserted or deleted, it shifts the reading frame, and all the subsequent triplets are read off differently. The entire portion of the polypeptide chain after insertion or deletion is translated wrongly, resulting in a "non-leaky" phenotype. Frameshift mutations therefore differ from base pair substitutions which produce "leaky" phenotypes by altering only one nucleotide in a single triplet, so that only one amino acid is wrongly placed.

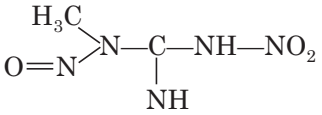
Nonsense and Missense Mutations: When a mutation in a triplet changes the codon so that it is recognised by another amino acid, it is called *missense* mutation. But if the mutation changes the codon for a specific amino acid into one which signals chain termination (nonsense codon), it is called a nonsense mutation. Missense mutations occur more frequently than nonsense mutations, and usually result in single amino acid replacements in the polypeptide chain. Such a chain may still have biological activity. Nonsense mutations result in premature termination of polypeptide chains so that only fragments of chains are formed. The lengths of fragments depend upon the distance of the nonsense mutation from the starting codon.

Mutations by Chemicals

Alkylating Agents: This is the most powerful group of mutagens. Chemicals of this group bind *in vitro* to the *N*–7 position of guanine. *In vivo* the 0–6 position of guanine and the 0–4 of thymine are preferred. They transfer alkyl groups to the nitrogen atoms of the bases in DNA. Alkylation of guanine causes ionisation of the molecule and changes its base pairing specificities. Alkylated guanine pairs with thymine instead of cytosine. Thus on replication there is a transition from a G = C to an A = T base pair. Alkylation of purines leads to hydrolysis of the sugar base linkage so that the purine base is lost from the backbone of the DNA molecule producing apurinic gaps. When DNA replication takes place, almost any base may be inserted in the gap. Insertion of a wrong base in a gap produces transitions as well as transversions. There are some bifunctional alkylating agents which form cross links between guanines on the same or opposite strands of the double helix. This causes more frequent production of apurinic gaps.

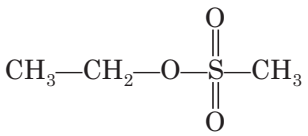
Examples of alkylating agents are nitrosoguanidine, mustard gas and mustard compounds, ethyl methane sulphonate (EMS) and ethyl ethane sulphonate (EES), alkyl halides, sulphuric and

phosphoric esters, ethylene imines and amides, and others. They are said to be electrophilic (electron deficient) reactants because they combine with nucleophiles which have electron rich centers. These compounds are also described as radiomimetic because their effects resemble those of ionising radiation.



N-methyl-N′-nitrosoguanidine

$Cl—CH_2—CH_2—S—CH_2—CH_2—Cl$ Sulphur mustard



Ethyl methane sulphonate (EMS)

EMS is one of the most powerful mutagenic agents known. It can add an ethyl group ($-CH_2CH_3$) to a guanine, and much less frequently to adenine. Due to ethylation guanine pairs with adenine leading to $A-T \rightleftharpoons G-C$ transitions. EMS also produces apurinic gaps into which any of the four bases may be inserted giving rise also to transversions.

Base Substitution by Tautomerism: The purine-pyrimidine base pairs in double helical DNA are determined mainly by the positions of hydrogen atoms which cross link the bases. Normally A pairs with T and G with C. The bases however, can exist in alternative forms due to rearrangements (*tautomeric shifts*) in the hydrogen atoms. *Tautomers* are rare and unstable and can revert to the common form. Watson and Crick suggested that if a base was present in its tautomeric form at the time of DNA replication, then a wrong base would be synthesised in the new strand. At the next replication cycle, the tautomer would revert back to its normal form, the two strands would separate, and this time the normal correct base would be synthesised on the new strand. This would result in substitution of one base pair for another, *i.e.,* $A \rightleftharpoons G$ and $T \rightleftharpoons C$ transitions would occur.

Base Analogues: These are chemicals with structure similar to bases in normal DNA and become substituted for normal bases during DNA replication. The first base analogue studied was 5-bromouracil (5-BU) which has chemical structure similar to that of thymine (5-methyluracil) except that the methyl group of thymine is replaced by bromine.

If 5-BU is supplied to the medium containing growing bacterial cells 5-BU gets incorporated into DNA instead of thymine. The cells remain alive and grow, and since 5-BU has the same pairing properties as thymine, it does not lead to mutation at all. However, there are two tautomeric forms of 5-BU, the normal keto form and the rare enol form (Fig. 20.5). When the rare enol form becomes incorporated into DNA, it pairs with guanine instead of adenine due to its hydrogen-bonding properties. The keto form however pairs with adenine. The enol form is short-lived and will eventually return to the keto form. When DNA undergoes replication in the presence of thymine, the strand containing 5-BU (now in keto form) will synthesise an adenine in the complementary strand. In the next round of replication, the *A* strand will synthesise a *T* strand opposite it. In this way a *GC* pair would be replaced by an *AT* pair resulting in mutation of the *transition* type.

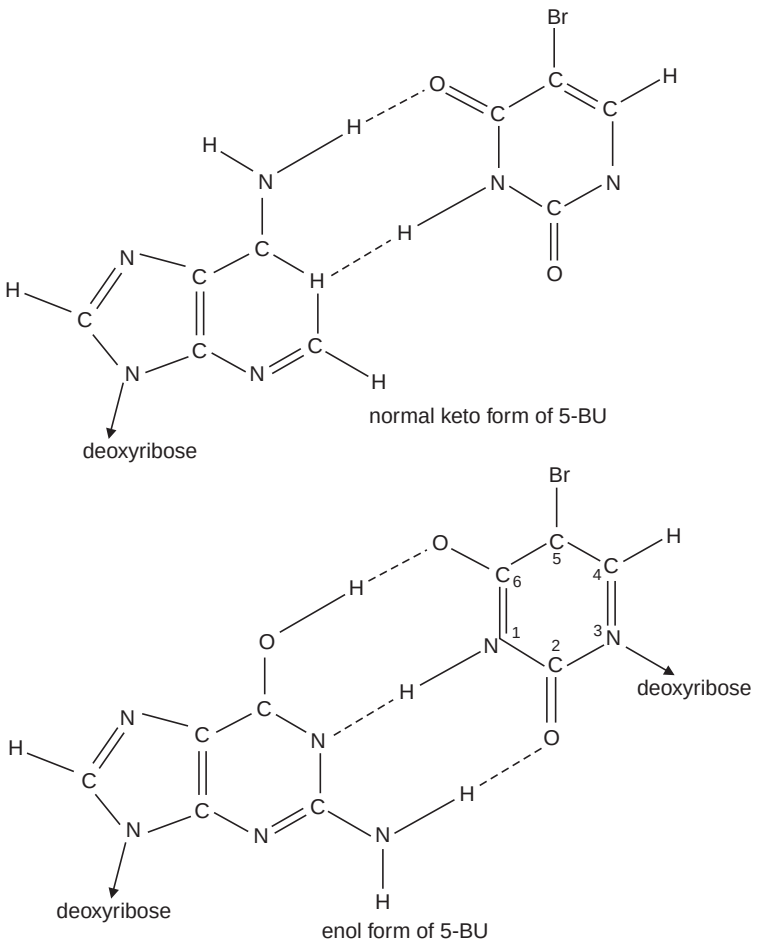


Fig. 20.5 Normal and tautomeric form of 5-BU.

A similar case is that of *2-aminopurin*e (2–*AP*) which is a chemical analogue of adenine inducing AT $\rightleftharpoons$ GC transitions. In the common form 2–*AP* pairs with thymine and there is no mutation. But when it exists in its tautomeric form (imino form) it forms two hydrogen bonds with cytosine. 2-*AP* therefore acts by first replacing adenine by shifting to its imino form, and then pairs with cytosine during replication. It also induces reversion in 5-BU induced mutants.

Nitrous acid (HNO_2) acts on non-replicating DNA by removing amino groups of nitrogenous bases, converting adenine to hypoxanthine, guanine to xanthine and cytosine to uracil (Fig. 20.6). The conversions lead to *AT* $\rightarrow$ *GC* and *GC* $\rightarrow$ *AT* transitions as explained in the Figure. In a similar way *hydroxylamine* (NH_2OH) changes hydroxymethylcytosine into uracil leading to *AT* $\rightarrow$ *GC* transition. These mutations are able to reverse their own effects (back mutation) as well as those of other base analogues.

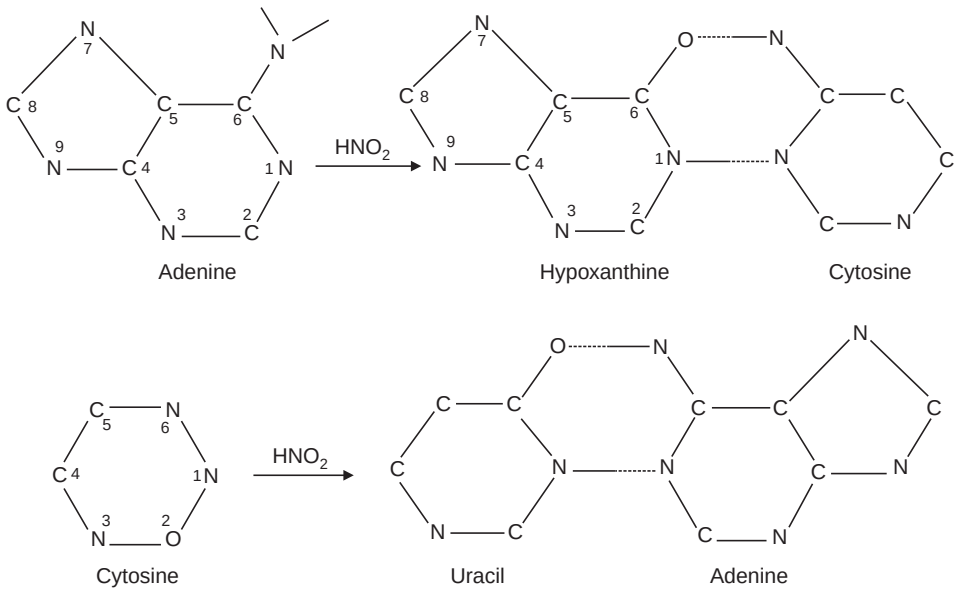


Fig. 20.6 Action of nitrous acid. It converts adenine to hypoxanthine which pairs with cytosine; cytosine is converted to uracil which pairs with adenine.

The acridine dyes act by intercalating themselves in DNA. Intercalating agents are planar polycyclic molecules which act by inserting themselves between the stacked base pairs of the double-stranded DNA molecule. This results in doubling of the distance between adjacent base pairs. Acridine mutations show a high rate of spontaneous reversion. Usually reversion is due to a second suppressor mutation within the same gene that carries the primary mutation. Acridine induced mutations are *'nonleaky'* as they result in total loss of function of the gene product. Acridines include important fluorochromes and antiseptics, phenanthridines (like ethidium bromide) used as trypanocides, and polycyclic hydrocarbons which are important carcinogens.

Certain chemicals like ethoxycaffeine, urethane and formaldehyde produce organic peroxides and free radicals leading to mutations. They probably cause destruction of nucleic acid bases resulting in breaks in single strands.

REVERSE MUTATIONS

When a mutation changes the wild type normal genotype to a mutant type, as is more often the case, the event is called a *forward* mutation. This is in contrast to reverse mutations in which the mutant genotype changes to the original wild type. In micro-organisms auxotrophs reverting to prototrophs are easily detected by plating cells of the originally auxotrophic strain on minimal medium. In bacteriophage *T*4 the *rII* mutants (strains which cause rapid lysis of host cells) frequently revert to the wild type rII^+ strain. In *Drosophila* a number of recessive mutant genes are known to revert to the wild type though with a lesser frequency than the forward mutations. The ability of mutant genes to revert suggests that mutation at least in some cases is not a permanent, irreversible process.

Reverse mutations could occur in different ways. In a true reverse mutation, the original base pair sequence of the wild type may be restored. Thus if a *GC* pair of the wild type sequence is

replaced by an *AT* pair to produce a forward mutation, a true reverse mutation could again substitute a *GC* pair in that position. Sometimes a different base pair may be inserted at the site of the altered pair which had produced the forward mutation. Thus when GC is replaced by *AT*, the reversion may be due to substitution by *CG* instead of *GC*. This produces a reverse phenotype even though its sequence differs from the wild type in a single base pair.

Sometimes an apparently reverse mutation is due to a second suppressor mutation which suppresses the effect of the primary mutation so that the phenotype appears like the wild type. There may be *intragenic* suppression when the second mutation occurs *within* the gene carrying the first mutation but in a different site. Or suppression may be *intergenic* (extragenic) when the second mutation lies in a different gene. In both types of suppression, the second suppressor mutation produces functional products of the gene which carries the first or primary mutation. For example suppose gene *A* is not able to produce A protein due to a mutation. A suppressor mutation in the same or in a different gene could result in the production of A protein, thereby reversing the mutation in gene *A*. In the case of intergenic suppression, the term *suppressor gene* denotes the gene which has the second mutation suppressing the primary mutation in another gene. The products of suppressor genes are usually components of the translation system; the suppressor molecules are frequently *t*RNA molecules.

In *intragenic* suppression, reversion to the normal phenotype might be caused by deletion and insertion of single nucleotides in the same gene. Thus if the reading frame of triplets is shifted due to a single deletion causing primary mutation, the frame would be restored if a single insertion occurs at the site of the second mutation. Only the triplets between the deletion and insertion would add incorrect amino acids, the rest of the chain would be normal. If the site of the second mutation is close to the first, the number of wrong amino acids would be small, and a full length functional protein would be produced.

Intergenic suppression by suppressor genes is due to changes in the process of translation. Most of the suppressor genes result from mutations in the tRNA genes and their products are mutant *t*RNA molecules. There are suppressor genes for mutations in each of the three chain terminating codons namely *amber* (UAG), *ochre* (UAA) and *opal* (UGA). Normally amber, ochre and opal are nonsense mutations located at sites within a message and result in fragments of polypeptide chains. The suppressor gene for each nonsense mutation allows insertion of an amino acid at the site of the nonsense codon and the chain is not terminated. The suppression of chain terminating mutants is brought about by reading of the nonsense codon as if it were a sense codon. For example, one kind of amber suppressor inserts the amino acid tyrosine against the nonsense codon UAG. The suppressor gene produces a mutant tyrosine-specific *t*RNA. The normal tyrosine *t*RNA has the anticodon GUA and recognises the triplets UAC and UAU in mRNA. The anticodon in the mutant *t*RNA is changed to CUA which pairs with UAG and tyrosine is inserted instead of the chain terminating precociously.

Another example illustrates intergenic suppression in *E. coli.* A primary amber mutation in *E. coli* had changed a base triplet into a nonsense codon UAG. Some other cells of *E. coli* carrying the same mutation had an intergenic amber suppressor mutation in a gene coding for serine *t*RNA. The suppressor gene produced mutant *t*RNA molecules with an altered anticodon that could recognise AUG. Thus the suppressor gene could insert serine at the site of the nonsense triplet AUG and full length polypeptide chains could be formed. Both UAG and UGA suppressors act by changing the anticodon in a specific tRNA and producing mutant suppressor tRNAs. The mechanism of UAA suppression is not fully known.

The amber suppressor gene also explained the occurrence of conditional mutants in *T*4 bacteriophage. The amber mutants of *T*4 were found to grow on one strain of *E. coli* (called

permissive host) but not in another strain (called *restrictive host*). This was because the permissive strain of *E. coli* carried the amber suppressor gene which reversed the effect of the amber mutation. Cells of *E. coli* which do not have the amber suppressor gene also show the amber mutation.

REPAIR OF DNA

Mutations cause damage to DNA, and if the genetic material has to continue to function normally, it must be repaired by special mechanisms. The term *lesion* indicates the damage caused initially. The various DNA repair mechanisms have been more extensively studied in prokaryotes than in eukaryotes and no essential differences seem to be present between the two groups.

Photoreactivation or Photorepair

This mechanism was discovered when bacteria that had been heavily irradiated with UV rays were found to have recovered their colony forming ability when stored in the presence of visible light. UV rays are a major source of damage to DNA and have been studied extensively for repair of DNA lesions (pyrimidine dimers). UV induces joining of adjacent pyrimidines on the same DNA strand by the formation of a cyclobutane ring resulting from saturation of the double bonds between carbons 5 and 6. The presence of dimers in DNA blocks transcription or replication at the site of damage. One method of repairing UV-induced dimmers involves reversal of dimerisation reaction. The process is called **photoreactivation**. It uses energy of visible light to break the ring structure. Thus the original pyrimidine bases remain in DNA but are restored to their normal state of individual pyrimidines. Repair of UV-induced damage by photoreactivation is common in prokaryotes and eukaryotes. A peculiar feature is that, photoreactivation may not be present in all tissues of an organism. In the crab *Gecarcinus* it is absent in the midgut gland but present in most other tissues. In chickens it is present in the fibroblasts and brain, but is absent in liver, kidney, skeletal muscle and some other tissues. Many species of animals as well as humans do not have this mechanism of DNA repair.

Dark or Excision Repair

Many bacteria can repair UV induced damage during storage in the dark. There are some mutant strains of *E. coli* which are extremely sensitive to UV. The bacterium *Micrococcus radiodurans* is highly resistant. After UV irradiation both sensitive and resistant bacteria show similar numbers of pyrimidine dimers in their DNA indicating differences in their repair mechanisms. When irradiated bacteria are incubated, the resistant strains show excision and release of pyrimidine dimers whereas sensitive strains do not.

It is now known that the process of excision repair involves several enzymatic steps: cutting of a single strand near the dimer; removal or excision of several bases including a dimer; synthesis of a new single strand complementary to the one opposite the excised dimer and linking the new strand to the original one (Fig. 20.7). The excision repair mechanism is also effective on other types of damage produced by UV, X and gamma radiation and on bases modified by alkylating agents in eukaryotic organisms.

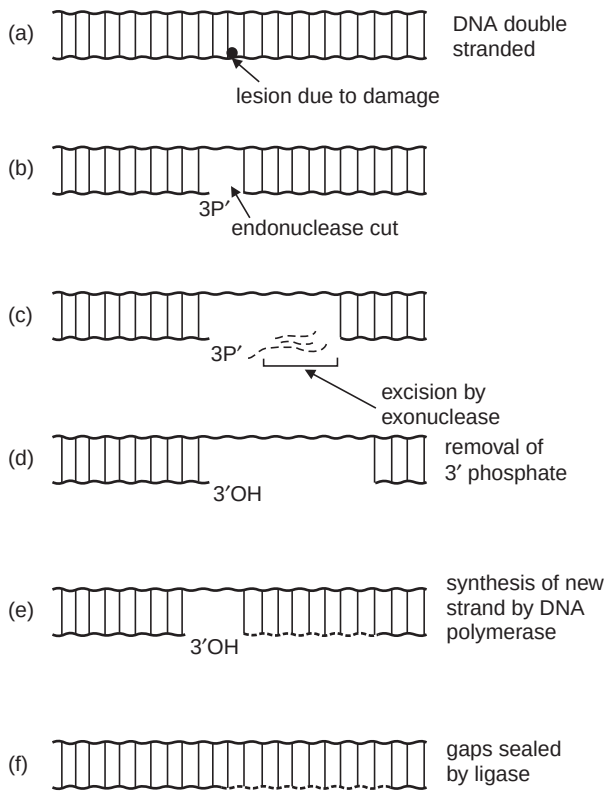


Fig. 20.7 The mechanism of excision repair.

Xeroderma pigmentosum (XP) is an autosomal recessive hereditary skin disease of man much studied for excision repair. Affected persons develop skin cancer when exposed to sunlight. When cultured cells from XP patients are irradiated with UV, they are found to be more sensitive than normal cells. The sensitivity has been associated with defective excision repair, XP cells showing 10–70% repair as compared to normal cells. Later studies by Cleaver (1969) and others showed that XP cells were deficient in the initial stage of excision repair, which involves production of single strand breaks by endonucleolytic enzyme. Normal cells could produce large number of single strand breaks which later disappeared. XP cells produced no such breaks. Thus pyrimidine dimers are not lost from DNA of irradiated XP cells.

Nucleotide Excision Repair

The repair of damage by removal of dimers by excision, described above, is also called **nucleotide excision repair** because the damaged bases, the pyrimidine dimer, are removed as part of a stretch of oligonucleotides. In *E.coli*, nucleotide excision repair involves protein products of three genes, *uvrA*, *uvrB*, *uvrC*. The protein uvrA recognises damaged DNA and recruits *uvrB* and *uvrC* to the damaged site. *Uvrb* and *uvrC* then cleave on the 3′ and 5′ sides of the damaged site, respectively, so as to excise an oligonucleotide consisting of 12 or 13 bases. The *uvrABC* complex is called the **excinuclease** owing to its ability to excise an oligonucleotide. A helicase enzyme then removes the damage-containing oligonucleotide from the DNA molecule, the resulting gap is filled by DNA polymerase I and joined by ligase. Nucleotide excision repair system has been found

in eukaryotes, and studied extensively in yeasts and humans. In yeasts several genes for DNA repair, called RAD (radiation sensitivity) genes have been identified. (In humans, DNA repair genes have been described through studies on patients with Xeroderma pigmentosum described above. At least seven different repair genes, XPA , XPB, ... XPG that are mutated in patients with Xeroderma pigmentosum. The enzymes encoded by these genes have been identified.

Mismatch Repair

A third excision repair system recognises mismatched bases that are incorporated during DNA replication. Many of the mismatched bases are removed by the proofreading activity of DNA polymerase. The remaining ones are corrected by the **mismatch repair** system which scans newly replicated DNA.

Postreplication Repair

In a strain of *E. coli* which was defective for excision of pyrimidine dimers from its DNA, it was found that the length of the newly synthesised DNA was equal to the distance between pyrimidine dimers. Obviously DNA replication is inhibited in regions where pyrimidine dimers are present. But if the cells are incubated subsequently the average length of newly synthesised DNA increases until it is similar to that found in unirradiated cells. The mechanism for repair is not well understood. Perhaps there are several processes involved in restoring normal replication between pyrimidine dimers, and the resulting gaps are subsequently joined.

There is some information on postreplication repair in eukaryotes. If mouse and Chinese hamster cells (which have very little excision repair) receive UV irradiation in two fractions, cell survival is greater than if the whole dose is given in one fraction. This is true only if the cells pass through the S phase between the two fractions. It appears therefore that the repair process and the S phase are associated.

It has been shown by several workers that the essential features of post replication repair in mammals are synthesis of new DNA strands in S phase leaving extensive gaps which are eventually closed. Since pyrimidine dimers inhibit DNA synthesis, the number of gaps corresponds to the number of pyrimidine dimers. The size of the gaps is approximately 800–1000 nucleotides long. The process of filling the gaps is inhibited by hydroxyurea, caffeine and theophylline.

In *E. coli* post replication repair is associated with genes at the *rec* loci namely *rec A, rec B* and *rec C*. A mutation in any of these loci results in loss of ability to integrate donor cell DNA by recombination. These loci are also sensitive to the effects of UV radiation.

SISTER CHROMATID EXCHANGES

Sister chromatid exchanges (SCEs) represent the interchange of DNA replication products at apparently homologous chromosomal loci. In the recent years SCE analysis has gained importance as a sensitive method for study of DNA damage; it seems that agents which induce SCEs are also active as mutagens and carcinogens (cancer causing agents). Some human genetic diseases deficient in DNA repair mechanisms show abnormalities in SCE formation and predisposition for cancer.

SCEs are induced in cells by incorporation of BrdU into DNA. The cells are treated with BrdU for one or two cycles. They are harvested at metaphase after the second cycle, stained and analysed. Besides BrdU, alkylating agents and proflavine also induce SCEs.

The exact mechanism resulting in SCE formation is not known. SCE analysis is useful for estimating the cytogenetic impact of some drugs given to patients in chemotherapy. A higher frequency of SCEs has been found in human beings exposed to environmental pollutants, or have cigarette smoking habits (Lambert, 1978).

MUTATIONS AFFECTING HUMAN BEINGS

Enzyme Functions and Proteins: A mutation in DNA can alter the primary structure of a corresponding polypeptide chain. Sickle cell haemoglobin (HBS) is the earliest example of a base pair substitution leading to an amino acid substitution in man. The proteins albumin, transferin, G6PD and carbonic anhydrase are further examples of amino acid substitutions. Mutations which cause substitutions of unlike amino acids, such as valine for a glutamate produce drastic phenotypic effects. Small deletions in hemoglobin (structural variants) lead to instability of the protein molecule by producing stresses that lead to rapid denaturation at lower temperatures. A group of mutant hemoglobins have additional residues attached to the C-terminal end of either the alpha or beta chain.

Human Disorders Susceptible to Genetic Damage: There are some human genetic disorders in which the cells are highly sensitive to DNA-damaging agents and the patients are prone to cancer. Many of these disorders follow the autosomal recessive mode of inheritance. Xeroderma pigmentosum (XP) mentioned earlier shows deficiency in cells to repair DNA damage by UV. There is a connection between the genetic damage and the symptoms of the disease (sunlight induced lesions). Bloom's syndrome, Fanconi's anaemia, Cockayne's syndrome and ataxia telangiectasia also show susceptibility of cells to genetic damage.

Hereditary retinoblastoma has a dominant mode of inheritance and the gene shows 90–95 % penetrance. Cells are sensitive to DNA damaging agents and patients show increased incidence of primary tumors.

ENVIRONMENTAL MUTAGENS AND CARCINOGENS

Mutagens and carcinogens may cause mutation of germ cells resulting in the accumulation of heritable abnormal genes in the population. Mutations in somatic cells may give rise to malignant cells in individuals.

Man has introduced many mutagens and carcinogens into the environment and nature has also provided these through plants and molds. Among the naturally occurring agents, some *fungi* produce *aflatoxin B1*, one of the strongest mutagens and carcinogens. Bracken fern (*Pteris aquilina*) which is eaten in Japan, induces tumours of the bladder in cows and rats. About 50 species of medicinal plants belonging to families Compositae, Boraginaceae, and Leguminosae contain pyrrolizidine alkaloids which are carcinogenic.

There are many flavonoids in plants of which quercetin, Kaempferol, rhamnetin and fisetin are strongly mutagenic. Flavonoids are mostly present in the outer tissues, skin and peels of

fruits, tubers and roots. Onion (*Allium cepa*) contains 11–24 gm of quercetin per kg fresh weight of outer epidermis. There is a high incidence of cancer among individuals taking large amounts of polyunsaturated fats (Pearce and Dayton, 1971). When polyunsaturated fatty acids are oxidised, malonaldehyde which is mutagenic is produced. Charged parts of broiled fish and meat are said to be mutagenic (Nagao *et al.*, 1977).

Tobacco tar in cigarette smoke is an important cause of human cancer. The mutagenicities of tobacco tars have been correlated with the contents of proteinaceous nitrogen and total nitrogen in the leaves.

Among synthetic compounds some food additives used for preserving meat, fish and other foods contain nitrofuran derivatives which are mutagenic. Certain cosmetics such as the synthetic hair dyes of the permanent type contain mutagenic compounds. Some of the anti-tumour antibiotics such as adriamycin, daunomycin, mitomycin-C and actinomycin-D are also mutagenic? Pesticides are known to contain potential mutagens.

QUESTIONS

1. Write short notes on (*a*) replica plating (*b*) somatic mutations (*c*) auxotrophs (*d*) mutable genes (*e*) sister chromatid exchanges.
2. Describe the mechanisms by which (*a*) X-rays can cause damage to DNA (*b*) frameshift mutations are caused (*c*) photorepair of damaged DNA can occur.
3. In how many different ways can a base in DNA be altered to produce a mutation? Give one example for each.
4. Distinguish between
   (*a*) nonsense and missense mutation
   (*b*) transition and transversion
   (*c*) ionising and nonionising radiation
   (*d*) spontaneous and induced mutation.
5. Define reverse mutations. Describe the mechanisms underlying intragenic and intergenic suppression of mutations.
6. Mention some potential mutagens in the environment that mankind is exposed to.
7. Most mutagens are carcinogenic (cancer causing) and *vice versa.* Discuss.

SELECTED READINGS

Ames, B. N. 1979. Identifying Environmental Chemicals Causing Mutations and Cancer. *Science* **204:** 587.

Auerbach, C. 1967. Changes in the Concept of Mutation and the Aims of Mutation Research. *In* Heritage from Mendel. Ed. Brink. R.A. Univ. Wisconsin, Press, Madison.

Auerbach, C. 1976. Mutation Research. Chapman and Hall, London.

Bootsma, D. and Hoejmakers, J.H. J. 1994. The Molecular Basis of Nucleotide Excision Repair. *Mutation Res*. **307:** 15–23.

Brink, R.A. 1973. Paramutation. *Annu. Rev. Genetics* **7:** 129.

Carrano, A.V. *et al.,* 1968. Sister Chromatid Exchange as an Indicator of Mutagenesis. *Nature* **271:** 551.

Drake, J . W. 1970. The Molecular Basis of Mutation. Holden. Day, San Francisco.

Ellis, N.A. 1996. Mutation-causing Mutations. *Nature* **381:** 110–111.

Hollaender, A. 1974. Chemical Mutagens. Plenum Press, New York.

Kochupillai, N. *et al.,* 1976. Down's Syndrome and Related Abnormalities in an Area of High Background Radiation in Coastal Kerala. *Nature* **262:** 60.

Latt. S.A. and Schreck, R.R. 1980. Sister Chromatid Exchange Analysis. *Amer. J. Human Genetics* **32:** 297.

Lehmann, A.R. 1995. Nucleotide Excision Repair and the Link with Transcription. *Trends Biochem. Sci*. **20:** 402–405.

Marx, J.L. 1978. DNA Repair: New Clues to Carcinogenesis. *Science* **200:** 518.

Reinhart, R.R. and Ratty, F.J. 1965. Mutation in *Drosophila melanogaster* Cultured on Irradiated Food. *Genetics* **52:** 1119.

Roth, J.R. 1974. Frameshift Mutations. *Annu. Rev. Genetics* **8:** 319.

Sancar, A. 1996. DNA Excision Repair. *Annu. Rev. Biochem*. **65:** 43–81.

Sutton, H.E. and Wagner, R.P. 1975. Mutation Add Enzyme Functions in Humans. *Annu. Rev. Genetics* **9:** 187.

Tanaka, K. and Wood, R. D. 1994. Xeroderma Pigmentosum and Nucleotide Excision Repair. *Trends Biochem. Sci.* **19:** 83–86.

Upton, A. C. 1982. The Biological Effects of Low Level Ionising Radiation. *Sc. Amer.* **246:** 29.

21

Human Genetics

When Mendelism was rediscovered, there was a natural tendency to apply Mendel's laws to inheritance of human traits. Shortly afterwards some heritable diseases of man were discovered. In 1908 Archibald Garrod could interpret inheritance of certain metabolic diseases in man (alkaptonuria and others) in Mendelian terms. In the decades that followed the genetic basis of conditions like haemophilia, brachydactyly, colour blindness, blood groups and a few others became well understood. A major breakthrough came in 1956 when J. H. Tjio and A. Levan established that the correct diploid chromosome number in man was 46 and not 48. The new technique developed by Moorhead *et al.,* (1961) for making chromosome spreads from cultured peripheral blood was exciting as it allowed identification of individual chromosomes (karyotype analysis). It also made possible the recognition of chromosome abnormalities. Thereafter, important developments in human cytogenetics occurred with great rapidity. Some significant advances were made when the association between chromosomal aberrations and abnormal human phenotypes were found out.

Meanwhile geneticists also found out that many of the genetic mechanisms existing in the lower groups (plants, animals and micro-organisms) could be applied to humans. This advanced our knowledge about the molecular basis of some human genetic diseases. Simultaneously pedigree studies have been done to follow the inheritance of a trait through several generations of a family. The technique of amniocentesis for prenatal diagnosis of a genetic disorder and foetal sex received much publicity. The importance of this technique in preventing future births of malformed children in 'high risk' families cannot be overemphasised. No less important is the technique of somatic cell hybridisation which holds great promise in the field of human gene mapping.

THE HUMAN CHROMOSOME COMPLEMENT

For karyotype analysis, a sample of peripheral blood is drawn from an individual, the leukocytes separated and cultured for about 3 days. To stimulate growth and cell division phytohaemaglutinin (PHA) is added. The dividing lymphocytes are arrested at metaphase stage by exposing them to

colchicine. The cells are then harvested and slides prepared. The metaphase spreads are photographed and the chromosomes cut out and rearranged according to size and location of centromeres. The study of a complete chromosome complement in this manner is called *karyotype analysis.* The karyotype of a normal human female is shown in Fig. 21.1.

The 23 pairs of human chromosomes are classified into 7 groups *A* through *G*. Group *A* includes chromosomes 1, 2 and 3, largest in size and metacentric. Group *B* has chromosomes 4 and 5, smaller than group *A* chromosomes, and submetacentric. Group *C* is largest containing chromosomes 6 through 12 and *X*, all of medium length and submetacentric. Group *D* has medium sized acrocentric chromosomes 13, 14 and 15. Group *E* chromosomes 16, 17 and 18 are shorter and either meta- or submetacentric. Group *F* has shorter metacentric chromosomes 19 and 20. Group *G* contains the smallest acrocentric chromosomes 21 and 22 as well as *Y*.

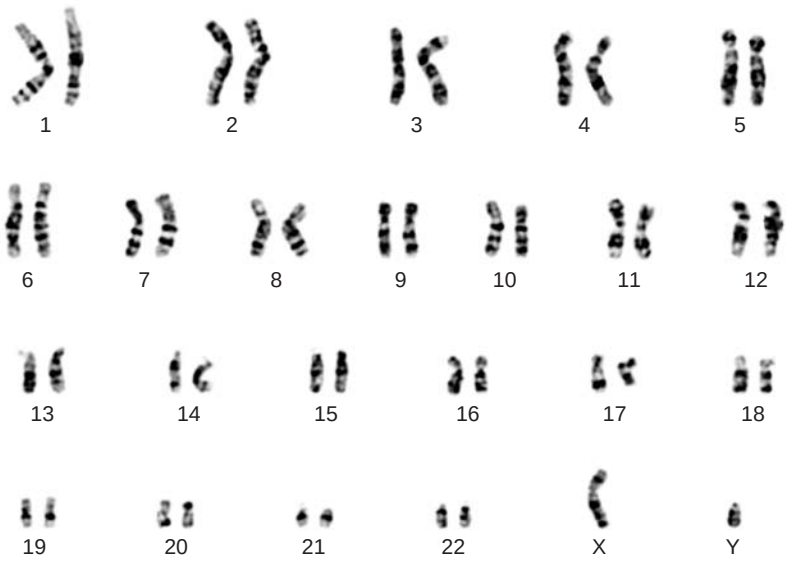


Fig. 21.1 Karyotype of a normal human male showing G-bands (Courtesy Arundhati Sharma).

The banding technique for chromosomes has proved useful in identifying abnormalities in chromosomes. It was first found by Caspersson and his associates in 1968 that metaphase chromosomes can be stained with fluorescent DNA binding agents such as quinacrine mustard to yield dark and light bands. Since then numerous staining methods have been found for producing specific banding patterns in chromosomes.

CHROMOSOMAL ANOMALIES AND HUMAN DISORDERS

A. Autosomal Anomalies

The first known case of a chromosomal anomaly associated with a clinical disorder was described by Lejeune. Gautier and Turpin in 1959 in Down's syndrome or mongolism. Characterised by mental retardation and a few other clinical symptoms, the condition is due to trisomy (presence of

an extra chromosome so that a certain pair is represented by three homologues instead of two) of an autosomal chromosome. So far three clinical syndromes have been found to be associated with trisomy of autosomal chromosomes. A few genetic disorders are related to some other aberrations of the autosomes and are described below.

1. *Down's Syndrome:* This is frequently due to trisomy of the *G* group chromosome 21 arising from nondisjunction during meiosis in one of the parents. The karyotype thus shows 47 chromosomes. *G* trisomy is the most common of all autosomal trisomies. Sometimes a translocation between a *D* group chromosome and 21, or between two G group chromosomes is associated with the syndrome (Fig. 21.2). Some patients of Down's syndrome show mosaicism. Metaphase spreads of such patients show two cell lines in peripheral blood, one cell line with normal chromosomes, another with 21 trisomy.

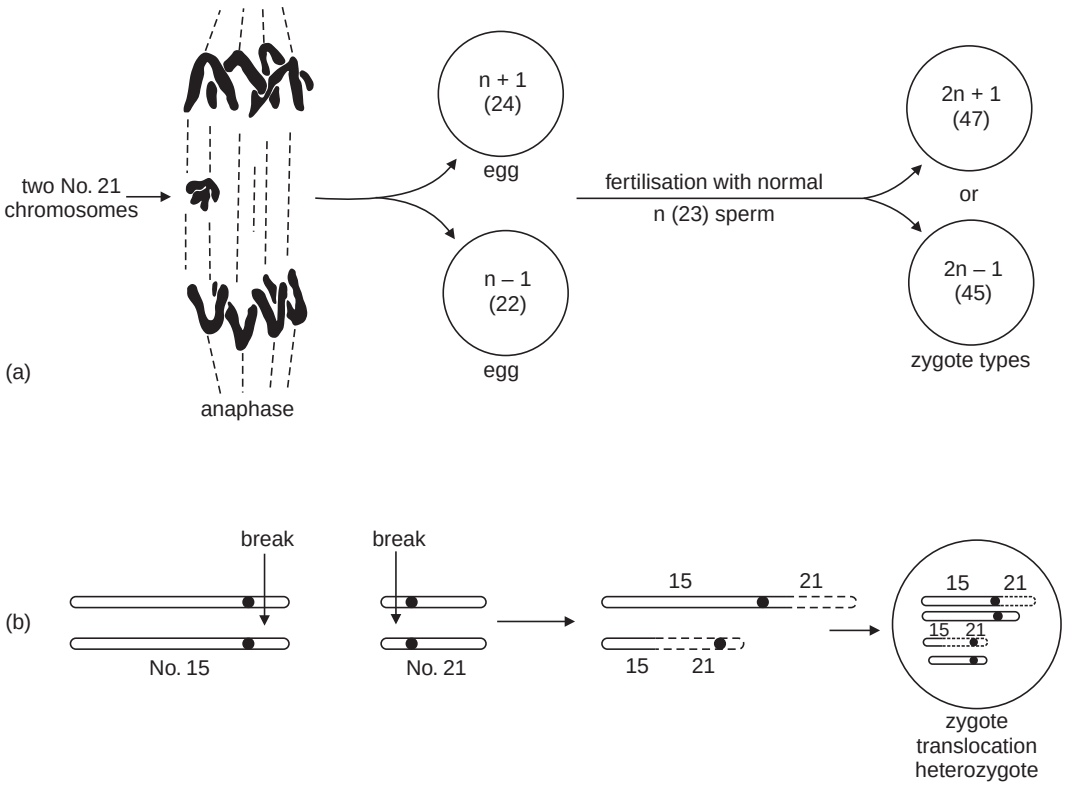
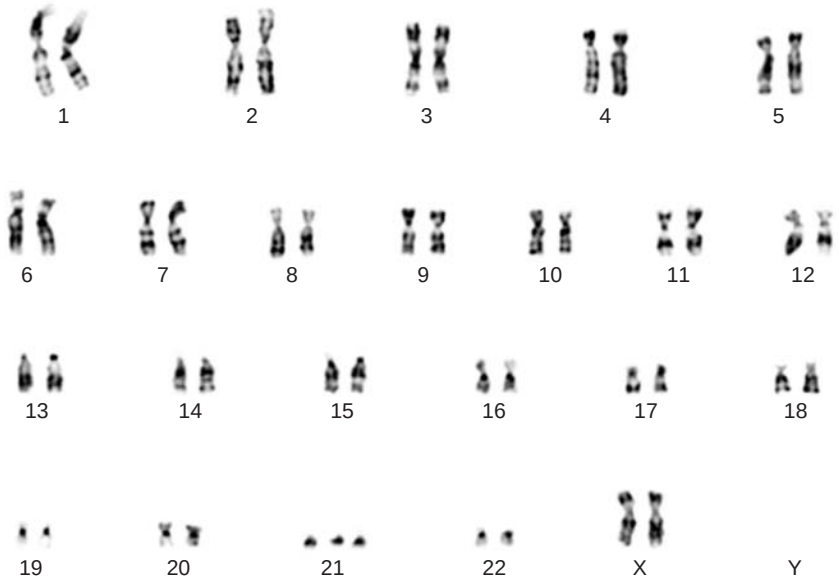


Fig. 21.2 (*a*) Origin of Down's syndrome through nondisjunction during meiosis and (*b*) through translocation.

About one in 700 births of both sexes could show Down's syndrome. The disorder is characterised by a typical mongoloid facial appearance, webbed neck, a flat nasal bridge, epicanthal fold (small folds of skin over inner corners of eye typical of the mongoloid race) brushfield spots around the iris, short fingers, and a gap between the first and second toe. Their striking feature is mental retardation with *IQ* ranging between 25 and 50 (normal average humans have an *IQ* of 80). Congenital heart disease and leukemia occur in many cases. Many have poor muscle tone during infancy. There is a higher incidence of Down's syndrome among children of older mothers. The patients live for a variable number of years. A Down's male is usually sterile, females are

fertile and rarely have produced offspring. The dermatoglyphic pattern (arrangement of lines on palm and fingers) shows in many cases a line called simian crease and distal axial triradius. Frequently all the ten fingers show ulnar loops.



Karyotype showing three no. 21 chromosomes (Courtesy Arundhati Sharma)

As in most genetic diseases there is no cure for a Down's patient. Affected individuals are usually institutionalised. However, after the birth of a mongol child it is necessary for the parents to have proper genetic counselling to prevent the birth of another child with mongolism. An accurate diagnosis through karyotype analysis of the affected child and both parents could provide an estimate of the recurrence risk. When mongolism is due to a translocation, the abnormality can be passed on to future generations through the gametes. A child that inherits the translocation is affected and could in turn produce victims of Down's syndrome. In contrast, Down's syndrome due to nondisjunction, which is a rare event during gametogenesis, is not familial and the condition is not inherited.

2. *Edward's Syndrome:* Edward and his colleagues in 1960 described a syndrome due to trisomy of an *E* group chromosome (16–18) and occurring more often in females than in males. The individual may also be a mosaic having a normal cell line and an 18 trisomic line. The incidence is about 1 in 3,500 live births.

The physical and mental growth is much retarded and death usually occurs in early childhood. Patients have hypertonicity of skeletal muscles resulting in a peculiar characteristic by which the affected person keeps the fingers tightly clenched against the palm of the hand. The typical features also include micrognathia (small jaws), deformed ears, small sternum and pelvis, a characteristic dermatoglyphic pattern and severe retardation.

3. *Patau's Syndrome:* In 1960 Patau and his associates described a clinical disorder originating from trisomy of *D* group chromosome (13–15). Most of the affected persons have a small head and abnormalities of the face, eyes and forebrain, cleft lip and palate, low set deformed ears, small chin, and the hands are often clenched in the manner described for Edward's syndrome.

In general the individual appears more severely malformed than in the previous two syndromes and the mean life span is only about 4 months.

4. *Cri du chat* (*cat cry*) *Syndrome:* An interesting abnormality in which the affected newborn cries in a manner resembling the mewing of a cat, was first described by Lejeune in 1963 in France, hence the name *cri du chat* (cat cry). A small head, widely spaced eyes, receding chin and congenital heart disease are some other typical features. The condition is due to a deletion in the short arm of the *B* group chromosome 5, and is very rare.

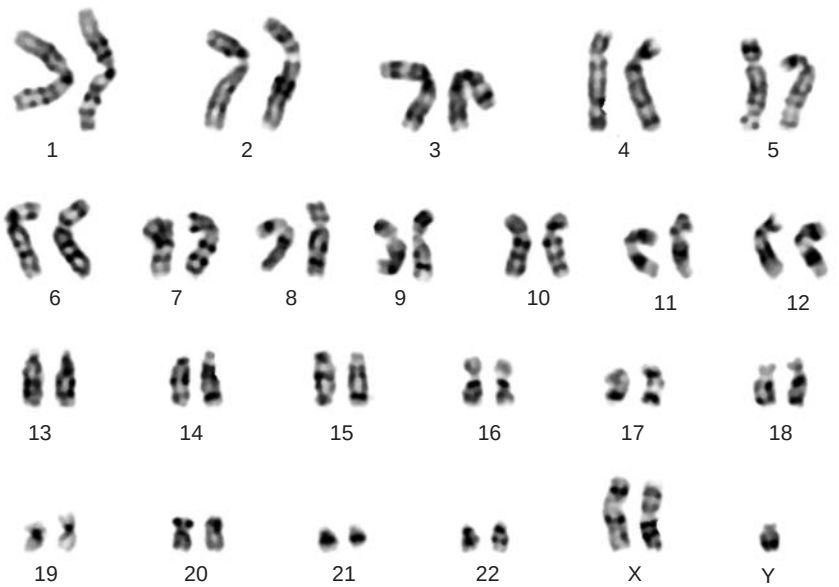
5. *Wolf Syndrome:* Another very rare syndrome is caused by a deletion in the short arm of a B group chromosome (no. 4). It is distinct from the cat cry syndrome.

6. *The Philadelphia* (*Ph1*) *Chromosome:* This is one of the best known cases of a specific association between a chromosomal anomaly and a human disorder. The bone marrow cells of most of the patients with chronic granulocytic leukemia show a deletion of the long arm of a *G* group chromosome (number 22) called Philadelphia or Phl chromosome. The banding technique has further revealed that a portion of the long arm of chromosome 22 is actually translocated to a longer chromosome (Rowley, 1973).

7. *Chromosome 13 and Retinoblastoma:* A deletion in the long arm of chromosome 13 is specifically associated with retinoblastoma, an autosomal dominant trait.

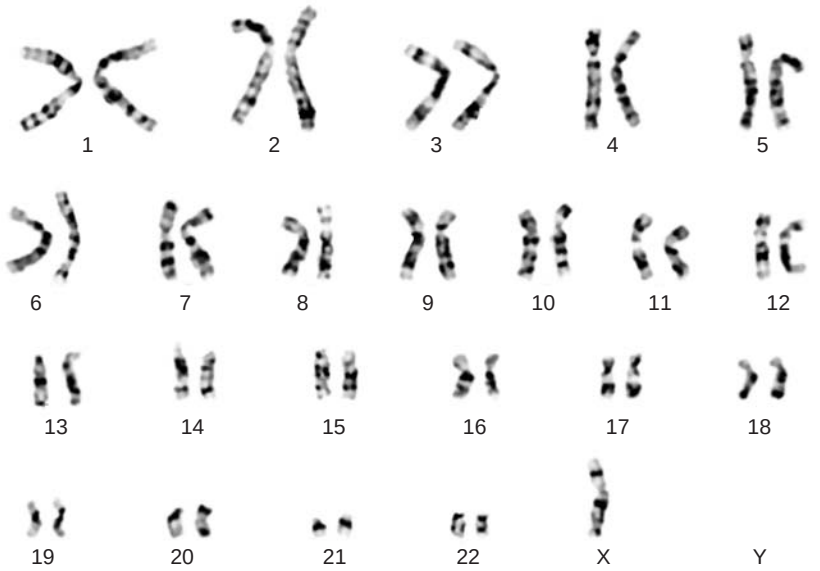
B. Sex Chromosome Anomalies

1. *Klinefelter's Syndrome:* In 1942 Klinefelter described a condition in phenotypic males which turned out to be due to an extra *X* chromosome (47, *XXY*). The affected individuals appear normal in childhood, the abnormalities becoming visible only in adult males. The syndrome is characterised by absence of spermatogenesis, gynaecomastia, and excessive secretion of gonadotropins in the urine. Since buccal smears of Klinefelter's males show Barr bodies, they are referred to as chromatin-positive males. Most of the patients are mentally retarded and develop a variety of psychiatric problems. Although many have the karyotype 47, *XXY*, some may have 48, *XXXY*, 49, *XXXXY*, or 48, *XXYY*, or they may be cytogenetic mosaics.



Klinefelter syndrome—47,XXY

2. *Turner's Syndrome:* This is shown by females characterised by a short stature, gonadal dysgenesis, sexual infantilism, webbed neck, prominent ears, cubitus valgus (increased carrying angle of the arms) dystrophy of the nails and hypoplastic nipples. Their sex chromosome constitution is *XO* and they have only 45 chromosomes. They are chromatin-negative females as they do not show Barr bodies. They are frequently mosaics with more than one cell line such as *XO*/*XXX*, *XO*/*XX*/*XXX*, and others. The incidence is one in about 5,000 births.



Turner syndrome—45,X
(Courtesy Arundhati Sharma)

3. *The XYY Male:* In 1965 Jacobs *et al* found that many of the men kept in institutions for the retarded due to aggressive and antisocial behaviour have 47 chromosomes with *XYY* sex chromosomes. They are usually tall but not always mentally retarded, frequently show hypogonadism and are sterile. The presence of two *Y* chromosomes can be recognized as two brightly fluorescent bodies by proper staining. The discovery of this syndrome received publicity because of the possible association of a chromosome anomaly with human behaviour.

4. *The triple-X Syndrome:* These are individuals with 3*X* chromosomes designated superfemales. They are mentally retarded, sexually normal and fertile. Although most triple-*X* females have 47, *XXX* karyotype, some may have 48, *XXXX*, 49, *XXXX*, and still others may be mosaics. They show 2, 3 or 4 Barr bodies in their buccal smears (always one Barr body less than the total number of *X*'s).

5. *Intersex:* Individuals with both male and female gonadal tissues are called hermaphrodites (from Greek Hermaphroditos, the son of Hermes and Aphrodite). Their karyotype analysis shows that they are mosaics having both *X* and *Y* chromosomes in their cell lines. Their buccal smears may or may not show a Barr body. Their external genitalia are often ambiguous, and they are almost always sterile.

The condition of pseudohermaphroditism is also included among intersexes. Such individuals are cytogenetically normal with 46, *XY* (male pseudohermaphrodites) or 46, *XX* (female pseudohermaphrodites) chromosomes and normal buccal smears for one sex only. But phenotypically they show both male and female characters. There are two classes. *Male* pseudohermaphrodites that have testes and either ambiguous or female-like external genitalia. The *female* pseudohermaphrodites have ovaries and either ambiguous or male like external genitalia. The pseudohermaphrodites have some defect in the biosynthesis of testosterone in the testes or in the adrenal glands or in both.

C. Chromosomes and Spontaneous Abortions

About 15% of all human pregnancies terminate before the 22nd week by spontaneous abortion. A study of aborted foetuses by D.H. Carr (1967, 1971) has shown that 5–6% of abortions are due to chromosomal aberrations. Triploidy is frequently observed in aborted foetuses. In fact the first cases of triploidy in humans were discovered from two aborted foetuses by Penrose and Delhanty in 1961. A characteristic clinical feature of triploid abortions is the hydatidiform mole of the placenta. Trisomy is also very common in abortuses. Trisomy for all the 22 autosomal pairs has been observed in abortuses. The most frequent however is trisomy 16. The *XO* constitution is common in abortuses, and unbalanced translocations are also observed (Boué and Boué, 1975).

D. Chromosomes in Disorders with DNA Repair Defects

There are four inherited disorders with DNA repair defects which show chromosomal anomalies. The affected persons are also predisposed to cancer. The most well studied is *Xeroderma pigmentosum* (XP described elsewhere). This is characterised by defects in repair of DNA lesions caused by UV rays and other mutagens. The patients often develop skin cancers. In Bloom's syndrome (BS) there is slow rate of DNA chain maturation during replication. Patients are sensitive to sun's rays and retarded in growth. Their cultured cells show an increased frequency of sister chromatid exchanges. In ataxia-telangiectasia (AT) the cells are not able to repair damage caused to bases in DNA by gamma rays. Fanconi's anemia (FA) affects bone marrow cells. There is faulty repair of cross links in DNA. In all the 4 disorders the cultured cells show one or the other chromosomal abnormality.

TRACING THE GENE IN FAMILY-PEDIGREE STUDIES

By recording the occurrence of a trait in several generations of a family, the inheritance of a gene can be studied. These are called pedigree studies and are based on the study of a pedigree chart. The individual presenting himself with a genetic disorder is called the propositus or proband. Male members of a family are shown in the chart by squares, females by circles. Sibs are represented horizontally on a line in the order of birth. Persons affected by the genetic disorder being studied are shown by filling the circle/square with black. Figure 21.3 shows the various symbols in a hypothetical pedigree chart.

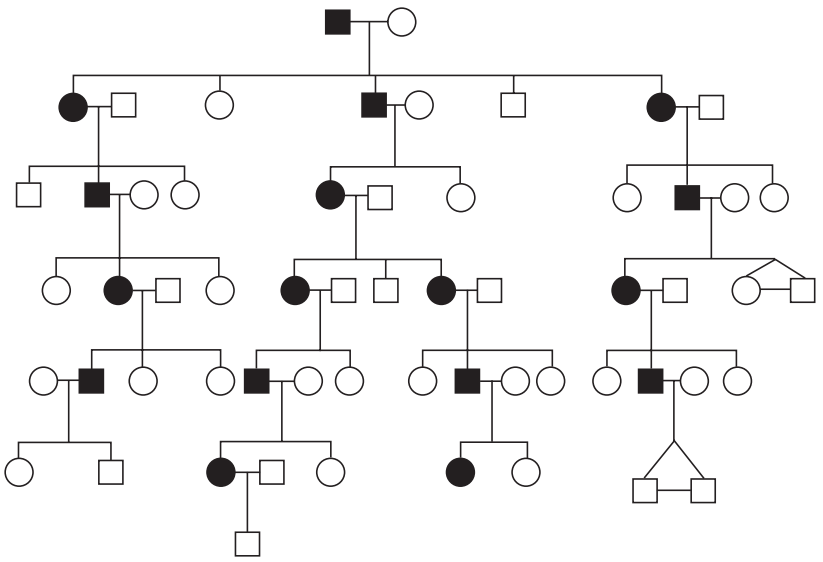


Fig. 21.3 A hypothetical pedigree chart.

Autosomal Dominant Inheritance

Farabee in 1905 gave the first description of a pedigree showing brachydactyly (short fingers and toes), an autosomal dominant trait . Some more examples are porphyria, Huntington's chorea (a degenerative disease of nerve cells), polydactyly, retinoblastoma (a malignant eye tumour of children) and others. The dominant transmission of a trait is also called manifest inheritance because whenever a gene is present its effect is produced.

Characteristics of Autosomal Dominant Inheritance

1. The trait appears in every generation.
2. An affected child must have at least one affected parent.
3. About one half of the offspring of an affected person are affected; the recurrence risk is 50% at each conception.
4. Both male and female persons are affected.
5. Individuals who appear normal do not transmit the trait to their offsprings.

Penetrance and Expressivity in Dominant Traits

Penetrance is a quantitative term and denotes the fraction of individuals carrying a gene that show the specific phenotype. Sometimes a pedigree shows absence of an affected person expected in a generation. The trait has skipped a generation because a heterozygous person is showing the normal phenotype. The gene though present was not penetrant. A gene is said to have low penetrance when it does not manifest in a large fraction of the individuals.

In other cases of dominant inheritance the gene may manifest in all the heterozygotes, but the *degree* of clinical *manifestation* may be different. This is called variable expressivity.

Autosomal Recessive Inheritance

In recessive inheritance the offspring inherits a trait from both parents. An albino offspring is born to parents who are normal in appearance, but each carries an albino gene (heterozygous *Aa*).

If each person had married a normal (*AA*) person, no albino would appear in their progeny. Abnormal recessive genes are thus transmitted for many generations through heterozygotes. Their existence is found out only when two heterozygotes marry and the homozygote appears, the ratio 1 normal : 1 affected.

The rare recessive conditions are more easily detected through consanguineous marriages which involve matings between blood relatives. In some parts of India consanguineous marriages are common, and the incidence of rare recessive disorders is also high. Microcephaly (small head), phenylketonuria, galactosemia and others are due to recessive genes.

Children affected with galactosemia are not able to metabolise galactose, which is a component of the milk sugar lactose. Normally, galactose is converted into glucose phosphate by the enzyme phospho-galactose uridyl transferase present in the liver. Homozygous children lack this transferase enzyme. The heterozygotes have an enzyme level intermediate between that of the normal and affected homozygotes. Affected babies have severe vomiting and diarrhoea, and consequently suffer from undernourishment and fail to grow normally. The condition can be treated if after birth the babies are kept on a lactose- and galactose-free diet, and are given specially formulated milk substitutes. Untreated children accumulate toxic amount of galactose–1–phosphate which lead to cataract in the lens of the eye, damage to liver and kidney tubules and some mental retardation.

Characteristics of Autosomal Recessive Inheritance

1. The trait is visible only in sibs, but not in their parents or other relatives.
2. The parents of an affected person may have been blood relatives (consanguineous).
3. About one fourths of the children of such parents are affected; the recurrence risk at each birth is 25 %.
4. Both male and female children have equal chance of being affected.

Sex-linked Recessive Inheritance

This type of inheritance is mostly *X*-linked and predominantly males are affected (due to hemizygous condition). Heterozygous females are carriers and are expected to produce affected and normal sons in the ratio 1 : 1. An affected male never produces an affected son. A famous example is haemophilia, the gene for which was passed on to the descendants of Queen Victoria. Some other examples are red green colour blindness, G6PD, Lesch-Nyhan syndrome and muscular dystrophy.

In haemophilia or bleeding disease, as it has been known for centuries, the blood is not able to clot within 4 to 8 minutes like the normal blood. Instead it takes an hour or more to clot. Failure to clot is due to the absence of a coagulation factor which is present in normal blood. Two different *X*-linked loci are involved in haemophilia. One causing the more prevalent haemophilia *A*, the other giving rise to haemophilia *B* or Christmas disease (so called because it was first noted in a person named Christmas). The coagulation factor which is absent in the *A* form of haemophilia is called antihaemophilic globulin (AHG), while in the *B* form the deficient factor is called plasma thromboplastin component (PTC). The affected individuals rarely live beyond the first decade, although with treatments available now they may live longer. The survivors still have problems due to internal bleeding in the joints. Being an *X*-linked recessive disorder, males are more frequently affected.

Haemophilia seems to have started in Britain's royal family through a mutation in one of Queen Victoria's parents. One of Victoria's four sons was affected and produced a carrier daughter.

Out of the two sons of this daughter, only one was a haemophiliac, the other was normal. Out of the 5 daughters of Victoria, two turned out to be carriers and produced in all 3 carrier daughters and 3 affected sons. The 3 carrier daughters further produced 6 carrier daughters and 5 haemophiliac sons.

Characteristics of *X*-linked Recessive Inheritance

1. Males are affected more frequently than females.
2. When the female parent is carrying the trait then 50% of her sons have a chance of being affected, and 50 % of the daughters would be carriers but phenotypically normal.
3. The trait is transmitted through several generations by carrier females.
4. The affected male parent cannot transmit the trait directly to his sons.

Sex-linked Dominant Inheritance

The *X*-linked dominant trait shows itself in hemizygous males and heterozygous females. All the daughters of an affected male also show the trait. An affected mother produces normal and affected children, both male and female in the ratio 1 : 1.

In *X*-linked dominant inheritance the males are affected more severely than females. For example dermal hypoplasia is apparently lethal in males. In females it produces characteristics like cutaneous pigmentation and papillomas, a few more skin defects and atrophy. Hypophosphatemia is also an *X*-linked dominant disease.

Characteristics of *X*-linked Dominant Inheritance

1. The affected male transmits the trait to all his daughters but not to the sons.
2. When affected females are homozygous, they transmit the trait to *all* their children of both sexes.
3. When affected females are heterozygous, only 50% of their children of both sexes have a chance of being affected.
4. Affected females transmit the trait to their progeny in a manner similar to that in autosomal dominant inheritance.

Polygenic Disorders and Multifactorial Inheritance

Some normal traits like height and intelligence, and disorders like cleft lip/palate, club foot, some allergies, diabetes mellitus, hydrocephalus, pyloric stenosis and others are inherited through polygenes and may be influenced by extraneous factors including drugs. The polygenes have small additive effects. The clinical features are due to cumulative effects of all the polygenes as well as other factors. For this reason the term multifactorial inheritance is preferred. In most cases the exact number of genes involved is not known.

The congenital conditions of cleft lip (CL) as well as of cleft lip with cleft palate (CLP) are found to be associated with a large number of syndromes (listed in Nora and Fraser, 1974). Some of these are related to chromosomal aberrations, a few are caused by mutant genes, the rest appear to be multifactorially determined. An individual may have a cleft lip (Fig. 21.4) due to defective closure of the primary palate in embryological development, or a cleft palate, due to faulty closure of the secondary palate; or both cleft lip and cleft palate (CLP) may be present in the same person.

(A) (B)

Fig. 21.4 Cleft lip (*A*) and cleft palate (*B*) in two human subjects (courtesy Dr. M.S. Grewal).

Studies on families have shown that cleft lip ± cleft palate are frequently associated with syndromes which are inherited as autosomal dominant, autosomal recessive, or *X*-linked conditions. It has also been revealed that some genes will cause cleft lip and cleft palate in some individuals, while in some other individuals they cause cleft palate. The combined influence of genetic and environmental factors have established multifactorial inheritance for CLP. More males than females are affected by CLP. The recurrence risk is higher in sibs of female probands. The condition is more common among orientals. Rarely cleft palate occurs in absence of cleft lip. This is more common in females than males, and in children of older mothers.

GENETICS OF HUMAN METABOLIC DISEASES

The biosynthetic pathways involve a number of enzymes and reactions each controlled by a structural gene. A mutation in a single gene can block a metabolic pathway and produce a heritable genetic disorder.

A. Defects in Amino Acid Metabolism

1. *Albinism:* The condition is due to lack of melanin pigments caused by the absence of tyrosinase enzyme. There are two types of albinism, one in which *only* the eye pigment is absent, called ocular albinism. The second where lack of pigmentation occurs in *both* skin and eyes. This is of two further types namely tyrosinase positive oculocutaneous albinism and tyrosinase negative oculocutaneous albinism.

The enzyme tyrosinase is normally present within structures called melanosomes which occur in melanocytes. In individuals having the autosomal recessive mutant gene for albinism, there is no tyrosinase in the melanosomes and the pathway for production of melanin is blocked (Fig. 21.5).

Tyrosine ⟶ DOPA ⟶ Dopaquinone ⟶ Various melanin pigments (black, brown, red, yellow)

Fig. 21.5 Pathway for production of melanin.

The tyrosinase negative persons show a more extreme form of albinism, are highly sensitive to sun's rays and may develop skin cancers. The tyrosinase positive patients show a milder form of albinism and may develop some pigment. There is a separate gene for each of the tyrosinase positive and negative forms. Because of two distinct loci, a marriage between a tyrosinase positive and a tyrosinase negative individual produces normal children.

2. *Alkaptonuria:* This was one of the first metabolic diseases described by Garrod in 1908. It was known to him that this was a recessive disorder occurring more frequently in offsprings of consanguineous matings. It was first noticed that if the diapers of some new-born babies were left exposed to air, the urine turned black. The affected children were found to excrete a substance called homogentisic acid (alkapton) which becomes oxidized in air to a dark coloration (Fig. 21.6). The enzyme homogentisic acid oxidase was lacking which normally causes breakdown of homogentisic acid.

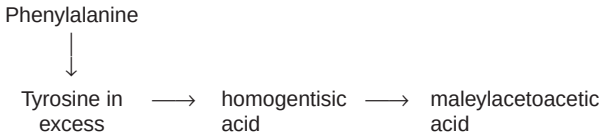


Fig. 21.6 Breakdown of phenylalanine in normal humans.

Children with alkaptonuria appear healthy. The disorder is relatively harmless except that the slow deposition of pigment in the joints sometimes leads to mild arthritis.

3. *Phenylketonuria (PKU):* Persons lacking yet another enzyme of the phenylalanine tyrosine pathway called phenylalanine hydroxylase have the condition PKU. This is one of the best understood of the inherited metabolic diseases. The hydroxylase enzyme which is present in the liver normally converts phenylalanine to tyrosine. When this enzyme is absent, there is a high level of phenylalanine in the body fluids like blood, cerebrospinal fluid and sweat. Besides tyrosine there are a few more breakdown products of phenylalanine such as phenylpyruvic acid, also formed. High levels of phenylalanine therefore produce deleterious secondary effects by accmulation of its products. Often fair skin and light hair and eyes are found in PKU patients due to lack of melanin pigment. This is a secondary effect caused by inhibition of enzymes by the abnormal by-products of phenylalanine.

There is progressive deterioration of the central nervous system starting a few months after birth. The affected babies may also show eczema on the skin, defective enamel of teeth and anomalies in bones. PKU patients are severely retarded mentally. The children are hyperactive and have highly increased muscle tone. They can be treated if placed on a diet with low levels of phenylalanine.

PKU also illustrates the phenomenon of pleiotropic (manifold) effects of a gene. A single defect in the gene which controls phenylalanine hydroxylase enzyme results in a primary block. This in turn produces manifold secondary effects.

B. Defects in Lipid Metabolism

1. *Tay Sachs Disease (TSD):* This is an autosomal recessive disease. Homozygous children show degeneration of the central nervous system due to accumulation of a fatty substance (sphingolipid) in nerve cells. This is caused by the enzyme hexosaminidase which in normal individuals exists in two forms *A* and *B*. In TSD only the *A* form is present, the *B* form is lacking.

The symptoms appear in infants in the first year after birth. They start showing paralysis, mental retardation, and other defects associated with degeneration of the neuromuscular system. After one year's age the child's condition deteriorates, there is general paralysis, loss of sight and hearing, and difficulties in feeding. By two years the child becomes immobile.

2. *Gaucher's Disease:* In this condition the breakdown of fatty substances is impaired leading to accumulation of lipid materials in body tissues and blood. It is caused by a recessive gene which inhibits the activity of an enzyme glucocerebrosidase. Consequently there is accumulation of cerebroside (a sphingolipid) in cells of the reticuloendothelial system. There is enlargement of the spleen and liver, and expansion of some of the limb bones. Of the three clinically distinct forms of this disorder, one called the early infantile form leads to death before two years age. Another form appears after the second year, while the third is present in adults in a chronic form.

3. *Mucopolysaccharidoses:* This includes a variety of conditions all related to defective breakdown of mucopolysaccharides in body tissues. A number of intermediate products of the degradative pathway thus accumulate in the lysozomes. A high level of acid mucopolysaccharides such as hyaluronic acid, dermatan sulphate and heparin are excreted in the urine. Two of the better known examples are Hurler's syndrome, caused by an autosomal recessive gene, and Hunter's syndrome, an *X*-linked recessive disorder. In both there is accumulation of dermatan sulphate and heparin sulphate in the tissues; symptoms include mental retardation and defects in the bone and facial appearance.

C. Defects in Sugar Metabolism

G6PD: Here the genetic condition is related to the drugs administered to the individual. The disorder has been studied extensively. It also has a historical background. During World War II some of the U.S. servicemen were found to be sensitive to the antimalarial primaquine and responded with hemolytic anemia. The blood counts became low and the patients became jaundiced. The haemolytic reactions lasted for a few days, then there was spontaneous recovery.

It was found out that the condition was related to low levels of reduced glutathione in the RBCs. The reduced coenzyme NADPH was also deficient in these cells. The appropriate levels of reduced glutathione in red cells are maintained by the enzyme glutathione reductase and its coenzyme NADPH. For adequate supplies of NADPH the enzyme G6PD (glucose 6-phosphate dehydrogenase) is required. Consequently, when activity of G6PD is low it leads to primaquine sensitivity.

There are many different variants of the enzyme G6PD. Some of the variants are normal, while others have low physiological activity so that they are not able to produce enough NADPH. This finally results in lower levels of reduced glutathione in the red cells. Two types of mutations have been recognised for causing haemolytic anaemias. In the African and Black American populations the enzyme G6PD A is synthesized. In Asian and Mediterranean subjects, the G6PD Mediterranean variant is prevalent. Since G6PD is an *X*-linked recessive trait, the affected persons are predominantly males. In females, due to inactivation of one of the two X chromosomes in each cell, there are two populations of cells, G6PD deficient and normal. There is also an indication that persons with an abnormal G6PD allele are less likely to have malaria. It is noteworthy that G6PD deficient persons appear healthy. It is only when they take certain foods or drugs like primaquine or chloroquine, sulpha and some others, that they respond with haemolytic reaction.

D. Defects in Purine Metabolism

Lesch-Nyhan Syndrome: This is an *X*-linked recessive disorder. It is caused by the absence of an enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT) which is required for the utilisation of purines in the synthesis of DNA and RNA. In absence of the enzyme, purines

accumulate and are converted into uric acid which passes into the blood stream and urine. Affected children (usually males) thus show high levels of uric acid in blood and urine and are detected due to the presence of 'orange sand' in the diapers of infants. They become mentally retarded and by six months the muscles start becoming weak and flabby. By about one year again the muscles develop excess tone and the children have spasms of arms and legs. Affected children about 2-3 years old have a tendency for self mutilation. They start biting their fingers, lips and inside of mouth. They also show aggressive behaviour towards others. Some teen-aged individuals who are affected develop gouty arthritis.

E. Defects in Metabolism of Sulphur-Containing Amino Acids

1. *Homocysteinuria:* This is an autosomal recessive disorder in which homocysteine and serine are not converted into cystathionine due to a defect in the apoenzyme cystathionine synthetase. During methylation of DNA (described in Chapter 16) or of RNA, the methionine is derived from S-adenosyl methionine. After giving away methionine, S-adenosyl methionine becomes converted into S-adenosyl homocysteine. The resulting homocysteine condenses with serine to form cystathionine, catalysed by the enzyme cystathionine synthetase which requires pyridoxal phosphate as the coenzyme. The enzyme action takes place in the liver. Homozygous persons affected with homocysteinuria have a structural change in the apoenzyme, so that it cannot bind to the coenzyme pyridoxal phosphate with the same efficiency as the normal apoenzyme.

The affected individuals have convulsions, mental retardation and abnormalities in the lens of the eye and bones (osteoporosis). The condition can be treated by administering large amounts of vitamin B6 (pyridoxin) in the diet, and by restricting intake of methionine and cystine.

2. *Cystathioninuria:* In this rare condition, there is deficiency of the enzyme cystathionase which is required for breaking down cystathionine into homoserine and cysteine. Consequently cystathionine accumulates in the plasma and is excreted. Various external symptoms are associated with this disorder. Treatment with pyridoxin improves the condition.

SOME COMPLEX TRAITS IN FAMILIES

Diabetes Mellitus

This is a complex, heterogeneous condition which is frequently seen to recur in certain families. It is due to partial or complete lack of insulin which leads to impaired carbohydrate metabolism, as well as abnormalities in lipid and protein metabolism. Insulin is required for utilisation of glucose and for maintaining low levels of glucose in the blood. In diabetes mellitus, the formation of glucose exceeds its utilisation giving rise to increased concentration of glucose in blood and tissues. Excess glucose appears to reach the kidneys where it forms glycoproteins resulting in proteinurea and progressive loss of renal function.

There are two forms, *juvenile* and *late onset diabetes*. The juvenile form affects young children and has a high recurrence risk among the young children in the family. In the late onset form people in the older age groups become diabetic. The other adults in the family have a lesser chance of being affected. The recurrence risk of diabetes also varies with the diet and environment. The juvenile form requires insulin intake for survival, while late onset diabetes can be managed by regulating diet and with drugs. The exact cause of diabetes is not known. External agents like infection with *Coxsackie* virus are perhaps involved with juvenile diabetes. Some other studies point toward the existence of a recessive gene.

Muscular Dystrophies

There are a group of disorders characterised in general by progressive deterioration in the structure and function of various groups of striated muscles. There are several types of muscular dystrophies of which only a few are hereditary. The Duchenne type is *X*-linked recessive. It appears in early childhood and the pelvic girdle muscles are first to be involved. The limb-girdle type is autosomal recessive and becomes manifest after puberty in the second or even third decade of life. The fascio-scapulo-humoral type is autosomal dominant. Being mild, this type does not affect longevity. It involves the muscles of the face and the pectoral girdle. The myotonic type is also autosomal dominant but with irregular expression. There is degeneration of the muscles of the tongue and forearm, as well as atrophy of the fascial muscles. Some more clinical features such as cataract, frontal baldness and cardiac involvement are associated with this condition.

GENETICS OF THE HAEMOGLOBINS

Normal Haemoglobin

Haemoglobin consists of 4 polypeptide (globin) chains each associated with a single haem group. There are 5–6 types of haemoglobins normally present at different stages of life, from zygote to adulthood. The embryonal haemoglobins (HbE) are heterogeneous and designated Gower-1, Gower-2 and Portland. In the fetal stages fetal haemoglobin (HbF) is present. After birth and in adult life the amount of HbF falls to less than 1 %. In adult stages 90% of the haemoglobin consists of HbA and about 2 % HbA2.

Out of the four globin chains constituting haemoglobin, two known as alpha (α) chains consist of 141 amino acids each and are identical in all types of haemoglobin. The other two chains are different in each type of haemoglobin. Thus HbE with two epsilon chains is written as $\alpha 2\varepsilon 2$, HbF with two gamma chains as $\alpha 2\gamma 2$, HbA as $\alpha 2\beta 2$ and HbA2 as $\alpha 2\delta 2$. The β chains each have 146 amino acids but they are structurally different. Each polypeptide chain is coded for by a different structural gene. The gene loci for the two α chains are on chromosome 16, and for β, γ, δ and ε chains on chromosome 11. The genes for the globin chains β and δ in adult humans contain intervening sequences (Chapter 22).

The Abnormal Haemoglobins

The defects in the globin chains can be divided into the following groups: (*a*) the thalassaemias caused by reduction in globin chain synthesis; (*b*) sickle anaemia and other structural variants in globin chains; (*c*) hereditary persistence of fetal haemoglobin (HPFH).

(*a*) *The Thalassaemias:* These are characterised by reduced synthesis of either the α or β chains, likewise designated as α- or β-thalassaemias. The first genetic disorder of this kind was described by Cooley in β-thalassaemia (also called Cooley's anemia or thalassaemia major). The reduced synthesis of β chains leads to accumulation of α chains which cause damage to the precursors of red blood cells in the bone marrow. Persons homozygous for the β thalassaemia gene suffer from severe haemolytic anaemia and usually die before puberty. Heterozygous persons are also not normal, but show the defect in a less severe form (thalassaemia minor).

Thalassaemia is of three types depending upon whether there is reduced or absent synthesis of one or more globin chains. Accordingly, there is α-thalassaemia, β-thalassaemia, and δ-β thalassaemia.

In complete absence of synthesis of a particular globin chain, the thalassaemias are referred to as α^0, β^0, and δ-β^0, respectively. When there is only reduction in globin chain synthesis, the conditions are designated α^+, β^+ and δ-β^+ thalassaemia respectively.

Thalassaemia is a heterogeneous disorder and patients present with a variety of clinical features. The disease is usually inherited as a recessive trait. Clinical presentations range from asymptomatic hypochromic microcytosis to severe anaemia which could be fatal in utero or early childhood. The heterogeneity may be attributed to several interrelated factors which modulate the globin genes, such as elevated synthesis of fetal globin subunits, and inheritance of other structural haemoglobin variants like HbS, HbD, HbE and others. Some individuals may inherit more than one defective globin chain gene. Thalassaemia seems to result from the collective consequences of inadequate haemoglobin concentration and accumulation of unbalanced globin subunits, which leads to haemolytic anaemia and inefficient erythropoiesis.

α-*Thalassaemia*

Among the 4 α loci in humans, one or more loci may be non-functional. Thus there are four kinds of thalassaemias: (1) α-thal-2 trait; only one of the four A-globin gene loci is non-functional. The individual may not express the trait but is a silent carrier.

(2) α-thal-1 trait; two of the four A-globin gene loci are non-functional.

(3) HbH disease in which three loci do not function.

(4) Hydrops faetalis with Hb Barts in which all four loci are defective. This condition is not compatible with life.

Most of these conditions are due to deletions of the α-globin genes; some non-deletional forms of α-thalassaemia have also been described. Haemoglobins that are structurally abnormal have been associated with α-thalassaemia.

β-*Thalassaemia*

Mutation in the single β -globin gene on chromosome 11 results in reduced synthesis of β-globin chains, giving rise to β-thalassaemia. More than 100 point mutations have been detected in the β-globin gene. The condition in homozygous patients is called Cooley's anaemia or thalassaemia major, is severe and the patient is dependent on transfusion. The nature of the mutation determines severity of homozygotes, so that β^0 patients (complete absence of β chain synthesis) are severely affected as compared to β^+ patients (reduced synthesis of β chain).

The various forms of β-thalassaemias have been correlated with several different mutations that could affect multiple steps in the pathway of globin gene expression, namely, transcription, processing of the precursor of mRNA, translation, and post-translational stability of the β-globin chain. Among these, 12 transcription mutations, some RNA modification mutants, as well as some RNA cleavage and polyadenylation (mutations in poly A tail) mutants have been described which produce a mild form of β^+-thalassaemia. The RNA splicing mutants result in blocking functional activity of mRNA, and as expected there is no chain synthesis, leading to β^0-thalassaemia. The same happens in the case of translation mutants for which 37 mutations have been found resulting in non-functional mRNA and absence of polypeptide chains. The mis-sense mutants produce highly unstable β-globin chains which are degraded rapidly after synthesis.

(*b*) *Sickle Haemoglobin:* Structural variations in α or β globin chains are due to amino acid replacements. These are caused by single base substitutions in the structural genes. In 1949 Linus Pauling and his colleagues discovered that the electrophoretic mobility of sickle haemoglobin

is different from that of normal haemoglobin. In 1957 Ingram demonstrated by the technique of "fingerprinting" that sickle hemoglobin and normal haemoglobin differed by a single amino acid substitution at position 6 from the N terminus in the β chain.

The word sickling is derived from the sickle shaped red blood cells present in this condition (Fig. 15.21). In homozygotes under low oxygen supplies most of the erythrocytes lose their normal disc shape and become crescent-shaped or sickled. The sickled cells increase the viscosity of blood. Their presence in smaller vessels leads to blockage of capillaries and tissue damage. The cells are fragile when exposed to mechanical trauma and easily destroyed in the blood vessels or spleen. This leads to the severe condition of sickle cell anemia. In heterozygotes roughly half the erythrocytes are sickle-shaped and half normal, and the individual is only mildly affected or normal.

Besides sickle haemoglobin a number of variants of haemoglobin have been described in which there are one or more (3, 4 or 5) amino acid substitutions in the β chain. Two of the more common ones are designated HbC and HbE; others are relatively rare.

(*c*) *HPFH:* Normally the synthesis of γ chains of foetal haemoglobin is much reduced in the first few months after birth; instead β and δ chains are synthesised in large amounts so that HbA and HbA2 are produced. A number of persons are found to show hereditary persistence of foetal haemoglobin. The condition does not produce any haematological disorder.

AMNIOCENTESIS

Prenatal diagnosis by amniocentesis is useful for prevention of birth of genetically abnormal children. Families having repeated abortions and/or congenital malformations (defects from birth) are likely to be carriers of a genetic anomaly ("high risk" families). The genetic counsellor can estimate the probability or chance for a future child in such a family to have a genetic defect. But probabilities are not certainties, and after conception the anxiety of the family grows regarding the normalcy of the foetus. As fetal cells in amniotic fluid of pregnant mothers originate from foetal membranes, skin, respiratory and digestive tracts, their study provides definite information on whether the foetus is actually affected or not. If positive results are obtained by amniocentesis, then termination of pregnancy can be suggested. As there is no cure for most genetic defects after the child is born, the importance of this technique is obvious.

Sometimes families are encountered that have children of only one sex, either all boys or girls. Such parents are desirous of having a child of the other sex. Amniocentesis can determine the sex of the unborn child so that the parents can decide whether or not to continue the pregnancy. In practice however, amniocentesis should not be encouraged for choosing the sex of the child.

Prenatal diagnosis by amniocentesis is advisable when parents are seeking genetic counselling in any of the following situations.

1. Previous birth of a child with trisomy or increased maternal age at the time of conception. When there is a trisomic child in the family, there is 1–2 % recurrence risk of another child being born with trisomy. The incidence of trisomy is also related to increased maternal age (above 37 years). It has been demonstrated that increased paternal age (above 55 years) also increases risk for trisomy (Stene, 1977).

2. If one of the parents is carrying a balanced chromosomal translocation, there is greater risk for producing a chromosomally abnormal fetus. The carrier parent produces unbalanced gametes that could lead to infertility, abortions and abnormal offspring. About 8–10% of couples who have had multiple recurrent abortions are found to be carriers of balanced rearrangements (Mennuti *et al.,* 1978). It is more serious if the chromosomal rearrangement is caused by a small fragment, because smaller degrees of imbalance are more likely to result in viable offspring.

3. When the woman is a carrier of a deleterious *X*-linked gene, there is a 50% chance that a male foetus would be affected. In such cases amniocentesis is performed to determine foetal sex. If it is a male foetus, abortion can be considered. A few of the *X*-linked disorders affecting males can be diagnosed through enzyme assays, for example Lesch-Nyhan syndrome, Fabry disease and Hunter's syndrome. In others such as haemophilia and muscular dystrophy (Duchenne type), diagnosis cannot be done antenatally, and, even if the foetus is male, it has equal chances for being normal or affected.

4. When parents are heterozygous for a recessive metabolic defect which can be detected from cultured amniotic cells. More than 40 disorders have been identified so far. The culture technique usually takes 4–8 weeks to grow sufficient cells for enzyme analysis. This is a long time for the family to wait for the result. Therefore sensitive micromethods for enzyme assay using microspectrophotometry and microfluorometry have been applied to prenatal diagnosis (Galjaard *et al.,* 1974; Hösli, 1974). These methods can detect Fabry disease, gangliosidosis and Pompe disease within 2 weeks after amniocentesis.

5. When there has been a previous birth of a child with a neural tube defect such as spina bifida or anencephaly. The levels of alpha-1-fetoprotein in amniotic fluid and maternal serum are higher when the fetus has a neural tube defect as compared to normal pregnancies. If a parent has a neural tube defect there is 2–5% risk of having a child with a similar defect.

6. When parents have had a consanguineous marriage and a recessive disorder is known to have occurred in a relative.

7. When a parent carries a deleterious gene and is also heterozygous for a closelylinked marker. The use of marker genes whose products can be detected in the amniotic fluid, when these genes are linked to genetic disorders has been suggested by McKusick and Ruddle (1977). In such cases the transmission of a genetic disorder to the fetus is determined indirectly through the transmission of the marker gene. This has been done for myotonic dystrophy (Insley *et al.,* 1976). The gene for this disorder is linked to the secretor gene which controls the secretion of the blood group substance.

8. When a parent is carrying a mutation for sickle cell anaemia or thalassaemia, it is possible to identify the defect at the level of the genes by the recent DNA sequence and molecular hybridisation techniques. The details are described later.

The procedure for transabdominal amniocentesis is as follows. In the 16th to 20th week of pregnancy about 5–10 ml of amniotic fluid is drawn by puncturing the uterus with a hypodermic needle. The foetal cells in amniotic fluid are analysed cytogenetically or biochemically as necessary.

For determination of foetal sex, cell smears are made directly. The slides are stained with Giemsa for Barr bodies, or with quinacrine for fluorescent Y bodies. For confirmation of foetal sex however, karyotype analysis is required. For this, amniotic cells are cultured for 10–15 days and

metaphase spreads prepared. Karyotypes are made from a number of different clones for accurate diagnosis. For correct identification of small deletions and translocations involving small fragments, chromosomes are stained by the banding technique. The sex chromosomes determine foetal sex.

Prenatal Diagnosis by DNA Sequencing Techniques

In the present era of DNA sequencing techniques, prenatal diagnosis has become possible at the level of the gene. The recent approaches make use of restriction enzymes which cut DNA at specific short nucleotide sequences. Combined with gel electrophoresis for separating DNA fragments on the basis of molecular weight, the technique is useful for characterising specific DNA fragments from nucleated foetal cells. The fragments are identified by preparing probes of complementary DNA or RNA and hybridising the probe with the DNA fragment. Since the sequence of the probe is known, that of the DNA fragment is easily determined. It should be noted however, that probes are not available for all known genes. So far probes for human globin genes only are obtainable. Such probes can be useful for prenatal diagnosis only if the fetus carries a risk for an alteration in the sequence of a globin gene.

How exactly restriction enzymes are utilised for prenatal diagnosis can now be explained. The presence of alterations in the DNA sequence of a mutant gene in the foetus changes the pattern of restriction endonuclease recognition sites. For example the restriction endonuclease Mnl 1 normally recognises the sequence GAGG. In a fetus carrying a mutation for sickle cell haemoglobin the sequence 5′-GAGG-3′ in the β globin gene changes to 5′-GTGG-3′ which the endonuclease Mnl 1 is not able to recognise. Such a fragment is detected because it is deleted. However, due to the small size of fragments generated by Mnl 1, there are technical difficulties in antenatal diagnosis of sickle cell anemia by this method.

An alternative approach for prenatal detection of sickle cell anaemia has been developed by Kan and Dozy (1978). The method is based on detecting a change in the DNA sequence within a region flanking (lying adjacent to) the β globin gene. This became possible when it was found that the sickle cell mutation was frequently linked to an altered DNA sequence in the flanking gene.

The endonuclease Hpa I recognises the normal sequence in the flanking region which is located 5,000 bases downstream from the β globin gene. When this sequence is altered, as in the case of homozygous sickle anaemia, Hpa I is not able to recognise the site and a fragment is deleted. In man the globin genes contain repeated DNA sequences. There are two α globin genes on chromosome 16 (Orkin *et al.,* 1978), while the genes for β, δ and γ chains are on chromosome 11. It is known that crossovers occur within human globin genes so that some sequences within or between the globin genes become deleted. Such deletions have been found to produce clinical disorders such as haemoglobin H disease, hydrops foetalis, the thalassaemias and HPFH. In the case of α-thalassaemia, the α chains of haemoglobin are not synthesised due to deletion of all the four α chain genes from DNA (Wong *et al.,* 1978). Using complementary RNA probes and hybridisation technique on cultured amniotic cells, Kan and Dozy were able to detect α-thalassaemia–1 where 2 genes are deleted and haemoglobin H disease having 3 deleted genes.

SOMATIC CELL HYBRIDS

It had been known since the 60's that somatic cells from the same or different species in culture could spontaneously fuse to form polyploid cells. The product of fusion was called *homokaryon* if the two parental cells came from the same species, and heterokaryon or *somatic cell hybrid* if the

fusion was interspecific. The hybrid cells could divide by mitosis and proliferate and thus could be maintained in culture (Fig. 21.7).

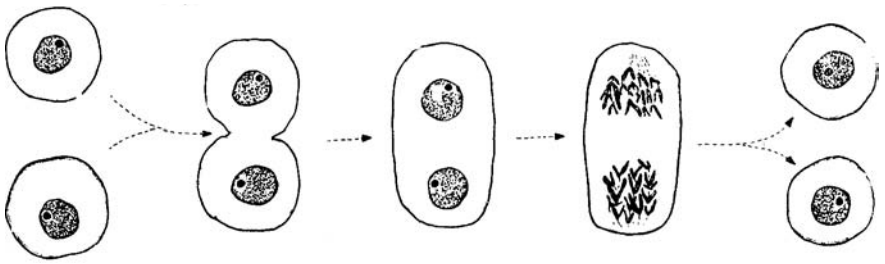


Fig. 21.7 Mechanism of formation of somatic cell hybrids.

Two further technical advances made human gene mapping by somatic cell hybrids possible. In 1962 Okada discovered that inactivated Sendai virus could greatly increase the rate of cell fusions. Since then several agents causing cell fusion have been tried among which polyethylene glycol has some advantages. The exact mechanism of cell fusion is not known. In the case of UV-inactivated Sendai virus, it seems that the virus absorbs to the cell surface leading to agglutination of cells. The protein coat of the virus forms the connecting bridge between the cells. The membranes of the two cells swell into this region and when they come in contact are dissolved. The cell contents mix up, the nuclei fuse and a heterokaryon is formed. When cell fusion is mediated by polyethylene glycol, the two cell membranes directly come in contact.

The second technical advancement was the finding that when sublines of hybrid cells are maintained in culture, there is gradual and preferential loss or retention of specific chromosomes (Weiss and Green, 1967). The association between the retention of a genetic marker and that of a specific human chromosome could be determined. In mouse-man hybrids, most of the chromosomes of the mouse are retained. By using a selective medium which allows growth of cells having a particular chromosome, it is possible to locate genes on a specific chromosome. This technique has been extensively applied for human gene mapping.

The following example illustrates the technique used by Littlefield (1964) for assigning genes to specific chromosomes. He used deletions and mutations affecting enzymes involved in the purine and pyrimidine salvage pathways, namely thymidine kinase (TK) and hypoxanthine guanine phosphoribosyl transferase (HGPRT). There are two pathways for DNA synthesis. In the first, which proceeds under normal conditions, DNA is synthesised from simple organic molecules and the necessary enzymes. The second is the alternate or salvage pathway which utilises nucleotide precursors for DNA synthesis. The salvage pathway is followed only if the first pathway is blocked by an antimetabolite (for example aminopterin) or by a mutation. Two enzymes are necessary for the salvage pathway, HGPRT and TK. If even one of the two enzymes is absent, DNA synthesis cannot take place by the salvage pathway.

Cells from a mutant mouse cell line deficient in the enzyme TK (*i.e.* TK^- /$HGPRT^+$) were mixed with cells from a human line deficient in HGPRT (*i.e.,* TK^+ /$HGPRT^-$) and allowed to grow on minimal medium. Under appropriate conditions the cells fused to form hybrid cells KT^+/TK^-; $HGPRT^+/HGPRT^-$). It should be noted that in the hybrid cells there is one normal allele for the enzyme TK (from the human cell line) and one normal allele for HGPRT (from the mouse cell line).

All the cell lines could grow on a minimal medium. The mouse cell line (TK^-) is not able to grow on a medium containing thymidine as the cells are deficient in TK. Similarly the human cell line ($HGPRT^-$) cannot grow on a medium containing hypoxanthine due to lack of the enzyme

HGPRT. It is also noteworthy that neither of the mutant cell lines is able to grow if aminopterin, an antimetabolite is present in the medium. Aminopterin acts by inhibiting the enzyme folic acid reductase which catalyses the synthesis of reduced folate. The latter is required in the various steps of the normal pathway leading to the synthesis of DNA. In the presence of aminopterin therefore, DNA is synthesised through the salvage pathway, but only if the enzymes TK and HGPRT are available. The mouse-man hybrid cells are thus able to grow in presence of aminopterin if thymidine and hypoxanthine are present in the medium.

There is a selective medium which allows growth of the hybrid cells but inhibits the parental cells. This medium contains hypoxanthine, aminopterin and thymidine and is called HAT medium. The hybrid cells will proliferate on HAT medium to form colonies as they alone have genes for both TK and HGPRT. Colonies of hybrid cells can be subcultured and cloned for mapping genes. Sublines which show progressive loss of human chromosomes are maintained. Only those cells that retain the specific chromosome having the gene for thymidine kinase would survive in HAT medium. It was found out that only cells retaining chromosome 17 could grow on HAT medium. Obviously the gene for thymidine kinase is located on chromosome 17.

The mouse-man cell hybrids have the following advantages for studies on gene mapping.

1. Preferential loss of human chromosomes.
2. Availability of cell lines with identifiable human phenotypes different from those in rodents.
3. The apparent distinction between rodent and human chromosomes in interspecific hybrids.
4. That both rodent and human genes are simultaneously expressed in the cell hybrids and the product proteins of each can be identified individually.
5. The linkage groups on mouse chromosomes are known.

IMMUNOGENETICS

Genetic mechanisms which control immune responses have opened up an entire new field of immunogenetics. Basically it is the study of antigens, antibodies and their reactions. An antigen is a substance present in the body or introduced from outside which can initiate an immune reaction. The antigens present on the surface of red and white blood cells are important in immunogenetics. The immune reactions in response to the antigen take place in the lymphoid organs (spleen, lymph nodes and tonsils) and result in the production of antibodies or sensitised lymphocytes which are effective in eliminating the antigen from the body. As there are millions of potential antigens, there are many millions of species of antibody molecules that are synthesised by the immune system.

Antibodies are serum proteins belonging to the group immunoglobulins (Ig). Each antibody is made up of two kinds of protein chains designated light and heavy (L and H) chains. The heavy chains are of 5 types: gamma, alpha, mu, delta and epsilon. There are thus five classes of immunoglobulins named according to the type of heavy chain present, that is IgG, IgA, IgM, IgD and IgE; of these IgG is most abundant in the serum. The structure and properties of the different immunoglobulins are genetically controlled. Basically IgG is made up of two light (L) and two heavy (H) chains linked by disulphide bonds. The L chains have 214 and H chains 440 amino acids. Each of the light and heavy chains consists of a constant (C) half and a variable (V) half (Fig. 21.8).

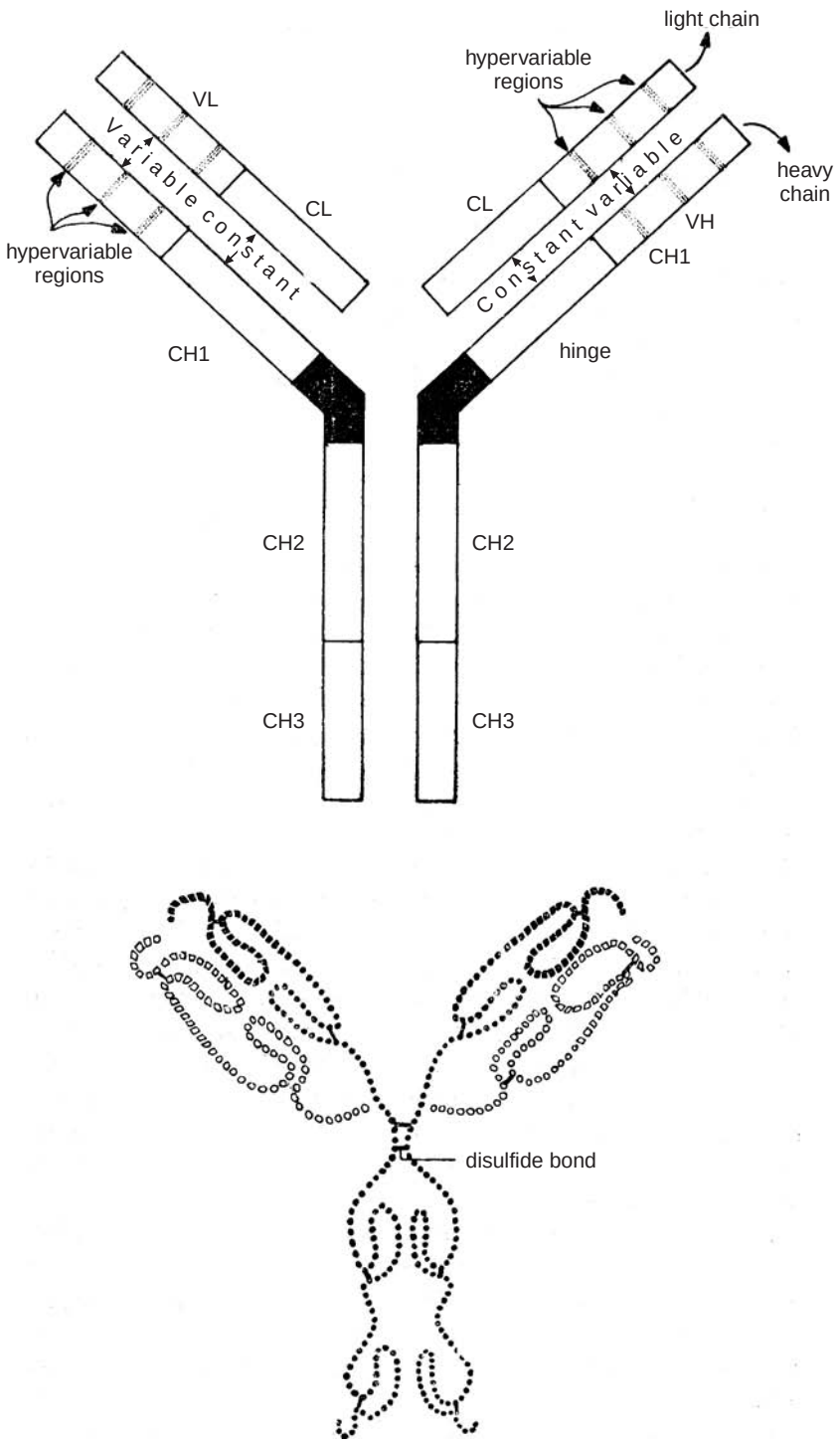


Fig. 21.8 Outline diagrams showing the general structure of immunoglobulin (IgG).

The ability of an organism to produce an enormous number of antibodies lies in the amino acid sequences of the V regions of the light and heavy chains. Thus there are a large number of immunoglobulin chains in which the V region is different but the C region is the same. The antigen binds to the antibody in the V regions of the light and heavy chains; the combining site determines the specificity for the antigen. The antibody molecule appears V-shaped in the electron microscope. The two antigen-binding sites containing identical amino acid sequences are located at the ends of the two arms. Thus two antigens of the same type can link to a single antibody molecule.

Two questions have been raised, firstly, what genetic mechanisms are responsible for the immense variability of the V region? Secondly are genes coding for the immunoglobulin chains present in the gamete or are produced during somatic differentiation? Proponents of the *germ line theory* believe that all the immunoglobulin genes are transmitted through the gametes. There would thus be many *different* V genes and a large number of copies of one or a few C genes. According to the *somatic differentiation theory* the variety of V genes is created during lymphocyte differentiation by somatic mutation or by somatic recombination. The current views combine both the above theories and envision many V genes containing coding sequences and spacers in the germ cells, and a small number of C genes. The innumerable variants of the V genes arise during somatic recombination in the cells of the immune system called B lymphocytes.

There are two types of light chains in the mouse as well as human beings, one called kappa the other lambda, each with a characteristic amino acid sequence. In mouse the variable regions of both kappa and lambda chains are coded by a large number of V genes which have been identified in the DNA of embryonal cells. There may be several hundred V genes, each gene consisting of two coding regions separated by a silent spacer. The first coding segment is translated into a leader sequence about 20 amino acids long and is thought to be involved in the movement of the antibody across the cell membrane. It is eventually cut off before the antibody molecule is transported. The second coding region is specific for about 95 amino acids out of the 108 constituting the V region. The remaining 13 amino acids, of the V region are specified by a J gene situated much downstream, that is toward the 3′ end near the gene for the C region. There is a long non-coded spacer between the V and J genes. When the light chains are of the lambda type in mouse, there are 4 genes for the C region each having its own J gene. The V and C sequences join at the J region to produce the immunoglobulin gene.

THE HLA SYSTEM

The genetic mechanisms controlling transplantation antigens play a significant role in the organ transplant technique. The treatment of certain human diseases requires transfer of grafts from a host to a recipient individual of a different genetic make up (allograft). The existence of transplantation antigens first became known from skin grafting experiments in mice. Whether a graft would be rejected or accepted by the recipient mouse depended upon the presence of histocompatibility antigens present on the skin and other tissue cells. From the results of genetic experiments it was concluded that the histocompatibility antigens were controlled by several different genes, each gene having multiple alleles.

That the red blood cells in humans contain blood group antigens on their surface has already been described (Chapter 2). These antigens are important in blood transfusions. The

white blood cells also carry several antigens which are important in organ transplantation. These are called histocompatibility antigens and are found to be controlled by four gene loci *A*, *B*, *C* and *D* on chromosome 6. The histocompatibility (histo meaning tissue) antigens and their corresponding genes constitute the HLA (human leukocyte antigen) system.

When a graft is transplanted, the lymph nodes of the recipient individual respond to the histocompatibility antigens of the genetically different tissue by producing lymphocytes which are of two types, the *T* lymphocytes (thymus dependent) and the *B* lymphocytes (cells equivalent to the bursa in birds). The *T* cells are important in graft rejection, and are involved in the *cell mediated immune* response (CMI). The *B* cells are involved in the humoral responses (HR) which produce antibodies against viruses, bacteria and such invasions, as well as against graft cells of the host. It is mainly the *T* cells which respond to the histocompatibility antigens present on the grafted tissue.

The genetics of HLA has been studied by performing breeding experiments in mice. When two mice homozygous for the histocompatibility antigens are crossed, the *F*1 mice are able to accept grafts from both parents. When *F*1 mice are inbred, three of the progeny mice accept grafts from either parent, while one mouse shows rejection. Applying the principles of Mendelian inheritance one can conclude that the parental mice differ at a single gene locus. From similar experiments a number of loci controlling histocompatibility antigens could be determined. Each locus has several alleles. One complex locus *H–Z* controls the strong transplantation antigens in mouse.

In humans, the strongest transplantation antigens are controlled by four distinct loci *A*, *B*, *C* and *D* on chromosome 6 (Fig. 21.9). Additional genes related to histocompatibility also lie in the adjacent regions. Studies on different individuals in the population have revealed that the *A* gene has more than 12 alleles, and *B* gene at least 20 alleles controlling more than 32 histocompatibility antigens. As the *A* and *B* genes are only one map unit apart, there is very little chance of crossing

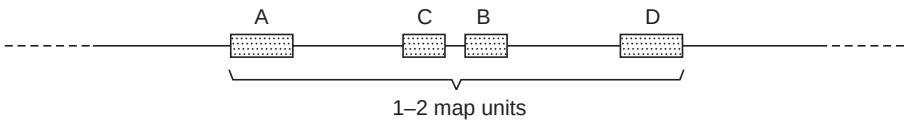


Fig. 21.9 Map positions of the transplantation antigens.

over between them; they are thus transmitted together in most cases. Since any allele at *A* can be associated with any allele at *B*, the number of possible combinations of the alleles of *A* and *B* genes in the population is about 240 (12 × 20). A combination of the alleles of the *A* and *B* genes is called the haplotype. Designating the alleles of the A locus as *A*1, *A*2, ... , *A*12, and at *B* as *B*1, *B*2, ..., *B*20, suppose one chromosome 6 of a particular individual is carrying the alleles *A*5 and *B*9. The haplotype in this case would be written as 5 9 (*A* allele is written to the left and *B* allele to the right). As there are two number 6 chromosomes in a cell, the HLA genotype of a person consists of two haplotypes. The presence of a particular allele in the individual is determined serologically from the existence of a specific antigen on the leukocyte surface by the technique of *leukocyte typing,* also called *HLA typing*. Thus individuals may be homozygous or heterozygous for any pair of alleles.

Due to the highly polymorphic and complex nature of the HLA locus, it is very unlikely that any two persons picked at random should have the same set of leukocyte antigens. These

genetic differences lead to rejection of allografts in organ transplantation. Due to this reason the host and recipient HLA genotypes have to be closely matched.

Some Other Loci Linked to HLA

The gene which is supposed to control the quantity and quality of the immune responses (*IR* gene) is said to be closely linked to the HLA genes. Some of the immunological disorders are also associated with HLA, such as psoriasis with alleles B8 and B15 of HLA; ankylosing spondylitis with B13 and B7; diabetes mellitus with an allele of the *D* gene; and multiple sclerosis with alleles of the *A*, *B* and *D* genes.

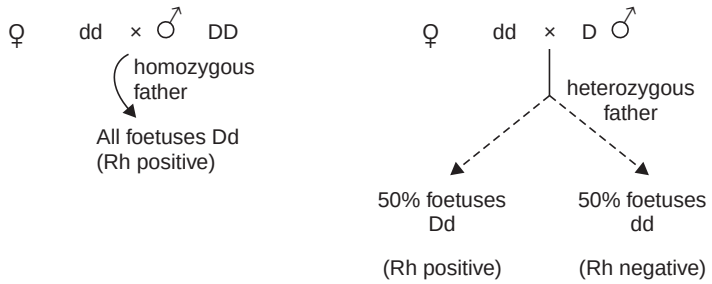
The Rh Factor

In 1940, Wiener and Landsteiner discovered that an antigenic factor Rh, which is found on the surface of red blood cells of the Indian brown monkey (*Macacus rhesus*), is also present in man. They injected the blood of rhesus monkey into rabbits and guinea pigs. Antibodies produced in the blood serum of these animals were found to agglutinate the red cells of the rhesus monkey. When Landsteiner and Wiener added the same antiserum to human blood, they found to their surprise that the red cells of about 85% of the persons tested were also agglutinated. It was thus revealed that human blood contained the same Rh antigen that was found in the rhesus monkey. In the same year, Wiener and Peters showed a similarity between the specificity of the animal antibody and an antibody found in the serum of some patients who had shown transfusion reactions. The animal and the human antibodies could both react with the same red cells. The name Rh (after rhesus) was given to the antigen and anti-Rh to the antibody.

Later studies on family pedigrees showed that Rh antigens in humans are controlled by an autosomal dominant gene *D*. Persons having one or both dominant alleles *(DD* or *Dd)* carry Rh or *D* antigens on their red cells and are said to be Rh positive (Rh^+). Persons having two recessive alleles (*dd*) of the gene lack Rh antigen and are said to be Rh negative (Rh^-). There are more than 30,000 Rh or *D* sites on the surface of each red cell.

As in the case of the ABO blood groups, Rh is also important in blood transfusions. If the blood of an Rh positive person is transfused into an Rh negative person, it will elicit the formation of anti-Rh antibodies in the Rh^- recipient. These antibodies will agglutinate the red blood cells of the transfused blood as they carry the Rh antigen on their surface. The first transfusion of Rh^+ blood into an Rh^- individual is without severe reaction. But in a subsequent transfusion with Rh^+ blood, the sensitised individual will produce increased amounts of antibodies against the Rh antigen in the transfused blood, resulting in severe reaction or even death. It should be noted however, that there are no *natural* anti-Rh antibodies in the blood of an Rh^+ or Rh^- individual. Therefore, an Rh^- person can donate blood to an Rh^+ individual without bad consequences.

Immediately after the discovery of Rh factor, an association was traced between Rh and haemolytic diseases of the newborn (*erythroblastosis fetalis*) by Levine (1942). These are conditions where newborn infants suffer from anaemia, jaundice, enlarged spleen or liver, in severe cases stillbirth or death soon after birth. This is due to Rh incompatibility between the mother and fetus. When the mother is Rh negative and carrying two recessive alleles *dd* of the Rh gene *D,* and the father is Rh positive, either homozygous *DD* or heterozygous *Dd,* then according to Mendelian inheritance the genotype of the foetus could be any one of the following:



If the foetus is *dd* (Rh^-) the pregnancy is normal. Only the Rh^+ foetus with *Dd* genotype creates problems at birth. The gene for Rh *(D* gene) is transmitted from the father to the foetus (when mother is Rh^-). If the father is homozygous for the *D* gene (*DD*), then all the conceptions will produce Rh positive foetuses; if the father is heterozygous (*Dd*), then there is 50% chance at each conception that the foetus would be Rh^+. In addition to the *D* antigen which distinguishes Rh positive and Rh negative persons, there are other antigens in the Rh system, the main ones being *C*, *E, c* and *e*.

As explained in the case of blood transfusions from Rh^+ to Rh^- persons, the first pregnancy with an Rh^+ foetus may develop low levels of antibody in the mother and remain free of complications. It is also possible for subsequent pregnancies with Rh^+ foetus to be normal. This is so because the circulatory systems of the foetus and mother are separate, eliminating any chance for the foetal Rh^+ antigen to reach the mother's blood. However, problems do arise in second or later pregnancies due to some defect or breakage in the capillaries of the placenta which allows foetal blood to leak into the maternal circulation. This usually happens about the time of birth. In such cases the Rh^+ blood of the foetus produces antibodies in the mother's blood, which may

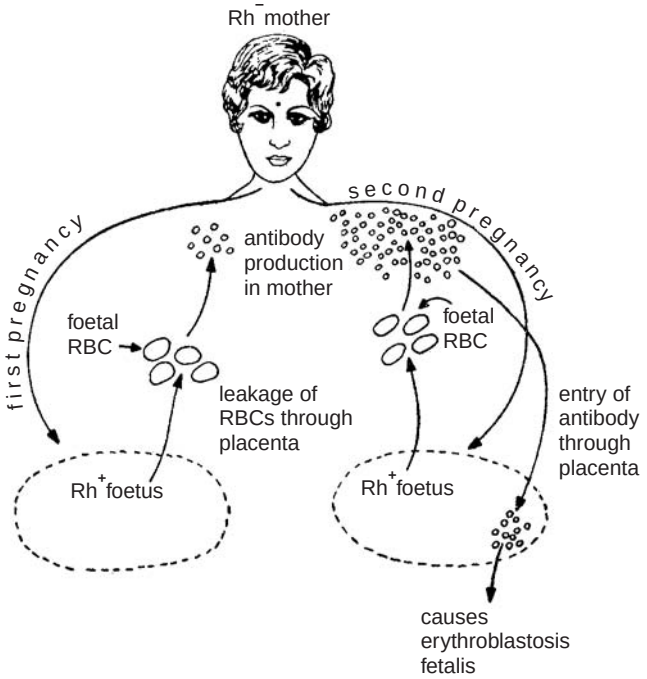


Fig. 21.10 Diagram showing the difference in the first two pregnancies of an Rh^- mother.

pass back across the placenta into the foetus. The antibodies react with foetal red cells and cause haemolysis. The reaction may be mild in the first pregnancy with low levels of antibody in the mother's blood. In subsequent pregnancies in the sensitised woman, another exposure to the Rh^+ antigen triggers an immediate response producing high levels of antibody in the blood (Fig. 21.10).

It is possible to treat the woman for prevention of Rh-caused disease in future births. Within 72 hours of the birth of every Rh^+ infant, the Rh^- mother is given an intramuscular dose of the anti-Rh antibody (Anti-D gammaglobulin), commercially available as RhoGAM. This antibody destroys the foetal red cells that have entered the mother's circulation so that they (foetal cells) cannot stimulate the production of anti-D antibodies in the mother. The process is called passive immunisation.

The Genetic Mechanism of Rh

Rh factor actually comprises a complex group of related antigens known as the Rh system. The genetic mechanisms controlling the Rh system remain confusing. Two genetic models have been proposed, one by Fisher and Race, the other by Wiener. In the *Fisher-Race Model,* inheritance of Rh is controlled by 3 closely linked genes, designated *C*, *D* and *E* which are present on an autosomal chromosome. There are two alleles of each gene: *C* and *c, D* and *d, E* and *e.* An individual could be homozygous or heterozygous for any of the 3 genes. It should be noted that the *D* gene produces the *D* antigen, but the allele *d* does not produce any antigen (silent gene or amorph). All the other alleles *C*, *c*, *E* and *e* produce their respective antigens. In the Fisher-Race system, the gene as well as the antigen it produces on the red cell are given the same symbol. If one chromosome of an individual carried the alleles *DCe,* and the homologous chromosome had *DcE* at identical loci, the antigens present on the person's RBCs would be *D, C*, *c*, *E* and *e.* Each of the antigens can be detected by testing the RBCs with antisera specific for the various antigens.

The Wiener Model: A somewhat different model has been proposed by Wiener. According to Wiener there is a single gene with multiple alleles (8 major alleles) controlling Rh. The 8 alleles are $R^0, R^1, R^2, R^z, r, r^1, r'' r^y$. Each gene produces an antigen on the red cell called an agglutinogen. Each agglutinogen can be recognised by reaction with specific antibodies.

Both the models of Fisher-Race and Wiener seem to arrive at the same conclusion. The 3 genes of Fisher-Race can produce 8 different combinations: *CDE*, *CDe*, *CdE*, *cDE*, *cdE*, *Cde*, *cDe*, *cde*. The combinations of the alleles of the 3 genes in a way correspond to the 8 alleles of Wiener. Most geneticists use the Fisher-Race system as it is more convenient.

CANCER

Cancer is fundamentally a disease of defects in the cell's regulatory mechanisms. In multicellular organisms, the regulation of proliferation, differentiation and survival of individual cells are carefully regulated. The loss of regulation in cancer cells leads to uncontrolled growth and division of cells and their spread in the body. Unlike normal cells, cancer cells fail to respond to signals that control normal cell behaviour and develop abnormalities in multiple cell regulatory systems.

Although cancer can result from abnormal proliferation and development of tumours, the key issue is, whether the tumour is benign or **malignant.** The malignant tumour is capable of invading normal tissues, spreading in the body via circulatory or lymphatic systems and establishing secondary tumours (**metastasis**). Most cancers fall into one of three main groups: **carcinomas** which arise from epithelial cells; **sarcomas** are solid tumours of connective tissues such as muscle, bone, cartilage and fibrous tissue; **leukemias** and **lymphomas** arise from blood

forming cells and from cells of the immune systems, respectively. Further classification of tumours is based on tissue of origin and type of cell involved in malignancy.

A key feature of tumours is their development from single cells that begin to proliferate abnormally, hence clones of cells are present in tumours. The development of malignancy in cell clones is a multistep process accompanied by a series of changes in the cells. In general, cancer is considered as a multistep process involving mutation and selection of cells with progressively increasing capacity for proliferation, invasion and metastasis. The first step, **tumour initiation** involves genetic alteration leading to abnormal proliferation of a single cell. **Tumour progression** continues as additional changes take place in tumour cell population**. Metastasis** occurs when tumour cells invade other organs and establish secondary sites of malignancy.

Substances that can cause cancer are called **carcinogens**, and include many agents such as radiation, chemicals, viruses and many more. Radiation and chemicals can initiate cancer by damaging DNA and inducing mutations in key target genes. Some carcinogens contribute to cancer development by stimulating cell proliferation, rather than by inducing mutations. Such compounds are designated **tumour promoters** because by inducing increased cell division they produce a proliferative cell population during early stages of tumour development. Classic examples are the phorbol esters that stimulate cell proliferation by activating protein kinase C. Hormones, particularly estrogens, are important as tumour promoters in the development of some human cancers. For example, the uterine epithelium responds to excess estrogen and increases the likelihood of development of endometrial cancer. Some viruses also induce cancer in experimental animals and humans, such as liver cancer and cervical carcinoma in humans.

Cancer cells display features that distinguish them from their normal counterparts. Cancer cells have abnormalities in the mechanisms that regulate normal cell proliferation, differentiation and survival. In culture, cancer cells can be distinguished from normal cells in displaying **density-dependent inhibition** of cell proliferation. Normal cells continue to divide until they reach a finite cell density. They then stop dividing and become quiescent, arrested in the G0 stage of the cell cycle. The proliferation of cancer cells is independent of density-dependent inhibition. Such cells do not respond to signals that cause normal cells to cease proliferation. Instead, tumour cells continue to grow to high cell densities in culture, this behaviour corresponds with their uncontrolled proliferation **in vivo**. The proliferation of many normal cells is controlled in part, by polypeptide growth factors. Cancer cells have reduced requirement for extracellular growth factors.

Both *in vitro* and *in vivo*, growth factor requirements of cancer cells are reduced contributing to unregulated proliferation of tumour cells. In some cases cancer cells produce growth factors that stimulate their own proliferation. Such an abnormal production of a growth factor leads to continuous stimulation of cell division, called **autocrine growth stimulation**, and makes cancer cells less dependent on growth factors from normal sources. In some cases, the reduced growth factor requirement of cancer cells is caused by abnormalities in intracellular signalling systems, such as unregulated activity of growth factor receptors, or proteins such as Ras proteins or protein kinases that are involved in signal transduction pathways leading to cell proliferation. Cancer cells also have reduced expression of cell surface adhesion molecules, for example E-cadherin, which is the principle adhesion molecule in epithelial cells.

Some properties of cancer cells facilitate their interactions with other tissue components, thus playing important roles in invasion and metastasis. For example, some malignant cells secrete proteases that digest extracellular matrix components, allowing cancer cells to invade nearby normal tissues. Carcinomas secrete collagenase by which they are able to break through the basal lamina and invade the connective tissue. Tumour cells also have the ability to secrete growth factors that promote the formation of new blood vessels, known as **angiogenesis**.

Angiogenesis becomes necessary for the tumour to support about a million cells, so that new blood vessels can supply oxygen and nutrients to the proliferating tumour cells. Growth factors secreted by the tumour cells induce the formation of new blood vessels. Growth factors stimulate proliferation of endothelial cells in walls of capillaries in surrounding tissue, resulting in an outgrowth of new capillaries into the tumour. These new capillaries are easily penetrated by cancer cells, which get an opportunity to enter the circulatory system and begin the metastasis process.

A very important characteristic of cancer cells is their failure to differentiate normally. In fact, defective differentiation is a fundamental issue and forms the basis for histological grading of tumours. Differentiation and cell proliferation bear an inverse relationship. A differentiated cell shows reduced cell proliferation, with cellular machinery directed toward synthesis of 'household' proteins. Actively dividing cells are not differentiated. **Programmed cell death**, or **apoptosis** is a normal self-destruction mechanism, and an integral part of the differentiation program of many normal cells. It is a normal form of cell death that eliminates damaged and potentially harmful cells. It plays a key role in maintaining adult tissues and embryonic development. In adults it maintains constant cell numbers in tissues undergoing cell turnover. Programmed cell death also provides a defense mechanism by which damaged and potentially dangerous cells, or virus-infected cells, as well as cells in which DNA is damaged or is carrying a mutation, can be removed from the system. During apoptosis, chromosomal DNA becomes fragmented as a result of cleavage between nucleosomes. The chromatin condenses and the nucleus breaks up into pieces. Eventually the cell also shrinks and breaks up into membrane enclosed fragments called apoptotic bodies. Three genes have been found to be associated with regulating and executing apoptosis. Two of these, *ced*-3 and *ced*-4 are required for apoptosis to occur; if these genes are inactivated, apoptosis does not take place. The third gene, *ced*-9 is a negative regulator of apoptosis. If *ced-9* is mutated and becomes inactive, cells that would normally survive fail to do so, and undergo apoptosis. Conversely, if *ced*-9 is over expressed, its high levels do not allow apoptosis to take place. Many cancer cells fail to undergo apoptosis, and have increased life spans compared to their normal counterparts. This failure of cancer cells to undergo apoptosis contributes to the resistance of cancer cells to irradiation and chemotherapeutic drugs, which act by damaging DNA. Therefore, both abnormal cell survival and cell proliferation play significant roles in uncontrolled growth of cancer cells.

A family of proteins called **caspases** have been found in mammalian cells that are homologous to *ced*-3. Caspases are a distinctive group of proteolytic enzymes that are activated at an early stage of apoptosis. Caspases are capable of removing small parts of other proteins, thereby inactivating or activating other proteins. Caspases could have the following targets. (*a*) Nuclear lamins, consisting of filamentous intermediate filament proteins lining the inner face of nuclear envelope. Cleavage of lamins disrupts the nuclear envelope; (*b*) Focal adhesion kinase (FAK). Inactivation of FAK disrupts cell adhesion leading to detachment of apoptotic cell from its neighbours; (*c*) Proteins maintaining cell structure. Inactivation of actin, gelsolin and intermediate filaments could lead to changes in cell shape and surface blebbing characteristic of apoptotic cells; (*d*) Endonuclease *CAD* which is activated when caspase inactivates an inhibitory protein. The activated *CAD* causes fragmentation of DNA.

ONCOGENES

Studies on tumour viruses had shown that specific genes called **oncogenes** are capable of converting normal cells into malignant cells. Further studies demonstrated that viral oncogenes

are related to genes of normal cells, led to the identification of cellular oncogenes which are involved in the development of non-virus-induced cancers. It was thought that the non-virus induced cancers might arise by mutations in normal cell genes, giving rise to oncogenes of cellular origin. But eventually it was molecular studies on retroviral (Rous sarcoma virus, RSV) oncogenes that elucidated the link between viral and **cellular oncogenes**. RSV transforms chicken embryo fibroblasts in culture into sarcomas. Deletion mutants and temperature-sensitive mutants of RSV were then isolated that could not induce transformation, but could replicate normally in infected cells, indicating that RSV contains genetic information that is required for transformation but not for virus multiplication. Further studies indicated that RSV contains a single gene by which RSV is able to induce tumors in chicken and transform fibroblasts in culture. This oncogene in RSV was called *src*; it encodes a 60 kDa protein tyrosine kinase.

All of the 40 different highly oncogenic retroviruses isolated have been found to contain one or more oncogenes. The oncogene is not required for virus multiplication but is responsible for cell transformation. More than 25 oncogenes have since been identified in retroviruses. Many of these genes encode proteins that seem to be key components of signalling pathways that stimulate cell proliferation. Retroviral oncogenes are not involved in viral replication, and seemed to be additions to the viral genome. What is the source of the retroviral oncogene? Experiments indicated that retroviral oncogenes are derived from genes of the host cell, and when a specific host cell gene becomes incorporated into the viral genome, it produces a new, highly oncogenic virus as the product of a virus-host recombination event. Thus, normal cells contain genes that are closely related to retroviral oncogenes. A complementary DNA (*c*DNA) probe for the *src* oncogene of RSV hybridised to closely related sequences in the DNA of normal chicken cells. Moreover, *src*-related sequences were found in normal cells of many vertebrates and humans. The genes in normal cells from which the retroviral oncogenes originated are called **proto-oncogenes**. They are cell regulatory genes which encode proteins that function in the signal transduction pathways controlling normal cell proliferation. Hence, oncogenes are considered to be abnormally expressed or mutated forms of corresponding proto-oncogenes, and as such induce abnormal cell proliferation and development of tumours.

Oncogenes are frequently seen to encode proteins that differ in structure and function from those encoded by their normal counterpart proto-oncogenes. Some oncogenes are expressed as fusion proteins having viral sequences at the amino terminus. Many other oncogenes differ from the corresponding proto-oncogene by point mutations, resulting in single amino acid substitutions in the oncogene products.

Oncogenes in Human Cancer

Evidence for the involvement of cellular oncogenes in human tumours first came from gene transfer experiments (Fig. 21.11). Weinberg in 1981 showed that human bladder carcinoma contained a biologically active cellular oncogene. Subsequently, active cellular oncogenes have been found in human tumours of many different types. Some oncogenes in human tumours are cellular homologs of oncogenes that were found in retroviruses, while others are new oncogenes, first detected in human cancers. It was through gene transfer experiments that the first human oncogene, *ras* gene, was identified in human cells. The *ras* gene was found to be a human homologue of the *ras*H oncogene of the Harvey sarcoma virus. Members of the *ras* gene family are encountered in many types of human malignancies.

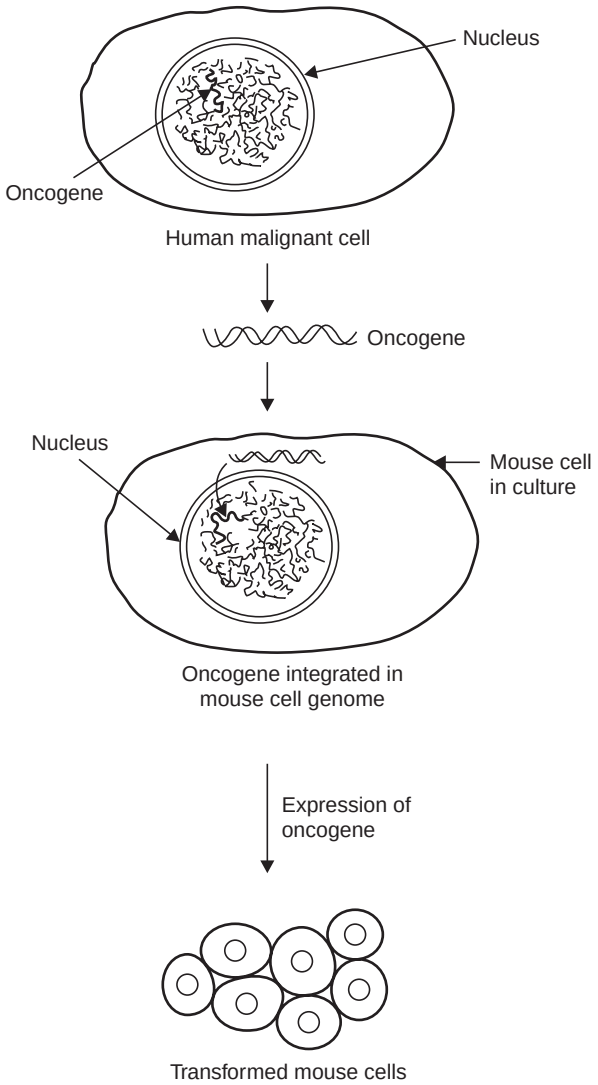


Fig. 21.11 Detection of a human tumour oncogene by gene transfer. DNA extracted from a malignant human cell induced transformation of recipient mouse cells in culture. The oncogene from human cell is integrated into mouse genome, and its expression observed in transformed mouse cells.

It is noteworthy that *ras* oncogenes are produced in tumour cells by mutations that occur during tumour development; hence *ras* oncogenes are *not* present in normal cells. The sequences of *ras* oncogenes differ from their proto-oncogenes by point mutations resulting in single amino acid substitutions at critical positions. Mutated *ras* oncogenes are present in about 20–30% of all human cancers. The action of *ras* oncogenes has clearly linked the development of human cancer to abnormalities in the signalling pathways that regulate cell proliferation. Experiments using animal models have indicated that mutations that convert *ras* proto-oncogenes to oncogenes are caused by chemical carcinogens. Mutations that convert normal *ras* genes to oncogenes substantially decrease GTP hydrolysis by *ras* proteins. The activity of the *ras* proteins is controlled by GTP or GDP binding, such that they alternate between active (GTP-bound) and inactive (GDP-bound) states. As a result of mutations, the mutated oncogenic *ras* proteins remain constitutively blocked

in the active GTP-bound conformation. Constitutive activity of the mutated *ras* proteins thus continuously stimulate the MAP kinase pathway (MAP for mitogen-activated protein kinases) and drive unregulated cell proliferation even in the absence of growth factors.

Besides point mutations, there are other ways by which proto-oncogenes are converted to oncogenes in human tumours. Many cancer cells are found to have abnormalities in chromosome structure, such as translocations, deletions and duplications. The resulting rearranged DNA sequences produce oncogenes. The first oncogene generated by chromosome translocation was *c-myc* oncogene in Burkitt's lymphomas. Here, chromosomal translocations involve genes that encode immunoglobulins. All Burkitt's lymphomas seem to have translocations of a fragment of chromosome 8 to one of the immunoglobulin gene loci residing on chromosomes 2, 14 and 22. Translocations of other proto-oncogenes could result in rearrangements of coding sequences, and abnormal gene products. Thus, translocation of the *abl* proto-oncogene from chromosome 9 to chromosome 22 occurs in chronic myelogenous leukemia (**CML**). There is fusion of *abl* with the translocation segment called *bcr* on chromosome 22. Consequently, a fusion protein *Bcr/abl* is produced in which the normal amino terminus of the Abl proto-oncogene protein is replaced by *Bcr* amino acid sequences. The fusion protein cause unregulated activity of the Abl protein-tyrosine kinase leading to cell transformation.

Gene amplification is another mechanism by which oncogenes become activated in human tumours, resulting in high levels of gene expression. Amplification of oncogenes could lead to more rapid growth of tumours and increased malignancy. For example, amplification of the **N-myc** oncogene in neuroblastoma (a childhood tumour of embryonal neuronal cells) causes neuroblastomas to progress rapidly into aggressive tumours and increasing malignancy. The amplification of another oncogene, *erb*B-2 which encodes a receptor protein-tyrosine kinase is also related to progression of breast and ovarian carcinomas.

Oncogene Products and their Function

Whereas proteins encoded by proto-oncogenes encode normal cell proliferation, their corresponding oncogene proteins drive the uncontrolled proliferation of cancer cells. Oncogene products play a role in defective differentiation, failure to undergo programmed cell death, or function as elements of the signalling pathways that regulate cell proliferation and survival in response to growth factor stimulation.

In cancer cells growth factors act as oncogene proteins by way of their abnormal expression. Consequently, the tumour cell which produces growth factor starts responding to it. Such autocrine stimulation of the growth factor producing cell drives abnormal cell proliferation and is responsible for the development of a variety of human tumours. Oncogenes also encode growth factor receptors which are usually protein-tyrosine kinases. These receptors are frequently converted to oncogene proteins by modifications at their amino terminals. The *ras* proteins play a role in mitogenic signalling through involvement of growth factor receptors. Mutations that convert *ras* proto-oncogenes to oncogenes lead to constitutive expression of ras oncogenes which activate the MAP kinase pathway.

The intracellular signalling pathways activated by growth factor stimulation result in regulation of the cell cycle components that promote progression through the restriction point in *G*1. Growth factor stimulation induces *D*-type cyclins which play an important role in coupling growth factor signalling to cell cycle progression. Notably, the gene encoding cyclin *D*1 is a proto-oncogene which can convert into an oncogene (*D*1) by chromosome translocation or gene amplification. The resulting constitutive expression of cyclin *D*1 then drives cell proliferation in absence of growth factor stimulation.

Transcription factors also show oncogenic activity through inhibition of cell differentiation. Tumours having differentiated cells have reduced capacity to divide and predict good prognosis for the cancer patient. Thyroid hormone and retinoic acid are known to induce differentiation in many types of cells. They can cross the plasma membrane and bind to their intracellular receptors which act as transcriptional regulatory molecules. Mutated forms of both the thyroid hormone receptor (ErbA) and retinoic acid receptor act as oncogene proteins in chicken and human leukemias, respectively. The mutated oncogene receptors interfere with their normal counterpart receptors and inhibit differentiation, so that leukemic cells continue to proliferate.

Tumour Suppressor Genes

Inactivation of tumour suppressor genes also contributes to development of tumours. Normally, **tumour suppressor genes** act to inhibit cell proliferation and tumour development. When these genes are inactivated or lost they lead to abnormal proliferation of tumour cells. The first tumour suppressor gene was identified through studies on retinoblastoma. Through studies on patients with retinoblastoma that have survived, it was found that some cases of retinoblastoma are inherited, 50% of the children of an affected parent have a chance to develop retinoblastoma. According to Mendelian inheritance, this suggests transmission of retinoblastoma by a single dominant gene. Although it is a dominant trait, inheritance of the gene is not sufficient to convert a normal retinal cell into a tumour cell. That is because tumour cells have further requirements in addition to inheriting the gene. In 1971 Knudson found out that development of retinoblastoma requires two mutations that would make both copies of the gene, that is the ***Rb*** tumour suppressor gene, to become non-functional. Thus both alleles of *Rb* on the two homologous chromosomes must be inactivated to induce retinoblastoma. One defective copy of *Rb* is not sufficient for tumour development. That *Rb* gene may be considered as a negative regulator of tumorigenesis was deduced from study of deletions in chromosomes 13 and 14 displaying loss of *Rb* gene. Gene mapping studies confirmed that the loss of *normal* alleles of *Rb* resulted in tumour development, suggesting *Rb*'s function as a tumour suppressor gene. Gene transfer experiments made it clear that introduction of a normal *Rb* gene into retinoblastoma cells reverses their tumorigenecity, thus indicating activity of *Rb* as a tumor suppressor. Later studies have shown that *Rb* is lost or inactivated in many other human cancers, such as bladder, breast and lung carcinomas. Additional tumour suppressor genes that contribute to development of tumours have subsequently been identified. Studies show that tumour suppressor genes are involved in the development of both inherited and non-inherited cancers of humans. Mutations in the tumour suppressor genes appear to be the most common molecular alterations resulting in human tumour development.

Subsequently, ***p53*** was identified as the second tumour suppressor gene that is inactivated in a wide variety of human cancers, including leukemias, lymphomas, brain tumours, sarcomas, and carcinomas of several tissues. Mutations in *p53* are said to play a role in about 50% of all cancers, making it the most common target for genetic alterations in human cancers. Cancer cells that have lost *p53* function cannot undergo apoptosis and they become highly resistant to further treatment. This may be the primary reason why tumours that typically lack a functional *p53* gene (example melanoma, colon cancer, prostate cancer, pancreatic cancer) respond much more poorly to radiation and chemotherapy than tumours that have a wild-type copy of this gene (example testicular cancer, childhood acute lymphoblastic leukemias). Like *p53*, the *INK4* and the *PTEN* tumour suppressor genes are also frequently mutated in several human cancers. Cancer of the colon may have mutation in two other tumour suppressor genes, namely, *APC* and *MADR2*. Additional tumour suppressor genes have been indicated in the development of brain tumours, pancreatic cancers and basal cell carcinoma of skin, as well as in several rare inherited cancers.

Survivin Protein and Cell Division

Survivin is a recently discovered (1997) small-sized protein that is essential for cell divison and also acts as an inhibitor for apoptosis. Because of its involvement in promoting cell proliferation and preventing apoptosis, it is considered to be the protein that interfaces life and death. Survivin has been found to be abundant in human cancers, where it has the potential as a prognostic marker for cancer, and is also a target for chemotherapy. The survivin gene, about 15 kb long is located on chromosome 17 at position $q25$.

Survivin is expressed in embryos and juveniles, but has not been detected in quiescent cells and terminally differentiated adult tissue. In actively proliferating cells, survivin expression is regulated through the cell cycle, such that it is absent in G_1 and S phases but with a peak level in G_2 and in mitosis. Attempts to localise survivin in proliferating HeLa cells using fluorochromes have indicated presence of survivin during prophase to prometaphase stages of cell division at the centromeres and associated with the microtubules. Not much is known about its prognostic importance. Importantly however, is the finding that survivin is expressed in many human malignancies, both solid and haematological and seems to be one of the most tumour-specific of all human gene products.

HUMAN PRION DISEASE

Prion diseases are a very small number of transmittable neurodegenerative disorders, very rare in humans, about one in a million per year. Prions are proteinaceous infectious particles that lack nucleic acids. Prion disease attracted attention owing to the new concept of transmission of a disease by a protein. Also because the mad cow disease, bovine spongiform encephalopathy (BSE) was thought to lead to an epidemic of prion disease. Human prion disease could be inherited or sporadic. The new variant Creuzfeldt Jakob Disease (CJD) is inherited. Sporadic cases possibly result from infection through diet and following medical procedures such as surgery, growth hormone injections, corneal transplants where the infectious agent is implicated. The disease Kuru, found in geographically isolated tribes of New Guinea, brought attention towards prion diseases in 1950s. It had been established that injesting brain tissue of dead relatives as a religious ritual was the likely route of transmission. Kuru resembles CJD clinically. Chronic wasting disease (CWD) in mule and deer, Feline spongiform encephalopathy (FSE) in cats, Gerstmann-Straeussler-Scheinker syndrome (GSS) in humans and Gatal Familial insomnia (FFI) also in humans, are also listed as prion diseases.

The human prions are composed of an abnormal isoform of a normal cellular protein PrP. The pathogenic isoform of the protein PrP is designated as PrP^{sc} where the superscript sc is derived from scrapie because scrapie is the prototypic prion disease. Thus PrP^{sc} is the scrapie-like isoform of PrP. The prion gene *PRNP* is located on the short arm of chromosome 20, is about 16 kb long and contains two exons. The *PRNP* gene normally encodes a protein PrP from a single exon, where it exists in a particular cellular isoform PrP^{c} (c denotes the cellular form of PrP). This protein is found predominantly on the surface of neurons attached by a glycoinositol phospholipid anchor, and is protease-sensitive. It is thought to be involved in synaptic function. A conformational change in PrP^{c} leads to an abnormal isoform PrP^{sc} which is the pathological transmitting agent of prion disease. PrP^{sc} is relatively resistant to proteases and accumulates in

cytoplasmic vesicles of diseased individuals. The term prion was introduced by Prusiner to imply 'proteinaceous infectious particles' and was awarded Nobel Prize in 1997.

Experimental studies on 3-dimensional conformation of murine PrP^c have indicated that PrP^c is predominantly in alpha helical form, whereas PrP^{sc} is predominantly in beta pleated sheets. It appears that this protein can adopt two quite different stable conformations. The alpha helical PrP^c in normal cells is safe, but can switch to the beta sheet PrP^{sc} form that causes disease. The inherited prion disease is said to be caused by a mutation in the *PRNP* gene. Familial CJD, GSS and FFI are all dominantly inherited prion diseases. Five different mutations of the *PRNP* gene have been shown to be genetically linked to the development of inherited prion disease. Disease produces serious behavioural and neurological symptoms and there is no effective treatment. The human variant CJD has a peak onset in the sixth decade of life. Death usually follows in about six months. Post mortem observations reveal damage in brain consisting of spongiform vacuolation, astrocyte proliferation and neuronal loss.

QUESTIONS

1. Mention the technique which can be used for the correct determination of chromosome number in man.
2. Comment on the significant contributions of the following:
   Landsteiner, Tjio, Klinefelter, Archibald Garrod Prusiner.
3. Give one example of a human disorder associated with each of the following abnormal genetic conditions:
   (*a*) non-disjunction of an autosomal chromosome in germ cell;
   (*b*) a defect in the lipid metabolism;
   (*c*) a single base substitution in DNA;
   (*d*) presence of a translocation in a parent;
   (*e*) deletion of a *B* group chromosome;
   (*f*) maternal fetal incompatibility;
   (*g*) beta sheet conformation of protein.
4. After the birth of a Down's child, an anxious couple approaches you for genetic counselling. How would you proceed to determine the risk involved in future pregnancies?
5. Under what circumstances in a family can prenatal diagnosis by amniocentesis be useful for preventing the birth of an abnormal child?
6. An Rh negative woman married to an Rh positive man need no longer worry about the birth of an Rh positive child. Explain.

SELECTED READINGS

Aaronson, S.A. 1991. Growth Factors and Cancer. *Science* **254**: 1146–1153.

Athwal, R.S. and McBride, D.W. 1977. Serial Transfer of a Human Gene to Rodent Cells by Sequential Chromosome Mediated Gene Transfer. *Proc. Natl. Acad. Sci. U.S.* **74**: 2493.

Bacchetti, S. and Graham, F.L. 1977. Transfer of the Gene for Thymidine Kinase Deficient Human Cells by Purified Herpes Simplex Viral DNA. *Proc. Natl. Acad. Sci. U.S.* **74**: 1590.

Bishop, J. O. 1974. The Gene Numbers Game. *Cell* **2**: 81.

Borgaonkar, D. S. 1974. The XYY Chromosomes. Male or Syndrome? *In* Progress in Medical Genetics. Vol. V. Eds. Steinberg and Bearn. Grune & Stratton, New York.

Bos, J. L. 1989. *Ras* Oncogenes in Human Cancer : A Review. *Cancer Res.* **49**: 4682–4689.

Boué, A. and Boué, J. 1974. *In* Physiology and Genetics of Reproduction, Eds. Coutinho, P. B. and Fuchs, F. Plenum Press, New York, p. 317.

Cantley, L.C. and Neel, B.G. 1999. New Insights into Tumour Suppression: PTEN Suppresses Tumour Formation by Restraining the Phosphoinositide 3-kinase/AKT Pathway. *Proc. Natl. Acad. Sci. USA* **96**: 4240–4245.

Carr, D. H. 1965. Chromosome Studies in Spontaneous Abortions. *Obstet. Gynaecol.* **26**: 308.

Davidson, R.L. *et al.,* 1976. Polyethylene Glycol Induced Mammalian Cell Hybridisation: Effect of PEG Molecular Weight and Concentration. *Somat. Cell Genet.* **2**: 271.

Deisseroth, A. *et al.,* 1975. Haemoglobin Synthesis in Somatic Cell Hybrids: Coexpression of Mouse with Human or Chinese Hamster Globin Genes in Interspecific Somatic Cell Hybrids of Mouse Erythroleukemic Cells. *Proc. Natl. A cad. Sci. U.S.* **72**: 2682.

Deshmukh, R. N. *et al.,* 1979. Dermatoglyphics in Cleft Lip and Cleft Palate Anomaly: Familial and Teratogenic Groups. *Ind. J. Med. Res.* **70**: 814.

Epstein, C. J. and Golbus, M. S. 1978. The Prenatal Diagnosis of Genetic Disorders. *Annu. Rev. Med.* **29**: 117.

Farrell, S. A. and Worton, R. G. 1977. Chromosome loss is responsible for segregation at the HGPRT locus in Chinese hamster cell hybrids. Somat. *Cell Genet.* **3**: 539.

Flavell, R. A. *et al.,* 1979. The Structure of the Human β-globin Gene in β-thalassaemia. *Nucleic Acids Res.* **6**: 2749.

Fearon, E.R. 1997. Human Cancer Syndromes: Clues to the Origin and Nature of Cancer. *Science* **278**: 1043–1050.

Fuchs, F. 1980. Genetic Amniocentesis. *Sc. Amer.* **242**: 37.

Garrod, A. E. 1909. Inborn Errors of Metabolism. Oxford Univ. Press, Oxford.

Goss, S. J. 1976. Radiation Induced Segregation of Syntenic Loci: A New Approach to Human Gene Mapping. *In human gene mapping* 3. Baltimore Conference 1975. S. Karger Basel, New York.

Goss, S. J. and Harris, H. 1975. New Method for Mapping Genes in Human Chromosomes. *Nature* **255**: 680.

Goss, S. J. and Harris, H. 1977. Gene Transfer by Means of Cell Fusion. I. Statistical mapping of the human X chromosome by analysis of radiation induced gene segregation. *J. Cell Sci.* **25**: 17.

Hartwell, L.H. and Kastan, M.B. 1994. Cell Cycle Control and Cancer. *Science* 266 : 1821–11828.

Hunter, T. 1997. Oncoprotein Networks. *Cell* **88**: 333–346.

Jones, R. W. *et al.,* 1981. Major Rearrangement in the Human β-globin Gene Cluster. *Nature* **291**: 39.

Kan, Y.W. and Dozy, A.M. 1978. Antenatal Diagnosis of Sickle Cell Anaemia by DNA Analysis of Amniotic Fluid Cells. *Lancet* **2(8096)**: 910.

Lancker, J. L. Van. Molecular and Cellular Mechanisms in Disease. Springer-Verlag, Berlin.

Leder, P. 1982. The Genetics of Antibody Diversity. *Sc. Amer.* **246**: 72.

Legauer, C., Kinzler, K.W. and Vogelstein, B. 1998. Genetic Instabilities in Human Cancers. *Nature* **396**: 643–649.

Levine, A. J. 1997. p. 53, The Cellular Gatekeeper for Growth and Division. *Cell* **88** : 323–331.

Mitiko, Go 1981. Correlation of DNA Exonic Regions with Protein Structural Units in Haemoglobin. *Nature* **291**: 90.

McReynolds, *et al.,* 1978. Sequences of Chicken Ovalbumin mRNAs. *Nature* **273**: 723.

Okada, Y. 1979. Virus-cell Interaction during Cell Fusion by Virus. *In* Mechanisms of Cell Change. Eds. Ebert. J.D. and Okada, T. S. John Wiley & Sons, New York.

Polani, P.E., 1979. DNA Repair Defects and Chromosome disorders. *In* Human Genetics: Possibilities and Realities. Ciba Foundation Symposium. Excerpta Medica. Amsterdam, Oxford.

Riccardi, V. M. 1977. The Genetic Approach to Human Disease. Oxford.

Ruddle, F. H. 1975. Parasexual Approaches to the Genetics of Man. *Annu. Rev. Genet.* **9**: 407.

Sherr, C.J. 1996. Cancer Cell Cycles. *Science* **274**: 1672–1677.

Siniscalco, M. 1979. Human Gene Mapping and Cancer Biology. *In* Human Genetics: Possibilities and Realities. Ciba Foundation Symposium. Excerpta Medica, Amsterdam, Oxford.

Sugden, B. 1993. How Some Retroviruses Got their Oncogenes. *Trends Biochem. Sci*. **18**: 233–235.

Thompson, C.B. 1995. Apoptosis in the Pathogenesis and Treatment of Disease. *Science* **267**: 1456–1462.

Tjio, J.H. and Levan, A. 1956. The Chromosome Number of Man. *Hereditas* **42**: 1.

Turpin, R. and Lejeune, J. 1969. Human Afflictions and Chromosomal Aberrations. Pergamon Press, Oxford.

Vogel, F. and Motulsky, A.G. 1979. Human Genetics. Springer-Verlag, Berlin.

Weatherall, W.J. *et al.,* 1979. Human Haemoglobin Genetics. *In* Human Genetics: Possibilities and Realities. Ciba Foundation Symposium. Excerpta Medica, Amsterdam, Oxford.

Wheatley, S.P. and McNeish, L.A. 2005. Survivin: A Protein with Dual Roles in Mitosis and Apoptosis. *Internat. Rev. Cytol*. **247**: 35–51.

Wiener, A. S. and Shapiro, M. 1965. Advances in Blood Grouping II. Grune and Stratton, New York, London.

22

Fine Structure of the Gene

The earlier geneticists considered a gene to be the smallest unit (a bead or a chromomere) on a chromosome which could be distinguished as a functional unit, or as a unit of recombination, or as a unit of mutation. It was also believed that crossing over occurred only between genes, *not* within a gene itself. This concept emerged from results of genetic experiments which indicated that crossing over occurred between genes which were spatially distant from each other. Now we know that all crossing over may occur by breakage and reunion of molecules in DNA. When genes are closely spaced, as in the case of multiple alleles, intragenic recombination occurs, with rare frequency, so that a very large testcross progeny is required for its detection. The occurrence of mutant alleles has given us insight into the functional composition of the gene. Such alleles are separated by small distances within a gene and are functionally related. The gene is therefore a unit of function called cistron. A genetic test has been devised to define a cistron. This test is applicable to both higher organisms and phages.

Let us consider a hypothetical gene a having two mutant alleles a^1 and a^2. When a^1 and a^2 are present on different members of a pair of homologous chromosomes (+ a^1 on one chromosome, a^2+ on the other), the mutant phenotype is produced. The alleles are said to be in *trans* arrangement (in opposite positions on the two chromosomes) and they are *non-complementing* because they produce a visible mutant phenotype. When both alleles are present on the same chromosome (a^1a^2/++) it is called *cis* arrangement (Fig. 22.1) and produces the wild phenotype. Now consider a third allele a^3 present in *trans* arrangement with a^1 on the paired chromosomes, and it produces the wild-type phenotype. In this case the alleles a^1 and a^3 are said to complement each other.

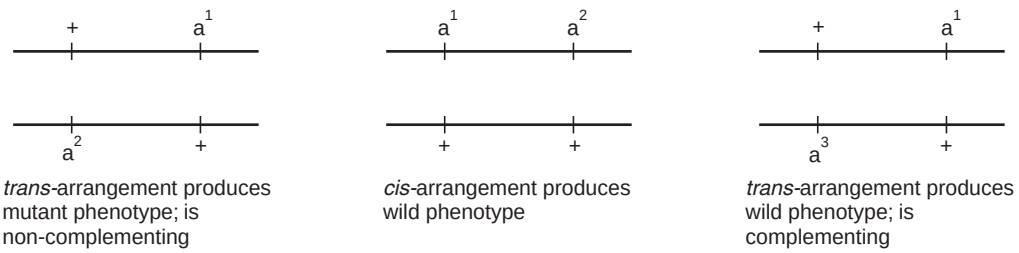


Fig. 22.1 Diagram to illustrate the *cis-trans* position effect.

Thus when two mutations present in the *trans* position produce a mutant phenotype, they are said to be members of the same functional unit called *cistron*. But if two mutations present in the *trans* position complement each other, they are said to belong to different cistrons. The concept of the cistron is based on the *cis-trans* position effect. From his studies on intragenic recombination in *Drosophila*, Lewis in 1951 devised the *cis-trans* test for complementarity between alleles. In essence it consists in comparing phenotypes produced by two mutations when the two mutations are present in *cis* and *trans* configurations.

In terms of complementation, the word *cistron* can be used in place of gene. The gene as a functional unit is a sequence of nucleotides in the DNA molecule that codes for one polypeptide chain. Genetic complementation is applicable to haploid organisms like *Neurospora.* In the case of higher organisms, complementation has been extensively studied in *Drosophila;* in some other cases it is difficult to understand and does not relate with the functional product of the gene, that is protein.

An abnormal eye condition in *Drosophila* called lozenge (Chapter 5) was used by Green and Green (1949) for mapping alleles of this locus. The sequence of the mutant sites was determined from the frequency of intragenic recombination between the alleles. They performed testcrosses of females heterozygous for the different lozenge alleles and obtained a linear map of the single lozenge gene. After that other investigators constructed linear maps of the gene in *Aspergillus, Neurospora* and in bacteria. By far the most refined analysis of intragenic recombination, in relation to nucleotides in the DNA molecule was done in bacteriophages by Seymour Benzer in 1955. He mapped thousands of independently arising rII mutants in *T*4. Such high resolution studies of intragenic recombination are called *fine structure* mapping.

MAPPING THE *r*II LOCUS IN T4

The wild type *T*4 phage produces small plaques with rough edges on both strains *B* and *K* of *E. coli.* The *r* mutants (rapid lysis) are easily distinguished by their large, sharp-edged plaques. The wild type phage do not lyse and release progeny phage as rapidly as the *r* mutants. The *r* mutants are classified into groups depending upon their behaviour in *E. coli* strains other than *B*.

The rII group of mutants analysed by Benzer differ from the wild type as they do not produce plaques on strain *E. coli K* which carries phage λ (lysogenic for λ). The wild type *T*4 (rII^+) grows on *E. coli K* (λ). Thus an rII mutant produces *r*-type plaques on strain *B* of *E. coli,* wild type plaques on strain *K* 12*S* (designated strain *S*) of *E. coli,* (that is one which does not harbour λ) and it forms no plaques on strain *K*(λ). The wild type *T*4 produces similar plaques on all the 3 strains *B*, *S* and *K*.

The rII mutants proved favourable for this study because it is possible to detect even a very small number of wild type particles among a very large group of mutants. The *r* group of mutants produces a distinct plaque type on strain *B*, and rII mutants are identified by testing on strain *K*. On strain *K*(λ) only wild type will grow. Thus when the progeny of a genetic cross between two different rII mutants are added to *E. coli K* (λ) only the wild-type recombinants will form plaques and will be detected even when their frequency is as low as one per 10^6 progeny.

The *cis-trans* complementation tests first devised by Lewis for studying the gene in the lower eukaryotes were applied by Benzer to phage. He crossed pairs of rII mutants and found that they belonged to two functional groups, due to presence of two separate segments rII A and rII B in the rII region of phage chromosome. The two groups of rII mutants, namely rII A and rII B can

be distinguished from their behaviour after mixed infection of strain *K* (λ). When mixed infection is done using two *different* mutants one belonging to rII A, the other to rII B group, the two phages multiply, cause lysis of host cell and plaques are formed. This means that normal polypeptides of both mutants are required to produce plaques—that is to say, the two *complement* each other (Fig. 22.2). But if the two mutants used for mixed infection of *K* (λ) both come from the same group, either rII A or rII B, they will not be able to form plaques (non-complementing).

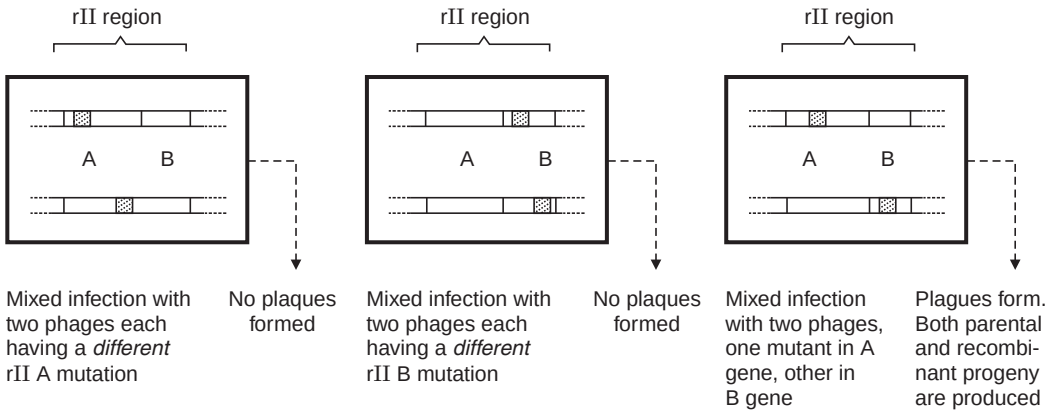


Fig. 22.2 The *cis-trans* complementation test which shows that *A* and *B* are distinct genes or cistrons in the rll region.

The fact that segments *A* and *B* show complementation demonstrates that *A* and *B* are independent, separate units. Each performs a different function resulting in distinct polypeptide end products. Polypeptides from both segments *A* and *B* are necessary for phage multiplication in strain *K* (λ). Thus *A* and *B* are two separate genes or cistrons (the term cistron was coined by Benzer). The entire rII region is a single functional unit in the sense that all mutations within this region (both segments *A* and *B*) produce the rII phenotype. The polypeptide product of *A* is not fully known; however *B* codes for a cell membrane protein.

The following procedure was used by Benzer for mapping the rII region. Strain *B* cells were infected with a 1:1 mixture of two rII mutants. After cell lysis, the progeny recovered consisted of the two parental types and recombinants of two types—double mutants and wild type. When progeny from many cells was considered, the two recombinant types were found in equal proportions. The progeny phages were made to infect *K* (λ) cells and *B* cells. The double rII mutants will not form plaques on *K* (λ), only the wild recombinant type will. Thus the total number of recombinants in the progeny are obtained simply by doubling the wild type plaques on *K* (λ). On strain *B* all types of progeny, parentals and recombinants will form plaques. The proportion of parentals to recombinants are thus determined. In this way (Fig. 22.3) by crossing rII mutants in pairs, Benzer was able to map a large number of mutations in rII region.

The method described above is sensitive enough to detect one wild type phage in a million particles. The lowest frequency of recombination between two loci is an index of the smallest distance separating two loci within which recombination could occur. But *r* mutants can revert to the wild type as first found by Hershey. Thus although Benzer had expected to detect a recombination frequency as low as 10^{-6}, the lowest he actually found was 10^{-4} (0.01 %). Since this represents half of the recombinant types, the lowest frequency would be 0.02 %. Therefore, the lowest frequency of recombination that can be accurately determined between two rII mutants is 0.02 per cent (that is a distance of 0.02 map units). Now the entire circular genetic map of *T*4 is 1500 map units

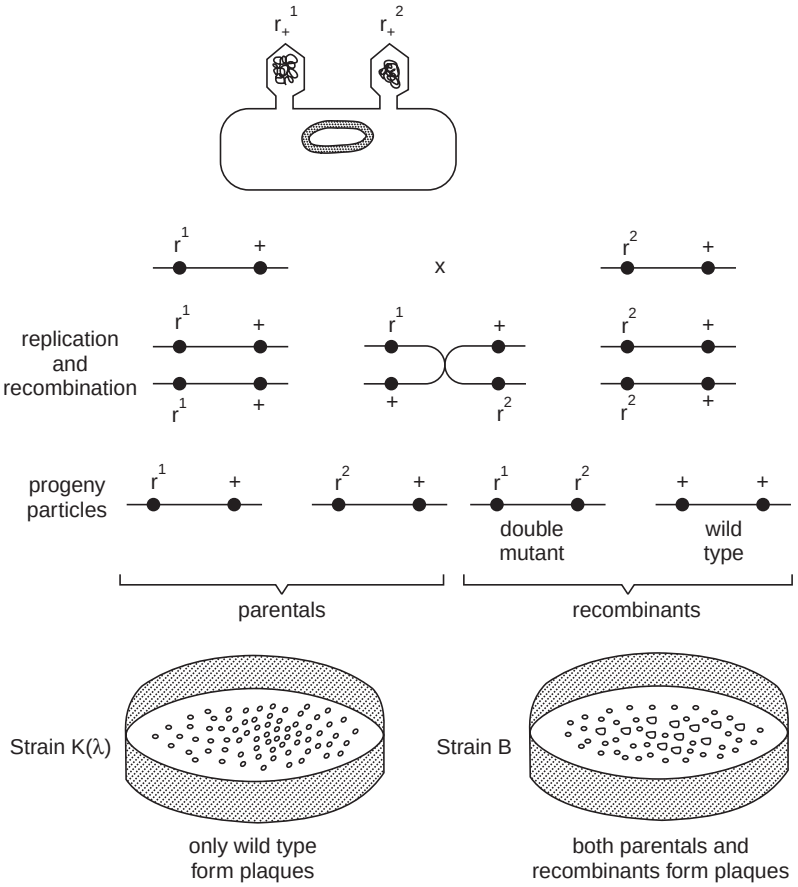


Fig. 22.3 Benzer's technique for fine structure genetic mapping of rII region of *T*4.

containing about 2×10^{-5} nucleotide pairs. Therefore a map distance of 0.02 map units would proportionately be 1.3×10^{-5} of the entire phage genome. Thus the minimum distance between two points within which recombination can occur would be $(1.3 \times 10^{-5}) \times (2 \times 10^{5})$ which amounts to about 3 nucleotide pairs. Later work on phage DNA by Yanofsky has shown that recombination can occur even between adjacent nucleotides.

With this technique Benzer mapped about 2000 mutations in rII region. Obviously it became difficult to cross such a large number of mutants in mixed infections. Fortunately there was a second class of mutations called *deletion mutations* which had a deficiency in either the *A* or *B* segment. In phage deletions are identified by their stability and their failure to have second mutations which revert them to the wild type. Further, recombination does not occur in the region of the deletion.

In *deletion* mapping a phage with a deleted segment and another phage carrying a mutation at the identical site as the deletion are used in mixed infection of strain *E. coli B* cells.

Their progeny is plated on *K* (λ). If the deletion and the known mutation occupy the same site, then no wild type recombinants will occur. But if the sites are different then wild type recombinants would appear (Fig. 22.4). By testing a deletion with a number of known mutations, the length of a deleted segment can be determined.

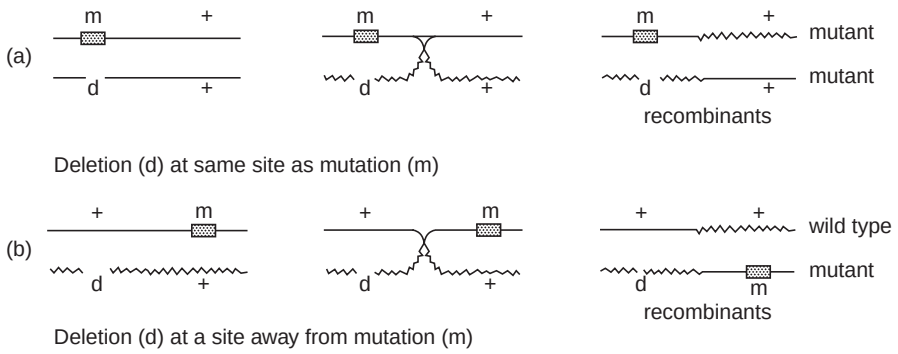


Fig. 22.4 Method of deletion mapping; mixed infections using phages with deletions and mutations at (*a*) same, and (*b*) at different sites.

In this way a *series* of deletions showing gradations in length can be mapped. They are then used to map an unknown rII mutation by crossing a mutant with a series of deletions and observing whether recombination occurs or not (Fig. 22.5). The method involves selecting a

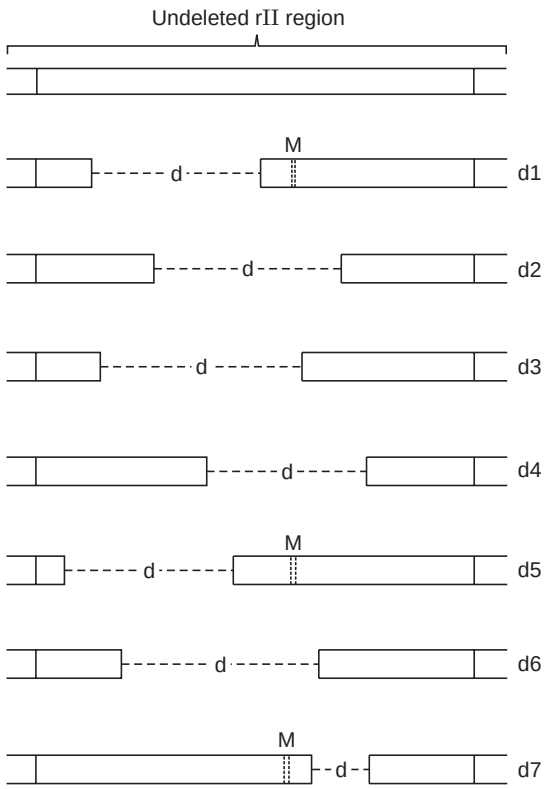


Fig. 22.5 Mapping an unknown rII mutation *M* by crossing the mutant with a series of delections.

group of deletion mutants whose lengths are known and crossing an unknown mutant with each one of them. As shown in the hypothetical Fig. 22.5, crosses between the unknown mutant *M* and deletion mutants *d*2, *d*3, *d*4 and *d*6, no recombinants are recovered because the mutant site lies opposite a deleted segment of the second phage DNA. But when the unknown mutant *M* is crossed

with *d*1, *d*5 and *d*7, recombinants are produced. In this way we can map the exact location of mutant *M*. The length of the DNA in rIIA gene is calculated to be 6 map units (800 base pairs) and in rII B region 4 map units (500 base pairs).

Intragenic Complementation

In complementation studies when each of two mutations impair the *same* function, the mutants do not complement each other. But sometimes there is *intragenic* or interallelic complementation between the different mutant alleles of the same gene, due to which a small proportion of normal phenotypes is produced. In some fungi and bacteria each of the alleles which are functionally related specifices the amino acid sequence of a different subunit of a single enzyme. Each subunit performs its own function. The association of subunits produces the secondary structure of the polypeptide; this secondary structure is responsible for enzyme activity. Thus intragenic complementation is observed through the secondary structure of a polypeptide, and not its primary structure which is determined by the amino acid sequence.

GENE CONVERSION

In eukaryotes reciprocal exchange between loci almost always produces reciprocal recombinants. In *Neurospora* the linear arrangement of ascospores in the ascus indicates directly which strands have exchanged segments during crossing over (Chapter 8). Thus after meiosis in a heterozygote Aa, due to *reciprocal* exchange the 8 ascospores are arranged in the order 4*A* and 4*a*. Sometimes there is non-reciprocal exchange between the paired strands so that instead of the 4:4 ratio, the ascospores are arranged in 5:3 or 2:6 arrangement. It is suggested that some *A* alleles are converted into *a* alleles and vice versa, and the phenomenon is known as *gene conversion* or nonreciprocal exchange (Fig. 22.6). When recombination occurs within a short chromosomal segment (intragenic) it may be sometimes nonreciprocal.

Tetrad analysis of intragenic crosses in the fungus *Ascobolus* in the 1960's showed that there are special sites in the chromosome which influence recombination in their neighbourhood. Thus if the order of linked loci is *a b c d e*..., when a cross is made between $a + \times + b$, wild type recombinants arise mainly due to gene conversion at site *a*. When the cross $b + \times + c$ is made, recombinant wild type arises due to conversion at site *b*. Similar results were subsequently obtained in other materials.

In yeast nonreciprocal recombination occurs in both mitosis and meiosis. When mutant loci lie very close to each other, nonreciprocal recombinations occur more frequently.

Molecular Mechanisms of Recombination

Studies on recombination at the molecular level have suggested two mechanisms called copy-choice and *breakage-fusion*. The second mechanism is generally accepted.

The copy-choice hypothesis was originally proposed for higher organisms by Belling in 1931. In this model recombination occurs at the time of chromosome replication. When homologous portions of two chromosomes are paired opposite each other, then during replication, each copies a segment of the other chromosome. The new strand formed along one member of the pair switches to the other chromosome and becomes a template for synthesis of the new strand. If this switching over event is repeated by the other member of a pair, then two reciprocally recombinant strands are formed (Fig. 22.7).

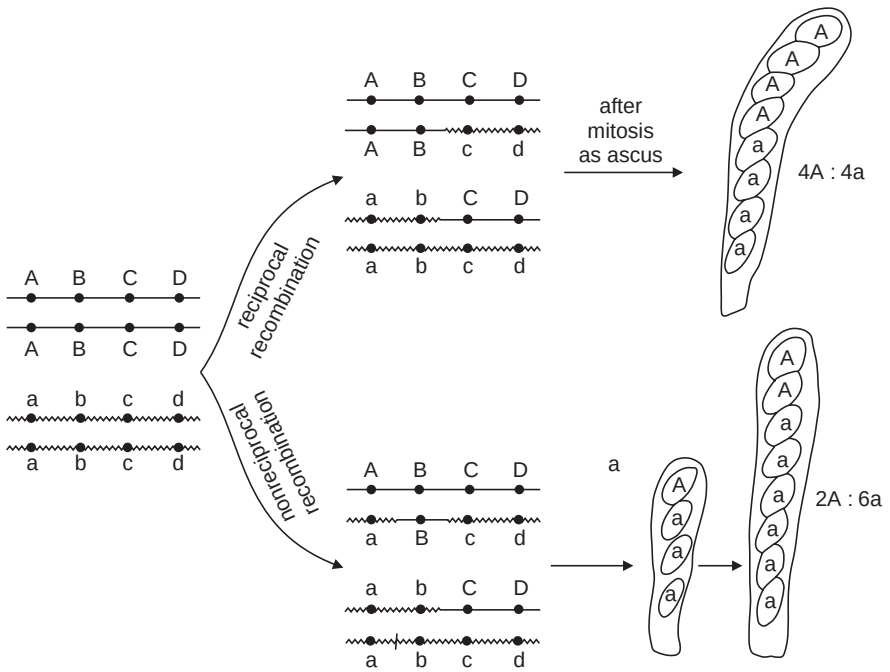


Fig. 22.6 Products of reciprocal and nonreciprocal recombination in *Neurospora*.

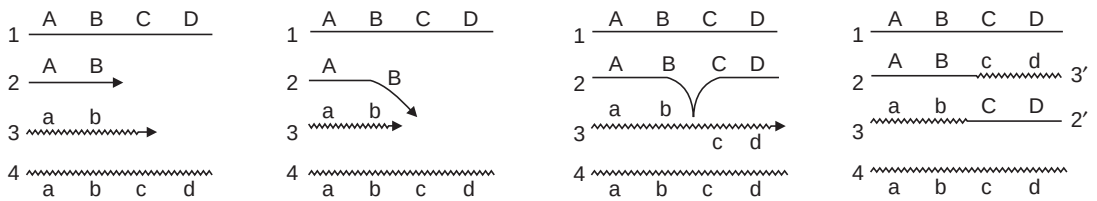


Fig. 22.7 Diagram showing the copy-choice model for recombination.

The copy-choice model has not been accepted as it cannot account for semi-conservative replication, nor the formation of heteroduplex molecules observed in viruses and bacteria. The mechanism also requires the replication of DNA prior to crossing over, which has not been detected by any of the biochemical techniques.

The *breakage-fusion* mechanism provides a better understood model of recombination. The experimental work of Meselson and Weigle (1961) and Meselson (1964) on phage λ gave support to the breakage-fusion model and evidence against the copy-choice model.

Meselson and Weigle took a mutant strain of lambda carrying linked mutant loci *a* and *b* labelled with heavy isotopes ^{13}C and ^{15}N. The label was incorporated into lambda by infecting *E. coli* cells growing on isotope containing medium. The wild type phage (+ +) did not contain density labels (^{12}C and ^{14}N). For the genetic cross, mixed infection was done by simultaneously infecting

light *E. coli* cells with heavy, density labelled lambda containing mutations *a* and *b,* and the wild type light phage particles (Fig. 22.8). The progeny particles were collected and subjected to CsCl density gradient centrifugation (Chapter 13). The different bands formed by the progeny virus particles were taken out and analysed for the presence of recombinant genotypes by infecting healthy *E. coli* cells.

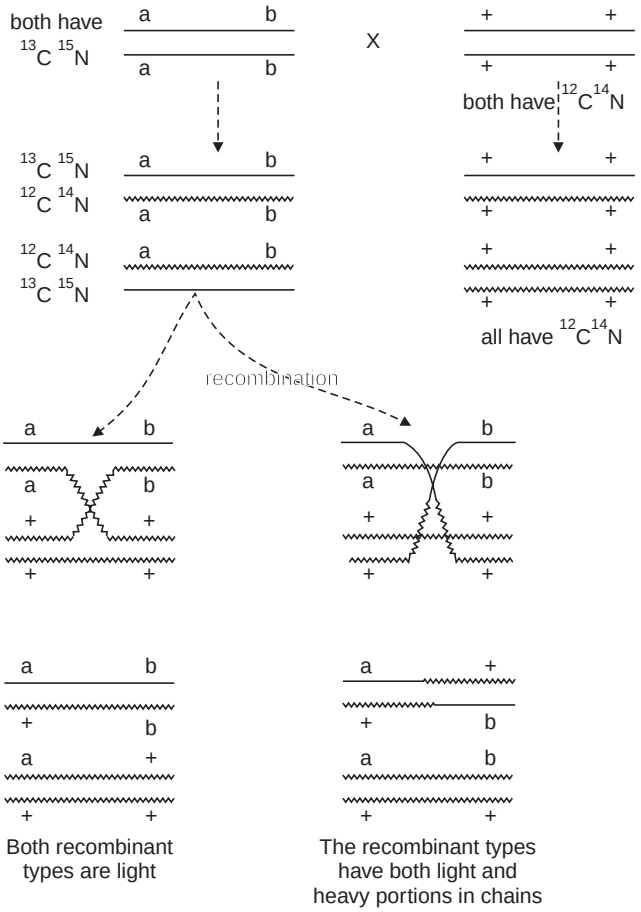


Fig. 22.8 Genetic cross between two mutant lambda strains by mixed infection of *E. coli* cells to show the breakage-fusion mechanism.

Under the conditions of the experiment, all the newly synthesised viral DNA molecules must be light (^{12}C and ^{14}N). Therefore, if recombination occurs by copy-choice, then all the recombinant phages should be light. Contrarily, if recombination occurs by breakage and fusion, then some of the recombinant phage particles should contain the heavy isotopes derived from the parental chromosome. The progeny particles from the CsCl bands showed that the recombinants contained the heavy isotopes. The experiment provided evidence for breakage and fusion as the mechanism for recombination in bacteriophage.

Site-Specific Recombination

Besides homologous recombination which takes place at any extensive region of sequence homology, **site-specific recombination** occurs between *specific* DNA sequences which are homologous

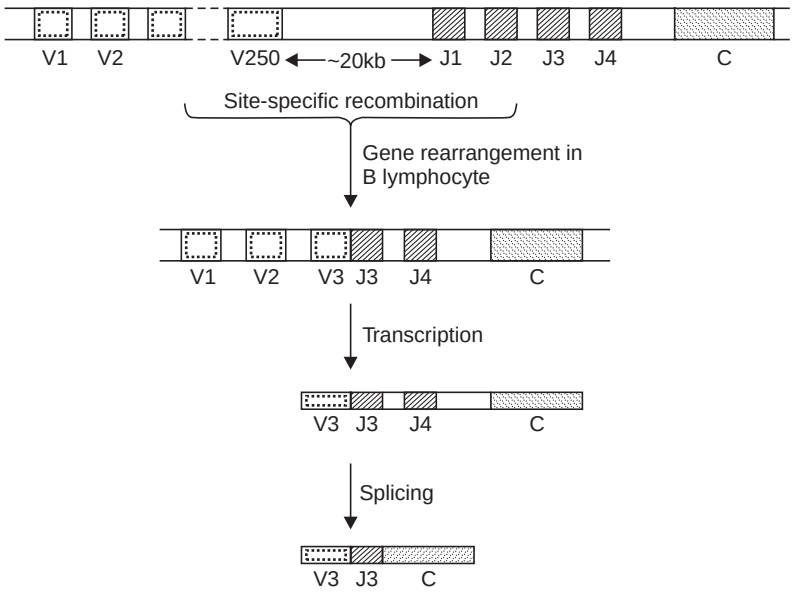
over only a short stretch of DNA. The process is mediated by proteins that recognise the specific DNA target sequences, instead of by complementary base pairing. When bacteriophage lambda (λ) infects *E. coli*, it can either replicate and kill host *E. coli* cell (**cell lysis**), or it can **integrate** into the *E. coli* chromosome forming a **prophage** that replicates with the *E. coli* genome (**lysogeny**). Under appropriate conditions, λ DNA can be excised and initiate lytic viral replication. Both integration and excision of λ DNA involve site-specific recombination between viral and host cell DNA sequences.

DNA of bacteriophage λ integrates into *E. coli* DNA at specific sites called attachment (*att*) sites. The process involves recombination between *att* sites of phage (*attP*) and the bacterium (*attB*) that are about 240 and 25 nucleotides long, respectively. The process is mediated by the phage protein integrase (Int) which specifically binds to both *attP* and *attB*. The phage and the bacterium then exchange strands within a core sequence consisting of 15 nucleotides present in both *attP* and *attB*. The Int protein produces staggered cuts in the core homology region of *attP* and *attB*, catalyses exchange of strands, and ligates the broken ends, thus integrating λ DNA into *E. coli* chromosome. The Int protein also acts in excision of the λ prophage by a process that is reverse of integration.

Site-specific recombination plays an important role in the development of the immune system in mammalian cells, which involves recognition of foreign substances (**antigens**) and provides protection against infectious agents. The two major classes of immune responses in humans and mammals are mediated by the *B* and *T* lymphocytes. The *B* lymphocytes secrete antibodies (**immunoglobulins**) that react with soluble antigens, while *T* lymphocytes express cell surface proteins (**T cell receptors**) that react with antigens expressed on the surfaces of other cells. Both immunoglobulins and *T* cell receptors are characterised by enormous diversity which enables different antibody or *T* cell receptor molecules to recognise a large variety of foreign antigens. For example, an individual is able to produce more than 10^{12} different antibody molecules, which exceeds the total number of genes (about 10^5) in the human genome. These diverse antibodies and *T* cell receptors are encoded by unique lymphocyte genes formed during the development of the immune system by site-specific recombination between distinct segments of immunoglobulin and *T* cell receptor genes.

The role of site-specific recombination in the formation of antibodies can be understood by first examining antibody structure (Fig. 21.8). The immunoglobulins consist of pairs of identical heavy and light polypeptide chains, both composed of *C*-terminal constant regions and *N*-terminal variable regions. The variable regions have different amino acid sequences in different immunoglobulin molecules, are responsible for antigen binding. It is the diversity of the amino acid sequences in the variable region that allows different individual antibodies to recognise unique antigens. In contrast to the vast array of different antibodies produced by an individual, each *B* lymphocyte produces only a single type of antibody. An important discovery by Susumu Tonegawa in 1976 indicated that each antibody is encoded by unique genes formed by site-specific recombination during *B* lymphocyte development. These gene rearrangements create different immunoglobulin genes in different individual *B* lymphocytes, so that the population of 10^{12} *B* lymphocytes in the human body includes cells that can produce antibodies against an enormous variety of foreign antigens.

The genes that encode immunoglobulin light chains consist of three regions, a *V* region that encodes the 95 to 96 *N*-terminal amino acids of the polypeptide variable region; a joining (*J*) region that encodes the 12–14 *C*-terminal amino acids of the polypeptide variable region; and a *C* region that encodes the polypeptide constant region (see Figure below) :



Immunoglobulin mRNA

In mouse, the major class of light-chain genes are formed from combinations of approximately 250 *V* regions and four *J* regions with a single *C* region. Site-specific recombination during development of lymphocytes results in a gene rearrangement in which a single *V* region recombines with a single *J* region to generate a functional light-chain gene. Different *V* and *J* regions are rearranged in different *B* lymphocytes, so that the possible combinations of 250 *V* regions with 4 *J* regions can generate approximately 1000 (4 × 250) unique light chains.

There is a fourth region in the heavy-chain genes called the diversity or *D* region which codes amino acids lying between *V* and *J*. The assembly of a functional heavy-chain gene requires two recombination events. First, a *D* region recombines with a *J* region, and next, a *V* region recombines with the rearranged *DJ* segment. In mouse there are approximately 500 heavy-chain *V* regions, 12 *D* regions, and 4 *J* regions, so that the total number of heavy chains that can be generated by recombination is 24,000 (500 × 12 × 4).

Combinations between the 1000 different light chains and 24,000 different heavy chains formed by site-specific recombination can give rise to approximately 2×10^7 different immunoglobulin molecules. The joining of immunoglobulin gene segments is often imprecise, so that one to several nucleotides may be lost or gained at the sites of joining, thus producing a further increase in diversity. The mutations resulting from the loss or gain of nucleotides increase diversity of variable regions about a hundredfold, attributable to the formation of about 10^5 different light chains and 2×10^6 heavy chains, which can then combine to form more than 10^{12} distinct antibodies.

The *T* cell receptors consist of two chains called α and β, each of which contains variable and constant regions. The genes encoding these polypeptides are produced by recombination between *V* and *J* segments (the α chain) or between the *V*, *D* and *J* segments (the α chain), analogous to the formation of immunoglobulin genes. Furthermore, site-specific recombination between these distinct segments of DNA, in combination with mutations arising during recombination, generates a similar level of diversity in *T* cell receptors to that in immunoglobulins.

The process of *VDJ* recombination that produces genes encoding polypeptides of *T* cell receptors is mediated by signal sequences adjacent to the coding sequences of each gene segment. These recombination signal sequences are recognised and cleaved by a complex of two proteins, RAG1 and RAG2 which are specifically expressed in lymphocytes. The subsequent joining of the coding ends of the gene segments then produce a rearranged immunoglobulin or *T* cell receptor gene.

SPLIT GENES

It has been generally believed that a gene is a continuous, uninterrupted sequence of nucleotides which codes for a single polypeptide chain. Some recent work has shown that this may not always be so. The sequences of some eukaryotic genes (globin, ovalbumin) are found to be interrupted by nucleotides that are not represented within the amino acid sequence of the protein. Although these sequences are transcribed into hnRNA, they are later excised (spliced) and removed, so that they are not included in the mature mRNA that is translated into protein (Fig. 22.9). Such interruptions within the sequence of a gene have been variously called as introns (Gilbert, 1978), inserts (Weissman, 1978), intervening sequences (Lewin, 1980) or 'silent' DNA by some. The sequences which are included in the mRNA and translated have been called exons; the eukaryotic genes being a mosaic of introns and exons (Gilbert, 1978). Although the coding regions are interrupted, they are present in the same order in the genome as in the mRNA. Hence the name *split genes.*

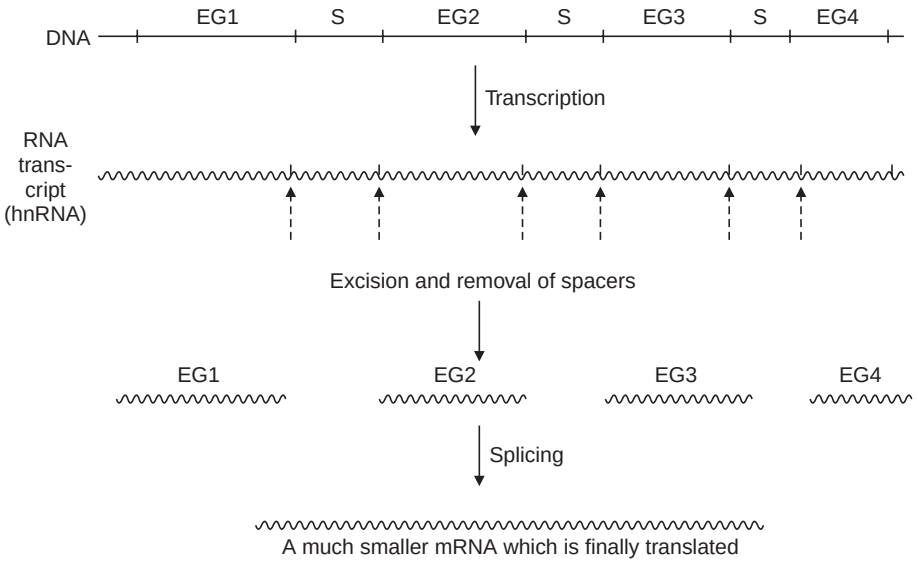


Fig. 22.9 Diagram to show relation between expressed genes (EG) and spacers(S) in split genes.

Intervening sequences were unequivocally demonstrated for the first time in the gene for β-globin in mouse and rabbit and then in chick ovalbumin (Tilghman *et al.,* 1978; Goodman *et al.*, 1977). In the β-globin gene of mouse and rabbit a single intervening sequence is present about 200–250 nucleotides inside from one end of the gene, in both somatic and germ line cells of the animal.

White and Hogness (1977) have developed an *R* loop mapping technique for identifying the site and length of an intervening sequence in the electron microscope. The method involves hybridisation of the RNA transcript with double helical DNA of the gene at high temperature and increased concentrations of formamide. Under these conditions the RNA-DNA hybrid molecules are more stable than the DNA duplexes. In *those* regions where RNA is complementary to a DNA strand, the two form a hybrid duplex; the unpaired DNA strand in the region forms a single-stranded loop visible in the electron microscope (Fig. 22.10). With this technique Tiemeier *et al* (1978) found out that the intervening sequence in the β-globin gene forms an *R* loop about 585 base pairs long and located about 200–250 base pairs inside of the 3′end. The formation of loops has also been observed in adenovirus and *SV* 40.

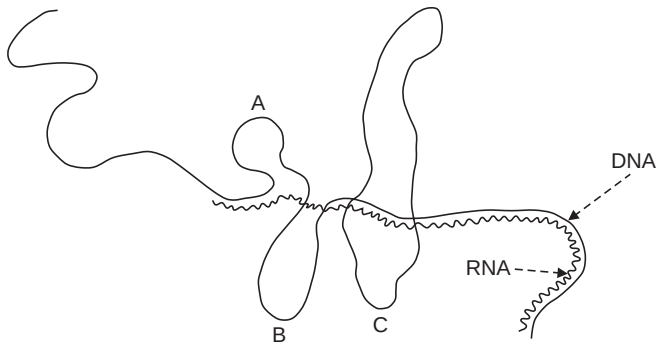


Fig. 22.10 Diagram from electron micrograph of the "hexon" gene of adenovirus. Loops *A*, *B* and *C* represent intervening sequences.

The ovalbumin gene in chicken has been shown to contain seven intervening sequences by the combined techniques of restriction enzyme and *R* loop mapping (Fig. 22.11). The insulin gene in mammals has two intervening sequences. Unlike mammals, rat has two insulin genes *I* and *II*; gene for insulin II has two intervening sequences, one 119 base pairs the other 499 base pairs long. Even in the chloroplasts and mitochondria of eukaryotes, the genes for the larger *r*RNA contain intervening sequences.

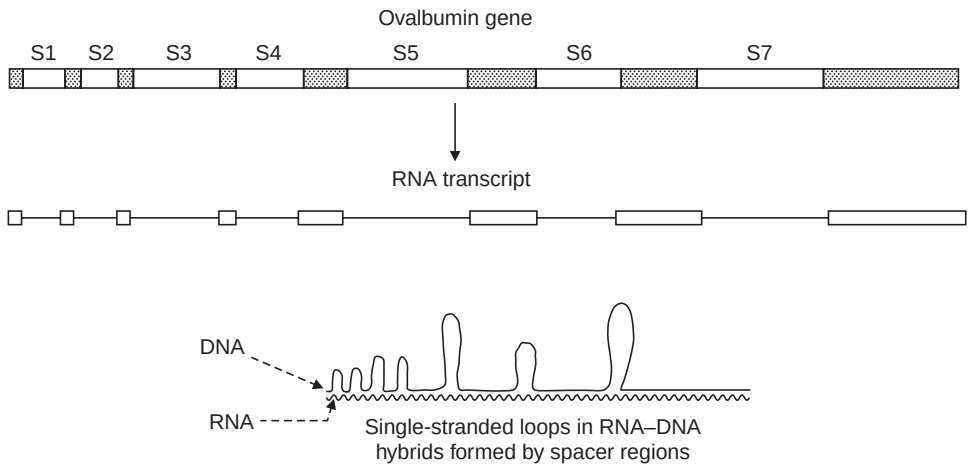


Fig. 22.11 Structure of ovalbumin gene and looping out of seven intervening sequences.

EM studies on cloned nucleolar *r*RNA genes have shown intervening sequences. In fact one of the first reports for split genes was on the 28S *r*RNA genes in *D. melanogaster* (White and Hogness, 1977; Pellegrini *et al.,* 1977). These genes were found to be of variable lengths. When the complementary *r*RNA was hybridised to the *r*DNA, the longer genes were found to contain sequences that were not represented in *r*RNA. The genes for tyrosine and phenylalanine *t*RNA in yeast contain a short intervening sequence of less than 20 bases in the region coding for the anticodon loop of *t*RNA. Lately split genes have been reported from various materials, and posttranscriptional removal of intervening sequences appears to be quite common in many eukaryotic genes.

Intron, Exon Structure of Split Genes

Electron microscopic images of RNA-DNA hybrids and subsequent nucleotide sequencing of cloned genomic DNAs and complementary DNAs (*c*DNAs) demonstrated that the coding region of the mouse β-globin gene (encodes the β subunit of haemoglobin) is interrupted by two introns that are spliced out from *m*RNA. In general, the intron-exon structure of eukaryotic genes is complicated, the amount of DNA in intron sequences usually exceeding that in the exons. For example, the chicken ovalbumin gene contains eight exons and seven introns distributed over 7700 base pairs (7.7 kilobases or 7.7 kb) of genomic DNA. The exons total only about 1.9 kb, which means that about 75% of the gene consists of introns. A notable example is the human gene that encodes the blood clotting protein factor VIII. This gene contains about 186 kb of DNA and is divided into 26 exons. The *m*RNA is only about 9 kb long, which implies that about 175 kb of the gene contains introns. On average, introns constitute about 10 times more DNA than exons in the genes of higher eukaryotes.

Although most eukaryotic genes contain introns, their presence is not universal. Histone genes do not contain introns which shows that introns are not required for gene function in eukaryotic cells. Most genes of the simple eukaryote yeast lack introns. Introns are however, present in some rare genes of prokaryotes. Most introns have no cellular function, although a few have been found to encode functional RNAs or proteins. In general, introns are considered to represent remnants of sequences that were important earlier in evolution. Introns may have helped to accelerate evolution by facilitating recombination between exons of different genes, the process is called exon shuffling. Recombination between introns of different genes could produce new genes containing novel combinations of exons. This hypothesis is supported by studies that have demonstrated that some genes are chimaeras of exons derived from several other genes, thus indicating that new genes can be formed by recombination between intron sequences.

GENE FAMILIES AND PSEUDOGENES

While most prokaryotic genes are represented only once in the genome, some genes in eukaryotes are repeated many times in multiple copies, called **gene families**, thus contributing to the large size of the genome. In specific cases, multiple copies of a gene are necessary to produce RNAs or proteins that are required in large quantities, such as ribosomal RNAs and histones. In other cases, distinct members of a gene family may be transcribed in different tissues or at different stages of development. For instance, the alpha and beta subunits of haemoglobin are both encoded by gene families in the human genome, but different members of these families are expressed in

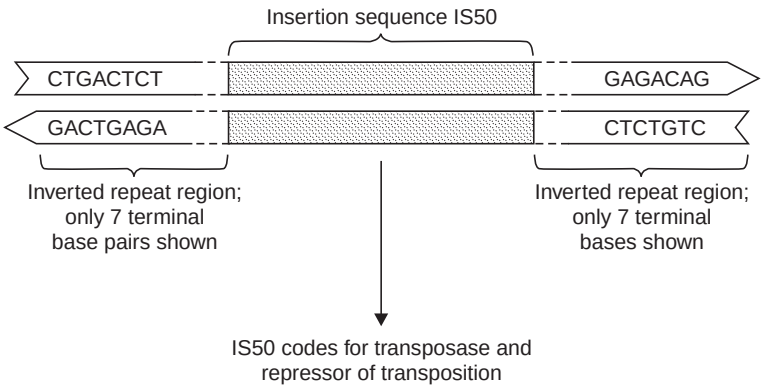
embryonic, foetal and adult tissues. Members of some gene families such as the globin genes are clustered within a region of DNA while other gene families are dispersed on different chromosomes.

Gene families are thought to have arisen by duplication of an original ancestral gene. The different members of the family then diverged during evolution as a consequence of mutation. Such divergence can lead to the evolution of related proteins that can function in different tissues or at different stages of development. For example, the foetal globins have a higher affinity for O_2 than adult globins which allows the foetus to obtain O_2 from the maternal circulation. However, not all mutations offer advantage in gene function. The mutations in some gene copies have resulted in their loss of ability to produce a functional gene product. For example, the human alpha and beta globin genes each contain two genes that have been inactivated by mutations. The non-functional gene copies are called pseudogenes, represent evolutionary relics which increase the size of the eukaryotic genome without making a functional genetic contribution.

TRANSPOSABLE ELEMENTS

The chromosomes of *E. coli* and other prokaryotes contain discrete moveable segments of DNA called **transposable elements**, which can be translocated to other locations. Such DNA sequences can move from one position on a chromosome to another, and from one DNA molecule to another. They are detected when they are present in the midst of sequences where they were not previously present. They are integrated into DNA by mechanisms that do not require recognition of sequence homology. Transposable elements were first described by McClintock in 1965 in maize (*Zea mays*) when controlling genes were found to be transposed to different locations. The controlling action of these genes is evident from their ability to suppress the function of neighbouring genes, known as position effect. For example, the gene *A*1 which codes for anthocyanin pigment in corn kernels is regulated by two types of controlling genes. One type regulates the function of gene *A*1; the other type called signalling gene is present elsewhere, on a different chromosome, and gives the signal for activity to which the regulating gene responds. DNA segments similar to transposable elements are also present in eukaryotes. These elements are involved in the formation of inversions, deletions and fusions of DNA segments. In fact they appear to be hot spots (vulnerable points) for these chromosomal changes. As such they are efficient tools for the study of mutagenesis. Transposable elements are divided into two general classes, depending on whether they transpose via *DNA intermediates*, or via *RNA intermediates*.

Many transposons in bacteria move via DNA intermediates. The simplest transposable elements are the **insertion sequences** (**IS elements**), ranging in size from 800 to 2000 nucleotides. Insertion sequences usually encode only the **transposase** protein required for transposition, and one or more additional proteins that regulate the rate of transposition. Like many transposable elements in eukaryotes, they possess inverted repeat sequences at their termini. Transposase acts at the site of inverted repeats for recognising and mobilising the IS element. Upon insertion they create a short, direct duplication of the target sequence at each end of the inserted element. The DNA organisation of the insertion sequence IS50 is shown in Figure below.



Other transposable elements in bacteria are more complex, containing one or more genes unrelated to transposition that can be mobilised along with the transposable element. Such complex transposons consist of two insertion sequences flanking other genes, which move as a unit. Insertion sequences move from one chromosomal site to another without replicating their DNA. Transposase introduces a staggered break in the target DNA and cleaves at the ends of the transposon inverted repeat sequences. Transposase has high specificity for transposon inverted repeats, but is less specific with respect to the sequence of the target DNA, hence it catalyses the movement of transposons throughout the genome. Subsequent to the cleavage of transposon and target site DNAs, transposase joins the broken ends of target DNA to the transposable element. The resulting gap in the target-site DNA is repaired by DNA synthesis, followed by ligation to the other strand of the transposon. At the end of this process there is a short direct repeat of the target-site DNA on both sides of the transposable element, indicating transposon integration.

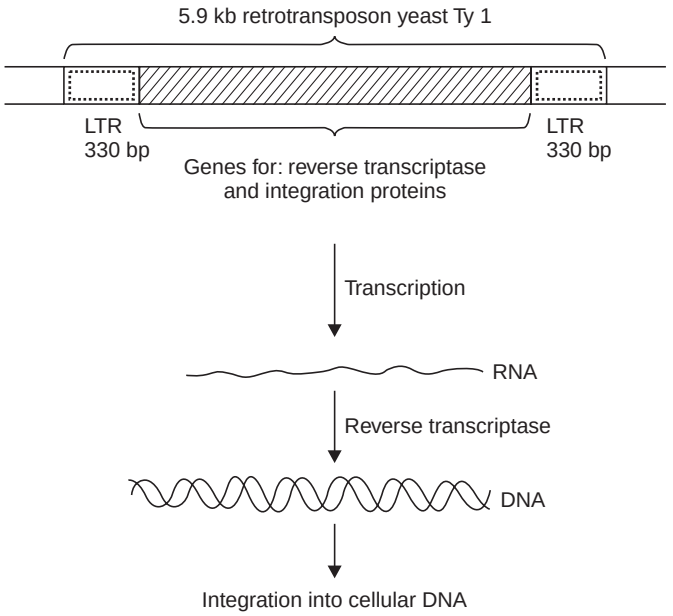
Some other types of transposons move by a more complex mechanism, in which the transposon is replicated accompanying its integration into a new target site. This results in integration of one copy of the transposon into a new position in the genome, while another copy remains at its original location. In yeasts and protozoans, transposition by a replicative mechanism is responsible for programmed DNA rearrangements that regulate gene expression. In these cases, transposition is initiated by the action of a site-specific nuclease that cleaves a specific target site, at which a copy of the transposable element is then inserted. Thus, transposable elements are capable of moving to non-specific sites throughout the genome. They can also participate in specific gene arrangements that result in programmed changes in gene expression.

Most transposons in eukaryotic cells move via RNA intermediates, the mechanism being similar to the replication of **retroviruses**. As already described (Chapters 15 and 18), RNA genomes in retroviruses replicate via synthesis of a DNA provirus, which is integrated into the chromosomal DNA of infected cells (Fig. 15.15). The enzyme **reverse transcriptase** catalyses the synthesis of a DNA copy of the viral RNA. This mechanism leads to the synthesis of a linear DNA molecule that contains direct repeats of several hundred nucleotides at both ends. These repeated sequences called **long terminal repeats** or **LTR** or **LTR-transposons** arise from *duplication* of sites on viral RNA at which primers bind to initiate DNA synthesis. Thus, LTR sequences play important roles in reverse transcription, and are also involved in the integration and subsequent transcription of proviral DNA. LTR transposons resemble retroviruses, have transcription control sequences, but lack sequences for forming the viral capsid.

The reverse transcriptase synthesises DNA in the 5′ to 3′ direction yielding a linear DNA with LTRs at both ends. The linear viral DNA integrates into the host cell chromosome by a process similar to the integration of DNA transposable elements, catalysed by a viral integration protein, and occurs at many different target sequences in cellular DNA. At the end of the process the integrated provirus is flanked by a direct repeat of cell sequences, similar to the repeats that flank DNA transposons. The viral life cycle continues with transcription of the integrated provirus, which produces viral genomic RNA and mRNAs that direct synthesis of viral proteins. The genomic RNA is packaged into viral particles that are released from host cell and can infect a new cell.

Retroviruses have been considered one type of **retrotransposon**, an element that moves via RNA intermediates. There are other retrotransposons that do not behave like retroviruses, are not packaged into infectious particles. These retrotransposons can move to new chromosomal sites within the *same* cell via mechanisms similar to those involved in retrovirus replication. Some retrotransposons present in yeast, *Drosophila*, mice and plants are structurally similar to retroviruses. They are called class I retrotransposons, have LTR sequences at both ends, encode reverse transcriptase and integration proteins, and transpose via transcription into RNA, synthesise a new copy of DNA by reverse transcriptase, and integrate into cellular DNA.

For example, retrotransposons **nts** that are about 5.9 kilo base long and include two directly repeated terminal sequences or **LTRs** (Figure below).



Each LTR contains a promoter and sequences recognized by transposing enzymes. The *Ty* elements encode a single 5,700 nucleotide long *m*RNA that begins at the promoter in LTR at the 5′ end of the element. The *m*RNA transcript contains two open reading frames (ORFs) designated Ty*A* and Ty*B* that encode two different proteins required for transposition. The number of copies of Ty elements varies between strains, though usually there are 35 in a strain.

The yeast *Ty* elements are similar to retroviruses that replicate by producing DNA using viral reverse transcriptase. The DNA integrates into host chromosome where it is transcribed to produce progeny RNA viral genomes and *m*RNAs for viral proteins. *Ty* elements are similar to retroviruses and are designated **retrotransposons**.

Some mammals, insects, plants and trypanosomes contain class II retrotransposons which do not have LTR sequences. In mammals most of these retrotransposons consist of highly repetitive long interspersed elements (**LINEs**), which are repeated about 50,000 times in the genome. Like other transposable elements, LINEs are flanked by short direct repeats of the target-site DNA. The **SINEs** that are about 100 to 400 bp long repeated sequences interspersed in the genome also belong to this group. LINEs and SINEs are retrotransposons. LINEs are autonomous elements that encode enzymes required for their own transposition. SINEs are non-autonomous elements that do not encode enzymes for their transposition, but depend upon enzymes encoded by LINEs. A very abundant SINE family in humans is the *Alu* family. The *Alu* sequences constitute a major family of these elements**,** of which there are a million copies in the genome. These sequences are about 300 bases long, have A-rich tracts at their 3′ end and are flanked by short duplications of target-site DNA sequences. SINEs arose by reverse transcription of small RNAs, including *t*RNAs and small cytoplasmic RNAs involved in protein transport. SINEs do not encode functional protein products and represent pseudogenes that arose via RNA-mediated transposition.

The transposition of SINEs and LINEs offers no advantage to the cell in which they are located. They induce mutations when they integrate at a new target site, and are thus harmful for the cell. In fact, mutations resulting from transposition of LINEs and *Alu* sequences have been associated with some cases of haemophilia, muscular dystrophy and colon cancer.

OVERLAPPING GENES

It is now becoming apparent that in both prokaryotes and eukaryotes a nucleotide sequence can code for more than one protein. The nucleotide sequence of a gene may be translated into a complete polypeptide chain. Then a part of the *same* sequence may also code for another protein of a short length. There are several ways of doing this. Sometimes the second polypeptide may be initiated or terminated *within* the sequence of the first protein. For example, the gene coding for the coat protein of the RNA phage *Q*β. The synthesis of most proteins is terminated at UGA codon when they are about 14,000 daltons. However, a protein of about 36,000 daltons may be synthesised when the sequence beyond UGA is also read in continuation and translated.

The tumour virus SV40 shows further complexities in the fine structure of a gene. Its genome is a single circular DNA molecule consisting of more than 5,200 nucleotide pairs (Fig. 18.12). Soon after infection, one half of one strand of viral DNA (designated early region) transcribes *m*RNA which is translated into 2 proteins: a small *t* antigen (mol. weight 17,000–20,000), and a large *T* antigen (mol. weight 94,000). Subsequent to this the viral DNA replicates itself. After replication, the second half of the viral genome from the opposite strand (late region) transcribes the late *m*RNA which leads to the synthesis of the following 3 structural proteins: *VP*1 (mol. weight 43,000), *VP*2 (mol. weight 39,000) and *VP*3 (mol. weight 27,000). If the total molecular weight of the proteins synthesised by the *SV*40 genome is considered, it is much more than the size of the DNA molecule (5,200 bases, *i.e.* 1733 triplets of bases) can code for. Actually there are a number of ways by which the virus uses the same gene sequence to code for more than one polypeptide chain. From these situations the concept of *overlapping genes* has emerged.

The nucleotide sequence that codes for *VP*3 also codes for the carboxy-terminal end of *VP*2. In this way *VP*3 has some amino acids identical to those at the *C*-terminus of *VP*2. Another kind of overlap is indicated by the presence of identical amino acid sequences at the *C*-terminal ends of both *VP*2 and *VP*3 on the one hand and at the amino terminal end of *VP*1. There is a sequence of about 120 nucleotides which first codes for the amino acids of *VP*2 and *VP*3, and then it codes for the amino terminal portion of *VP*1 (Weissman, 1979). Similarly, it has also been found that the

amino terminal sequence of both large *T* and small *t* antigens (specified by the early region) are identical.

Overlapping Genes in ϕX174

The bacteriophage ϕ*X* 174 is an extremely small icosahedral virus containing a single stranded DNA molecule about 5,400 nucleotides long. Nine genes *A*, *B*, *C*, *D*, *E*, *J*, *F*, *G* and *H* have been identified on ϕX DNA and they code for 9 specific proteins. The combined weight of these proteins is 250,000 daltons. A genome of about 5,400 bases has a maximum coding capacity for about 1800 amino acid residues with a combined weight of about 200,000. Obviously, the total mass of proteins coded for is significantly greater than is expected from the amount of DNA contained in ϕX genome. The works of Sanger and Coulson (1975), Barrell *et al.,* (1976) and Weisbeek *et al* (1977) have revealed several mechanisms underlying the compact genetic organisation in ϕ*X* DNA.

The authors have employed the techniques of restriction mapping using several different enzymes such as *Hind II*, *Hae III*, *Hpa II* and *Alu I*. By applying Sanger and Coulson's 'plus and minus' technique for determining DNA, sequences, Barrell *et al.,* have found the sequences of genes *D*, *E* and *J* and in fact the entire sequence of ϕ*X*174 DNA. Weisbeek *et al.,* have sequenced genes *A* and *B*.

The termination codon of gene *D* overlaps the initiation codon for gene *J* by one nucleotide. Further, the mutation *amb* 6 determined genetically to lie in gene *J* actually lies 179 nucleotides before the initiating codon for *J*. Gene *J* therefore lies in another gene (Weisbeek *et al.*, 1977).

The location of gene *E* has been identified from two amber mutations am^3 and am^6. Both the mutations lie within the sequence of gene *D*. Analysis of the sequence around the amber mutations has shown that the sequence of gene *E* (273 nucleotides long) overlaps the sequence of gene *D* (456 nucleotides long); that both genes are translated in two reading frames in two different phases. The sequence for gene *E* is displaced one base to the right from that of gene *D*. By identifying the initiation and termination codons of gene *D*, it is concluded that the sequence of gene *E* lies in the latter part of gene *D*, and specifies a protein of about 10,000 daltons. Thus gene *E* overlaps a portion of gene *D*; the sequence in the overlap region codes for two proteins in two phases (Fig. 22.12).

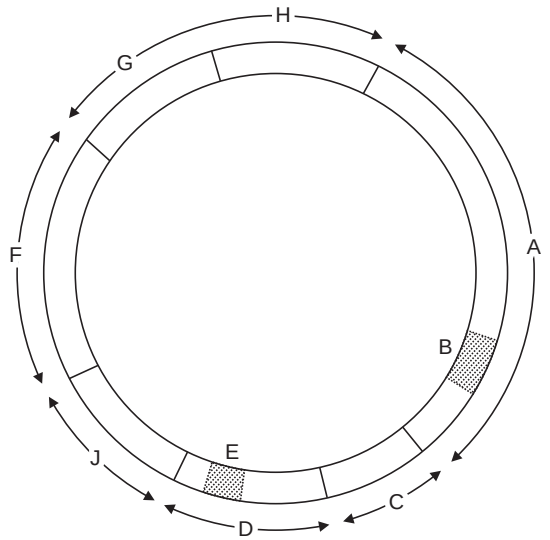


Fig. 22.12 The genetic map of ϕX174 showing overlapping genes.

The genes *A* and *B* have been characterised by Weisbeek *et al.,* (1977) by mapping positions of several mutations by a marker rescue technique. All the mutations in gene *B* have been identified within the sequence of gene *A*. However, the nonsense mutations in gene *A* do not impair the function of gene *B*; similarly nonsense mutations in gene *B* do not affect the activity of gene *A*; the genes *A* and *B* belong to different complementation groups. The study of amber mutations in gene *A* which results in synthesis of shorter chains of *A* and *A** proteins has proved that gene *B* is completely contained within gene *A* and is translated in two different reading frames: one which leads to synthesis of *A* and *A* * protein, the other for synthesis of *B* protein. Smith *et al.,* (1977) working in Sanger's lab have shown that gene A not only overlaps gene *B* but even extends beyond gene *B*.

Overlapping genes have also been identified in the single-stranded DNA virus G4 which is closely related to and has the same order of genes as $\phi X174$. In G4 also gene *A* overlaps gene *D*, and gene *E* overlaps gene *D*. It also has a gene *K* containing portions of sequence of gene *A* and *C*. The last 86 nucleotides of gene *A* and the first 89 nucleotides of gene *C* together constitute the sequence of gene *K*.

QUESTIONS

1. Can nucleotide sequences within a gene be exchanged by recombination? Describe a genetic test for a cistron.
2. Taking the example of rII locus in *T*4 bacteriophage explain the following:
   (*a*) complementation; (*b*) deletion mapping.
3. In how many different ways can you prove that all the codons in the sequence of a gene are not represented in its protein product?
4. Comment on the statement: the genome of some organisms can code for more polypeptide chains than the *number* of triplets of its bases would allow.
5. Correct or modify the earlier statements mentioned below, in the light of recent knowledge on genetic fine structure;
   (*a*) The genetic code is nonoverlapping.
   (*b*) Genes are arranged continuously in linear order on a chromosome.
   (*c*) Genetic exchange takes place in DNA segments between genes; smaller the distance between two genes, lesser the chances for exchange to occur.
   (*d*) Genes are present in definite fixed locations on chromosomes.
   (*e*) The one gene–one polypeptide chain concept.
   (*f*) Genetic exchange in eukaryotes is always reciprocal.

SELECTED READINGS

Barrell, B. G. *et al.,* 1976. Overlapping Genes in Bacteriophage ϕXI74. *Nature* **264**: 34.

Benzer, S. 1955. Fine Structure of a Genetic Region in Bacteriophage. *Proc. Natl. Acad. Sci. U.S.* **41**:344.

Borst, P. and Greaves, D.R. 1987. Programmed Gene Rearrangements Altering Gene Expression. *Science* **235** : 658–667.

Breathnach, R. *et al.,* 1978. Ovalbumin Gene is Split in Chicken DNA. *Nature* **270**: 314.

Calos, M.P. and Miller, J.P. 1980. Transposable Elements. *Cell* **20**: 579.

Cameron, J.R. *et al,* 1979. Evidence for transposition of Dispersed Repetitive Families in Yeast. *Cell* **16**:739.

Cohen, S.N. and Shapiro, J.A. 1980. Transposable Genetic Elements. *Sc. Amer.* **242**: 36.

Craig, N. L. 1995. Unity in Transposition Reactions. *Science* **270** : 253–254.

Dobritsa, A.P. *et al.,* 1981. Transposition of a DNA Fragment Flanked by two Inverted *Tn*1 Sequences. *Gene* **14**: 217.

Dugaiczyk, A. *et al.,* 1978. The Natural Ovalbumin Gene Contains Seven Intervening Sequences. *Nature* **274** :328.

Engler, J. A. and VanBree, M. P. 1981. The Nucleotide Sequence and Protein Coding Capability of the Transposable Element IS5. *Gene* **14**:155.

Federoff, N. V. 1979. On Spacers. *Cell* **16**: 697.

Gilbert, W. 1978. Why Genes in Pieces? *Nature* **271**: 501.

Gilmour, R.S. 1978. Structure and Control of the Globin Gene. *In* the Cell Nucleus. Vol. VI. Ed. Busch, H. Academic Press, New York.

Green, M.M. 1980. Transposable Elements in *Drosophila* and other Diptera. *Annu. Rev. Genetics* **14**:109.

Kazazian, H.H. Jr. and Moran, J. V. 1998. The Impact of L1 Retrotransposons on the Human Genome. *Nature Genetics* **19** : 19–24.

Levine, H. L. 1997. It's Prime Time for Reverse Transcriptase. *Cell* **88** : 5–8.

Lewin, B. 1976. DNA Sequences Coding for More than One Protein. *Nature* **264**: 11.

McClintock, Barbara 1965. The Control of Gene Action in Maize. *Brookhaven Symp. Biol. No.* **18**, p. 162.

McReynolds, L. *et al.,* 1978. Sequence of Chicken Ovalbumin *m*RNA. *Nature* **273**: 723.

Roth, D. B. and Craig, N.L. 1998. VDJ Recombination: A Transposase Goes to Work. *Cell* **94** : 411–414.

Starlinger, P. 1979. Transposons and Insertion Sequences. *In* Recombinant DNA and Genetic Experimentation. Eds. Morgan, J. and Whelan, W.J. Pergamon Press, Oxford, New York.

Weisbeek, P.J. *et al.,* 1977. Bacteriophage ϕ*X*I74: Gene A Overlaps Gene B. *Proc. Natl. Acad. Sci., USA* **74**: 2504.

Willetts, N. S. *et al.,* 1981. The Insertion Sequence IS21 of R68.45 and the Molecular Basis for Mobilization of the Bacterial Chromosome. *Plasmid* **6**:30.

23

Recombinant DNA Technology

Recombinant DNA technology has revolutionised life sciences, opening new vistas for research in molecular biology. It allows genetic manipulation using techniques for synthesising, amplifying and purifying individual genes from any type of cell. Genomics has emerged as an extension of recombinant DNA technology for high resolution mapping and characterization of *whole* genomes and gene products on a large scale, referred to as global analysis. These rapidly advancing disciplines promise new insights into sequence data, organisation, expression, and regulation of the genetic material. Genetic engineering is a natural fallout of these techniques, exploring application in medical, biotechnological industry and agricultural fields.

The success of recombinant DNA technology is based on some key discoveries: restriction enzymes that can cut and join DNA fragments (molecular scissors) for manipulations in a test tube; use of plasmids and bacteriophage DNA as vectors (vehicles) for foreign DNA that can replicate into identical copies and produce clones containing recombinant DNA; introduction of Southern blotting technique; and development of polymerase chain reaction for amplification of a specific sequence. Recombinant DNA molecules can be purified and investigated for understanding gene structure and sequence, and can be inserted into another genome.

Recombinant DNA molecules are constructed *artificially* by incorporating DNA from two different sources into a single recombinant molecule. The process is distinct from recombination which is a *natural process* in sexually reproducing organisms (description in Chapter 8), whereby a single individual gets a combination of genes from two parental organisms. Natural recombination involves the coming together of similar nucleotide sequences in chromosomal DNA, breakage and exchange of corresponding segments and rejoining. This type of recombination, notably, produces new arrangements of *alleles* and usually occurs in closely related species. It does not take place in unrelated organisms due to natural barriers.

Restriction Endonucleases

Restriction endonucleases are a class of enzymes that can recognise and bind to specific DNA sequences of four to eight nucleotides, then cleave the sugar-phosphate backbone of each of the two strands at the site of binding. All restriction enzymes cut DNA between the 3′ carbon and the phosphate moiety of the phosphodiester backbone. Therefore, fragments produced by restriction

enzyme digestion have 5′ phosphates and 3′ hydroxyls. Most restriction enzymes are present in bacteria, only one has been found in the green alga *Chlorella*. In bacteria, restriction enzymes protect the bacterium against foreign DNA such as that in viruses, by cutting up the invading viral DNA. Thus they *restrict* entry of foreign DNA into the bacterial cell. The bacterium modifies its own restriction sites by methylation, so that its own DNA is protected from the restriction enzyme it makes. Three scientists who discovered restriction enzymes and their applications namely, Adams, Nathan and Smith were awarded Nobel Prize in 1978.

More than 400 different restriction enzymes have been isolated. The restriction enzyme *Eco*RI (from *E. coli*) recognises the following six nucleotide base pair sequence in DNA of any organism: 5′-GAATTC-3′ and 3′-CTTAAG-5′.

This type of a nucleotide sequence is said to be symmetric, called a **palindrome** because both strands have the same nucleotide sequence in antiparallel orientation. Several different restriction enzymes recognise specific palindromes. Many restriction enzymes including *Eco*RI cut the sequence producing staggered ends. *Eco*RI cuts between G and A (Fig. 23.1).

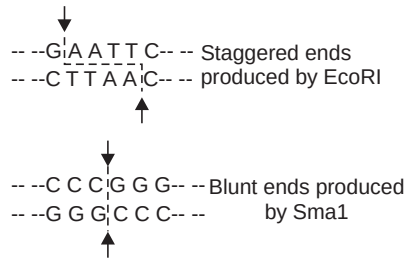


Fig. 23.1 Staggered and blunt ends produced by cleavage using restriction enzymes.

The staggered cuts give rise to a pair of identical five nucleotide long single-stranded "sticky ends". The ends consist of an overhanging piece of single-stranded DNA that are said to be sticky because they can base pair by hydrogen bonding to a complementary sequence. **Sticky ends** produced by restriction enzymes are desirable in recombinant DNA technology. If two different DNA molecules are cut with the same restriction enzyme, the fragments of each will have the same sticky ends that are complementary, enabling them to hybridise with each other under appropriate conditions. Some restriction enzymes such as *Sma*I cut the target sequence in the middle producing **blunt ends** which lack sticky ends. These can also be used for making recombinant DNA molecules with the help of enzymes that join blunt ends, or other special enzymes that synthesise short single-stranded sticky ends on the exposed 3′ strand of the blunt end.

Digestion of DNA by a restriction enzyme yields fragments of different lengths. For example, bacteriophage λ DNA cut with restriction enzyme *Eco*RI produces six fragments ranging in size from 3.6 to 21.2 kilo bases in length (one kilobase or kb = 1,000 base pairs). These fragments can be separated according to size by gel electrophoresis (Figs. 23.2, 3) or by other methods. The fragments together provide a map of the *Eco*RI sites in phage λ DNA. If multiple different restriction endonucleases are used, the locations of their cleavage sites can be used to generate detailed **restriction maps** of the DNA molecule.

Table. Some restriction enzymes and their recognition sites.

Enzyme	Source	Recognition site
*Eco*RI	*E. coli*	GAATTC CTTAAG
*Sma*I	*Serratia marcescens*	CCCGGG GGGCCC
*Pst*I	*Providencia stuartii*	CTGCAG GACGTC
*Hind*III	*Haemophilus influenzae*	AAGCTT TTCGAA
*Taq*I	*Thermus aquaticus*	TCGA AGCT

Gel Electrophoresis for Separation of DNA Fragments

Purification of DNA: The cells are homogenised and nuclei isolated for extraction of DNA. To lyse nuclei and release DNA, a negatively charged detergent sodium dodecyl sulphate (SDS) is used. The next step involves purification of DNA from contaminants such as RNA and proteins. To remove proteins, the mixture is shaken with phenol or chloroform. Phenol makes the proteins insoluble and precipitates them out of solution. Because phenol and buffered saline are immiscible, they form two separate phases. Centrifugation gives DNA or RNA in the upper aqueous phase, while protein precipitate forms the boundary between the two phases. The aqueous phase containing the nucleic acid is removed from the tube and shaken with phenol, followed by centrifugation. The process is repeated until no further protein can be removed from solution. Cold ethanol is then layered on top of the aqueous DNA solution, and the DNA is taken out with a glass rod at the interface between the ethanol and saline. RNA forms a precipitate that settles at the bottom of the vessel. To make the DNA preparation free from RNA, the DNA is treated with ribonuclease. The ribonuclease is destroyed by treatment with protease, and the protease is removed by using phenol. Purified DNA is then reprecipitated with ethanol.

Gel electrophoresis : Electrophoresis depends on the ability of charged molecules to migrate in an electric field. The *small* DNA fragments of a few hundred nucleotides or less, produced by a restriction endonuclease, can be separated by **polyacrylamide gel electrophoresis** (**PAGE**). The *larger* DNA fragments ranging in size from a few hundred base pairs to about 20 kb can be separated on **agarose gels**. The mechanism for separation of DNA fragments involves migration of DNA molecules through the pores of the matrical gel. Agarose consists of a complex network of polymerised molecules, the pore size of which is determined by the composition of the buffer and concentration of agar used. Visualisation by fluorescence microscopy has revealed that during electrophoresis DNA molecules display stretching in the direction of applied field and then contract into dense balls. The size of the DNA ball must be smaller than pore size in the gel to pass through. If the volume of the DNA ball exceeds that of pore size in gel, then the DNA molecule can only pass through by a serpentine motion resembling that of a snake. DNA molecules up to 20 kb can migrate through pores in gel by this mechanism. Studies have shown that both fragment length and molecular weight determine rate of migration of DNA molecules in a gel. For accurate size determination of a DNA molecule, marker DNA samples of known size are electrophoresed in the same gel.

The procedure for polyacrylamide gel electrophoresis is as follows. A mixture of linear DNA fragments of different sizes are driven by the current through the gel composed of organic molecules of acrylamide. Cross linking of acrylamide molecules forms a molecular sieve in a thin slab of the polyacrylamide gel between two glass plates. The gel slab is suspended between two compartments containing buffer in which the positive and negative electrodes are immersed (Fig. 23.2). The

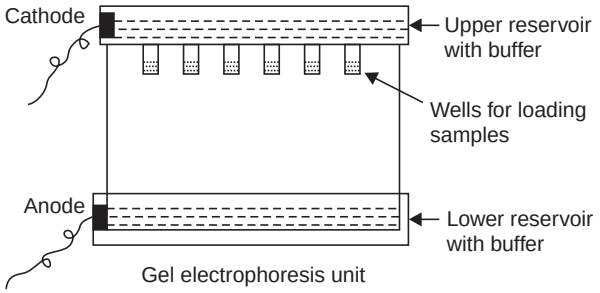


Fig. 23.2 Diagram showing main components of the gel electrophoresis unit.

sample containing linear DNA molecules (fragments) is placed in wells along the top of the gel, near the cathode of the electric field. Voltage applied between the buffer compartments allows current to flow across the slab. Because of its negatively charged phosphate groups, DNA migrates

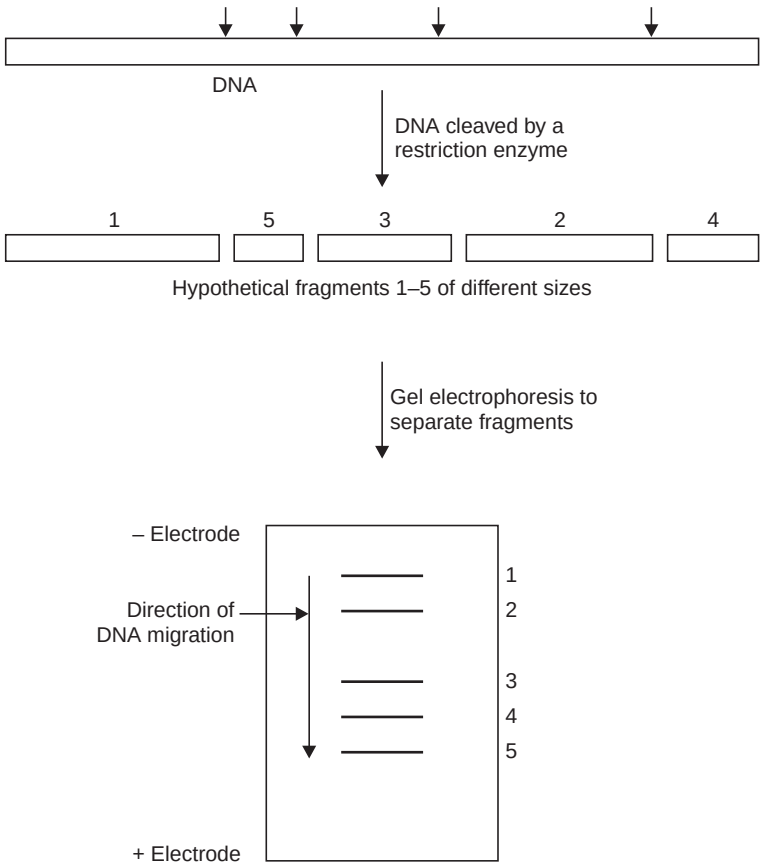


Fig. 23.3 Digestion of DNA by a restriction enzyme followed by separation of fragments on gel electrophoresis.

towards the positive electrode (anode) at speeds inversely dependent on their size. That is, the smaller fragments migrate most rapidly. All DNA molecules regardless of their length, have a similar charge density, that is, the number of negative charges per unit of mass, and therefore, all have equivalent potential for migration in the electric field. The greater the molecular mass of the DNA fragment, the more slowly it moves through the gel. Hence, the smaller the fragment, the farther it migrates in the gel. Therefore, if there are distinct size classes in the mixture of DNA fragments, these classes will form distinct bands on the gel. The fragments of different sizes thus get separated from each other (Fig. 23.3). The bands can be visualised by staining the DNA with ethidium bromide. The migration of the DNA fragments can be compared with a set of control size standards that were loaded in the same gel, to determine the exact size of each fragment in the mixture. Moreover, if the bands are well separated, a single band can be cut from the gel and the DNA sample can be extracted and purified.

Larger DNA fragments that do not pass through the pores of polyacrylamide, are fractionated on **agarose** gels which have greater porosity. Agarose is a sea weed polysaccharide, it is dissolved in hot buffer and poured into the mould and a comb placed in the molten agarose. Lowering the temperature solidifies agarose as a gel. The comb is removed producing wells in the gel. DNA samples are loaded in the wells. Gels having 0.3% concentration of agarose are used to separate larger DNA fragments. After running the gel, DNA fragments can be identified by using a labeled probe, for a specific fragment. Alternatively, the gel can be stained with a solution of ethidium bromide which intercalates into the bases in the double helix. The DNA bands can be viewed under ultraviolet light using a transilluminator.

Pulse-Field Gel Electrophoresis

The separation of DNA molecules greater than 20 kb, up to 10 Mb can be accomplished by use of **pulse-field gel electrophoresis** (**PFGE**). In PFGE, the orientation of the electric field with respect to the gel is changed in a regular manner, which causes the DNA to periodically alter its direction of migration on the gel. Each time there is a change in the electric-field orientation, the axis of the DNA must realign before it starts migration in the new direction. The difference between the direction of migration of DNA induced by successive electric fields determines the angle through which DNA must turn in order to change its direction of migration. To make the sample run in straight lines, improved methods for alternating the electric field have been devised. PFGE is used extensively in molecular biology laboratories.

Southern Blotting

When *genomic* DNA is subjected to restriction enzyme digestion it results in a very large number of fragments. A stained gel containing numerous fragments separated by electrophoresis shows a *continuous smear* of DNA instead of distinct bands. The technique of **Southern blotting** developed by E. M. Southern allows a single DNA fragment or a specific gene to be identified in this mixture. This is a widely used technique that is based on labelling of DNA and hybridisation on membranes. It is referred to as a **blotting technique** because it involves transfer of nucleic acids from gels to a solid support of a membrane and immobilisation of DNA on to the membrane. Detection of the DNA fragment is done by using a complementary strand of DNA or RNA that is called **probe** (something that detects is a probe) which hybridises with the DNA fragment.

The transfer of DNA from gel to membrane is accomplished by the *flow of buffer* through the wick, gel, membrane, and above onto adsorbent paper layers. The gel is overlaid on a filter

paper wick (3 to 4 sheets) which dips into the buffer contained in a vessel (Fig. 23.4). The hybridisation membrane is placed above the gel. A stack of paper towels comes above the membrane. The adsorbent paper towels serve to draw the buffer through the gel by capillary action. The flow of buffer carries the DNA molecules out of the gel on to the membrane (nitrocellulose or nylon membrane). Nylon membranes are considered superior in having greater binding capacity for nucleic acids and being stronger than membranes consisting of nitrocellulose. Large DNA fragments require longer time to transfer out of gel than shorter fragments. To overcome such time differences, the electrophoresed DNA on the gel is pretreated for depurination, which also denatures the fragments into single strands, making them accessible for hybridisation with the probe.

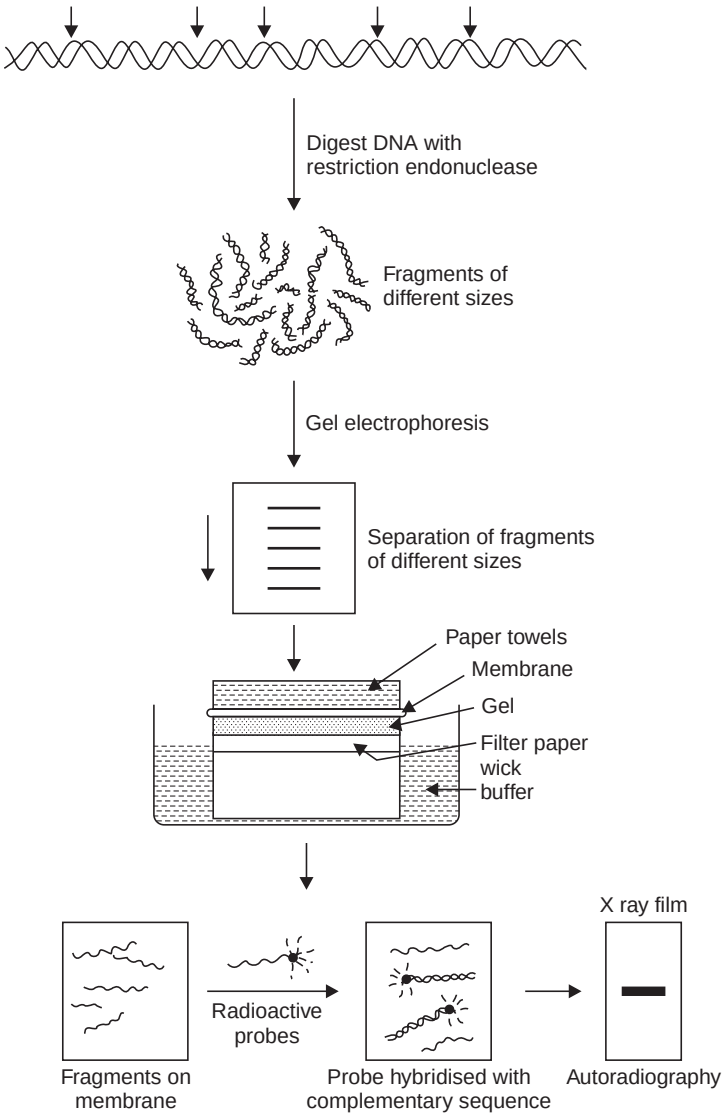


Fig. 23.4 Procedure for Southern blotting (explanation in text).

After transfer from the gel, the DNA fragments are attached (permanently fixed) to the membrane by heating at 80°C in the oven or by cross-linking using UV radiation. The DNA fragments become fixed or imprinted to the membrane at the same locations as in the gel (replica). In other words, the spatial arrangement of DNA in the gel is preserved in the sheet. Following fixation, the membrane is incubated in a solution containing labelled single-stranded DNA or RNA probe that is complementary to the blot transferred DNA sequence to be detected. Under appropriate conditions, the labelled nucleic acid probe hybridises with the DNA on the membrane. After hybridisation the membrane is washed to remove unbound probes, so that the only labelled molecules left are those that hybridised with target DNA. The membrane is then placed in contact with X-ray film for autoradiography that will reveal the position of the desired DNA fragment. By using size-calibration controls, the size of any fragment from the mixture of fragments can be determined.

Variations of Southern blotting are called **Dot** and **Slot blots**. The sample is blotted directly on the nylon or nitrocellulose sheet without prior separation on the gel.

Northern Blotting

This technique is similar to the Southern blotting technique and is used for identifying a specific **RNA molecule** from a mixture of RNAs separated on a gel. The separated RNA molecules on a gel are blotted on to a membrane and probed in the same way as DNA is blotted and probed in Southern blotting. A cloned gene (described below) having complementary sequence to the RNA being searched, can be used as a probe.

Stated briefly, it is a combination of the recent remarkable techniques that are finding widespread application. Generation of restriction fragments, gel electrophoresis, blotting techniques, cloning, together with nucleic acid hybridisation allow a specific DNA sequence to be analysed in a whole genome. To identify an individual gene, the DNA from the organism is purified and cleaved by restriction enzymes. The restriction fragments are separated by gel electrophoresis. The fragments are denatured to get single strands, then labelled with a radioactive probe. The probe consisting of single-stranded DNA or RNA must have a sequence that can base pair with the gene of interest. The probe hybridises with its complementary sequence under appropriate conditions of temperature, salt and pH. Exposure to X-ray film indicates the binding of the probe through the presence of the radioactive signal.

Recombinant DNA Molecules

The generation of recombinant DNA molecules is a process of inserting a gene of interest or donor DNA into a DNA molecule from a different source that acts as a vector or vehicle for the donor DNA. The first step is to obtain donor DNA, that is, isolate the gene of interest. Genomic DNA from the donor organism is isolated, purified and cut with a restriction endonuclease that makes staggered cuts with sticky ends in the donor DNA. The sticky ends consist of overhanging or **cohesive single-stranded tails.** Thus donor DNA will be cut into a set of restriction fragments according to the locations of the restriction sites. The vector DNA, which could be a bacterial plasmid or genome of bacteriophage λ (described later), is also cut with the *same* restriction enzyme that was used for donor DNA. The fragments thus produced would have the same complementary sticky or cohesive ends, enabling them to hybridise with sticky ends in the donor DNA fragments. The hybridised molecules produced do not have complete covalently joined sugar

phosphate backbones. To seal the backbones, the enzyme DNA ligase is used which makes phosphodiester bonds and links the two covalently to form a **recombinant DNA molecule** (Fig. 23.5).

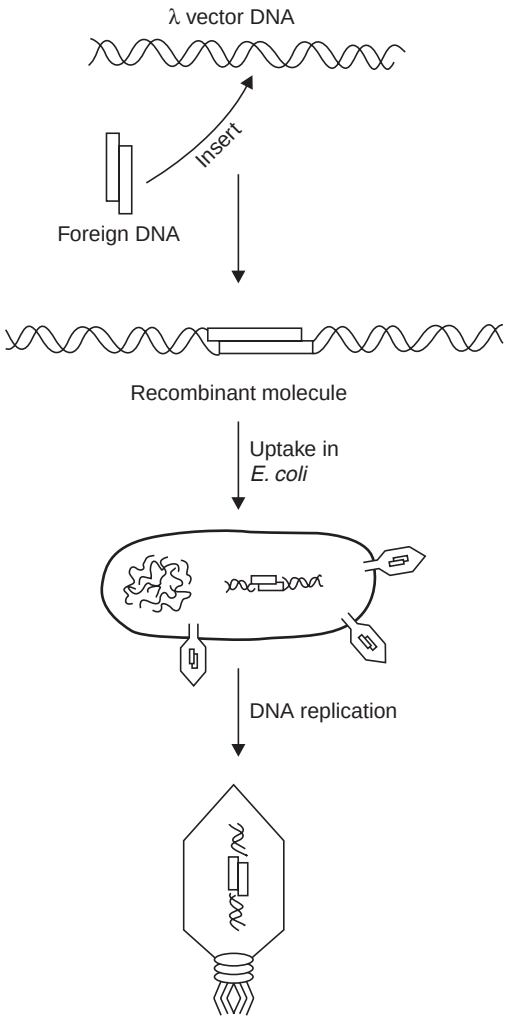


Fig. 23.5 Method for generation of a recombinant DNA molecule.

Other Sources of Donor DNA: Besides genomic DNA from a donor organism, it is possible to use RNA sequences as donors. The first step is to synthesise a DNA copy of the RNA using the enzyme **reverse transcriptase**. The **complementary DNA** or **cDNA** obtained can be used as donor DNA by incorporating it into vector DNA and making a recombinant DNA molecule. In studies that require characterisation of the gene transcript, that is *m*RNA, a *c*DNA is prepared by using *m*RNA as a template for reverse transcriptase. This is particularly useful because *m*RNA is short-lived and techniques for isolating individual *m*RNA molecules are not available. *c*DNAs can be used to learn about the variety of *m*RNAs in a cell, and the number of copies of different *m*RNAs present in a cell.

If the sequence to be inserted for making a recombinant DNA molecule cannot be isolated from natural genomic DNA, nor as *c*DNA, then chemically synthesised DNA can be used. Chemical synthesis of oligonucleotides, that is, DNA fragments 15 to 100 nucleotides in length can be developed by highly automated techniques.

Vectors for Generating Recombinant DNA

A vector must be a small molecule capable of independent replication in a living host cell; must have convenient restriction sites that can be used for insertion of the DNA to be cloned; must permit easy identification and recovery of the recombinant molecule. **Vectors** are also referred to as **cloning vehicles** or **replicons**. The basic vector systems use **bacterial plasmid** or **bacteriophage λ DNA** as vectors.

Plasmids are small circular DNA molecules that replicate independently of the bacterial chromosome. Plasmid molecules are partitioned accurately to daughter cells. Most plasmids exist as double-stranded DNA molecules. If both strands of the DNA are intact circles, the molecule is described as a **covalently closed circle** DNA (CCC DNA), if only one strand is intact, then **open circle DNA** (OC DNA). Not all plasmids are circular, some exist as linear molecules such as those in *Streptomyces* and *Borrelia*.

Plasmid DNA is 2 to 4 kb in length and has a sequence which is **origin of replication**, that signals the host cell DNA polymerase to replicate the DNA molecule. Plasmids have the advantage of carrying genes for resistance to antibiotics, so bacteria carrying antibiotic-resistance phenotype (plasmids) can be selected. Suppose we want to insert gene *X* in human genome into a plasmid vector. Fragments of human DNA are prepared by cutting with a restriction enzyme, for example *Eco*RI, and the *same* restriction enzyme is used to cut plasmid DNA (Fig. 23.3). The sticky ends will have complementary single-stranded cohesive ends that would hybridise with each other. Among the large number of fragments produced from human DNA, only a very small fraction would have gene *X*. To isolate the fragment with gene *X*, plasmid and human DNA restriction digests are incubated together, along with DNA ligase. During incubation, the sticky ends of the two types of DNA become hydrogen-bonded to each other, their broken ends sealed by ligase to form circular recombinant DNA molecules. Now there would be a large number of different recombinant DNA molecules, each containing a bacterial plasmid with a human DNA fragment. To isolate those plasmids that have human gene *X* , the process of DNA cloning has to be done.

Bacteriophage λ vectors can carry foreign DNA inserts as large as 15 kb. There are two basic types of phage λ vectors, insertional vectors and replacement vectors. The wild type phage λ DNA contains several target sites for most of the commonly used restriction endonucleases, hence is not suitable as a vector. Derivatives of the wild type phage have, therefore, been produced that either have a single target site at which foreign DNA can be inserted (insertional vectors), or have a pair of sites defining a fragment that can be removed and replaced by foreign DNA (replacement vectors). Since phage DNA can accommodate only about 5% more than its normal complement of DNA, vectors are constructed with deletions to increase space within the genome. The shortest λ DNA molecules generally used are 25% deleted.

Foreign DNA is first ligated to phage DNA (Fig. 23.6). The recombinant molecules are then introduced into *E. coli* cells. DNA replication produces numerous phage progeny containing the foreign DNA fragment (described below). This fragment can be isolated from the rest of phage DNA by restriction endonuclease digestion and gel electrophoresis.

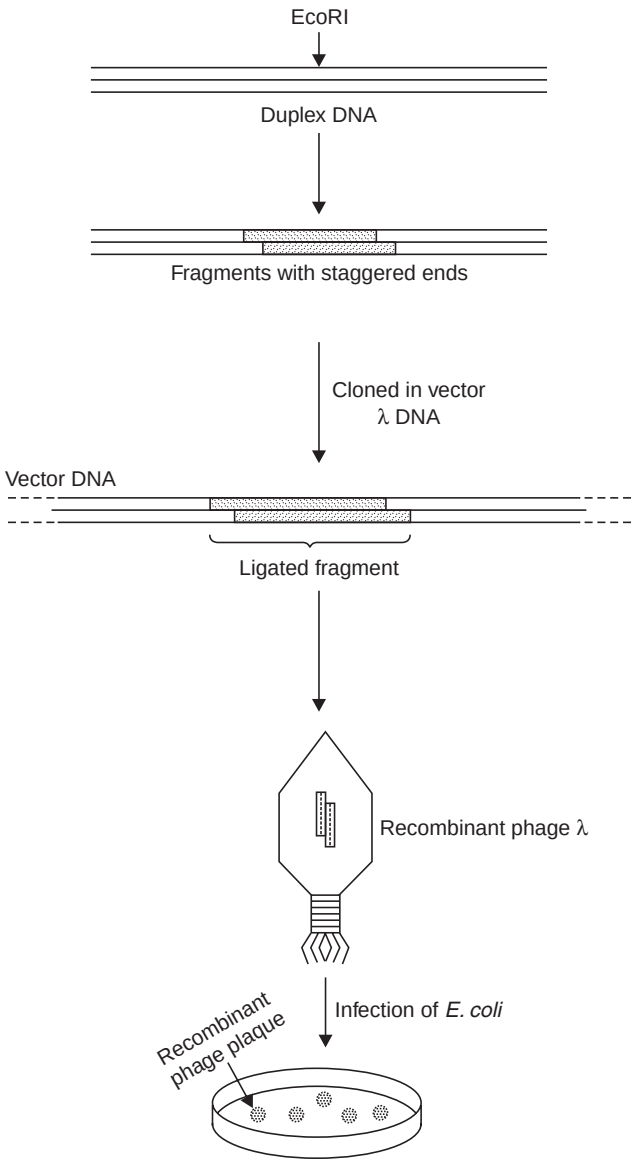


Fig. 23.6 Procedure for cloning in lambda bacteriophage vector.

Phagemids are plasmid vectors that carry the origin of replication from the genome of a single-stranded filamentous bacteriophage such as M13 or f1. Phagemids combine the best features of plasmids and single-stranded bacteriophage vectors. They have two separate modes of replication: as a double-stranded DNA plasmid, and as a template to produce single-stranded copies of one of the phagemid strands. A phagemid can therefore, be used in the same way as a plasmid vector, or it can be used to produce filamentous bacteriophage particles that contain single-stranded copies of cloned segments of DNA.

Some studies require large fragments of foreign DNA to be inserted for which phage λ vectors are not suitable. Artificially prepared **cosmid** and **yeast artificial chromosome** (**YAC**) vectors are useful for this purpose. Foreign DNA up to 45 kb in length can be accommodated in

cosmid vectors. Cosmid vectors can exist as plasmids but they also contain the complementary overhanging single-stranded ends of phage λ. The presence of bacteriophage λ sequences in cosmid vectors permit packaging of the recombinant DNA into phage particles. The λ phage then introduces these large sized recombinant DNA molecules into recipient *E. coli* cells. Cosmids also contain origins of replication and genes for drug-resistance, so that they can replicate as plasmids in bacterial cells (Fig. 23.7). However, because cosmid vectors lack the essential phage sequences necessary to form progeny phage particles, the recombinant DNA molecule depends on the plasmid sequences in the cosmid. Once inside the recipient *E. coli* cell, the recombinant cosmids form circular molecules that replicate extrachromosomally in the same manner as plasmids.

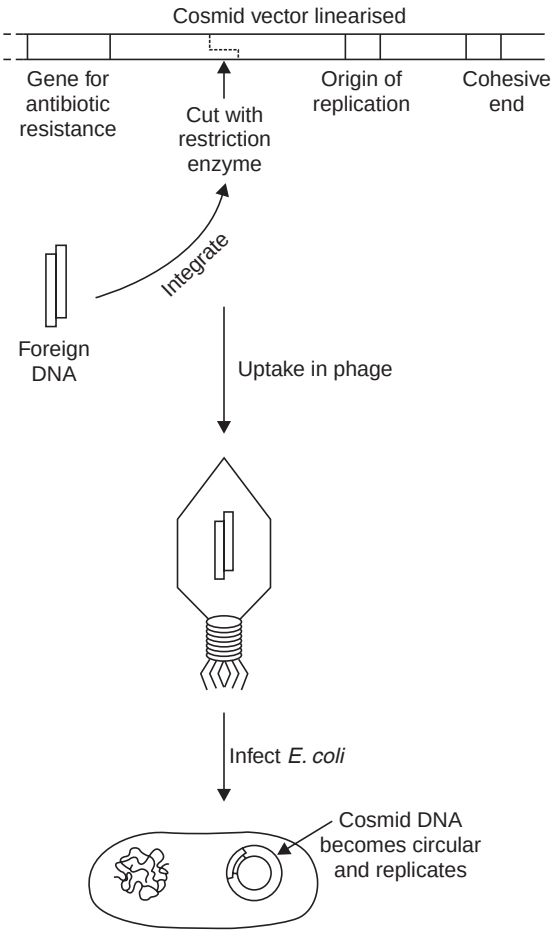


Fig. 23.7 Cloning in artificial cosmid vector.

The YAC vectors can accommodate still larger fragments , from a hundred to a thousand kilobases in length, that is one million base pairs. As the name implies, YACs are artificial versions of normal yeast chromosomes and replicate as chromosomes in yeast cells. They contain all the elements of a yeast chromosome that are required for replication during S phase, one or more origins of replication, and segregation to daughter cells during mitosis, telomeres at the ends of the chromosome, as well as centromere to which spindle fibres can attach during chromosome separation. In addition to these elements, YACs are constructed to contain the following: a gene whose encoded product allows those cells containing the YAC to be selected from

those that lack the element; the DNA fragment to be cloned. Yeasts can take up DNA from the medium, allowing YACs to be introduced into host yeast cells.

For introducing such large-sized DNA fragments (100 to 1000 kb in length), it is necessary to use restriction enzymes that recognise particularly long nucleotide sequences, 7 to 8 nucleotides containing CG dinucleotides. For example, the restriction enzyme *Not*I recognises the 8 nucleotide sequence GCGGCCGC, which cleaves mammalian DNA into fragments approximately one million base pairs long. These fragments can then be incorporated into YACs and cloned within host yeast cells. YAC technology has been used extensively in the Human Genome Project.

Bacterial artificial chromosomes (**BACs**) are based on the *F* factor of *E. coli*, and are among the most widely used vectors for very large DNA fragments of up to 300 kb. BAC vectors are 6 to 8 kb in length and include genes for essential functions such as replication (genes *rep*E and *ori*S), genes for regulating copy number (*par*A and *par*B), and genes for resistance to the antibiotic chloramphenicol.

Replicating the Recombinant Molecules

In order to use recombinant molecules for analysis, sequencing and other purposes, it is necessary to produce many *identical* copies of the specific DNA sequence of interest. A homogeneous preparation of identical copies of a desired DNA molecule is known as a **clone.** The first cloning experiments were done in *E. coli,* later extended to other micro-organisms such as *Bacillus subtilis*, *Pseudomonas* spp, yeasts and filamentous fungi. Still *E. coli* remains the most widely used host. The process of **DNA cloning** is carried out in two ways: *A.* inside a unicellular microbe such as *E. coli*; B. chemically by **polymerase chain reaction** (**PCR**).

A. DNA Cloning Inside Host Cell

DNA cloning is the selective amplification of a specific DNA fragment or a sequence for producing large quantities of the DNA fragment or sequence for detailed analysis of its structure and function. The DNA segment to be cloned is first linked to a vector DNA which can carry the foreign DNA into a host cell such as *E. coli*. As described above, two types of vectors are commonly used to clone DNAs within bacterial cells, either plasmid or the genome of bacterial virus λ. In both cases, after insertion of foreign DNA, the recombinant plasmid or phage DNA is allowed to infect a culture of bacterial cells. When the foreign DNA segment is inside the bacterium, it is replicated along with the plasmid or the viral DNA and passed on to the daughter cells (Fig. 23.8). By this procedure, a single recombinant plasmid or viral genome present in one bacterial cell can be amplified to produce millions of copies of DNA.

Cloning DNA in Bacterial Plasmids

A population of recombinant plasmids containing different fragments of foreign DNA from the genome of an organism are added to a culture of *E. coli* cells. Bacterial cells can take up DNA from the medium, undergoing **transformation**. The bacterial cell membrane is normally not permeable to large molecules such as DNA fragments, but can be made permeable by a variety of methods such as exposure to salts, heat, or high voltage for becoming **competent** for transformation. Presence of calcium ions in the medium and a brief heat shock induce uptake of DNA by bacteria. Usually only a small percentage of bacteria become **competent** for uptake of recombinant DNA molecules. Moreover, usually only a single DNA molecule will be taken up by a host bacterial cell undergoing transformation. Inside the *E. coli* cell, the recombinant plasmid replicates autonomously and is passed on to the progeny cells. Thus, large quantities of the original DNA, that is clones, would be produced. Bacterial progeny is grown in the presence of the antibiotic, so that bacteria carrying the recombinant plasmid (with genes for antibiotic-resistance) are easily selected against bacteria without plasmid.

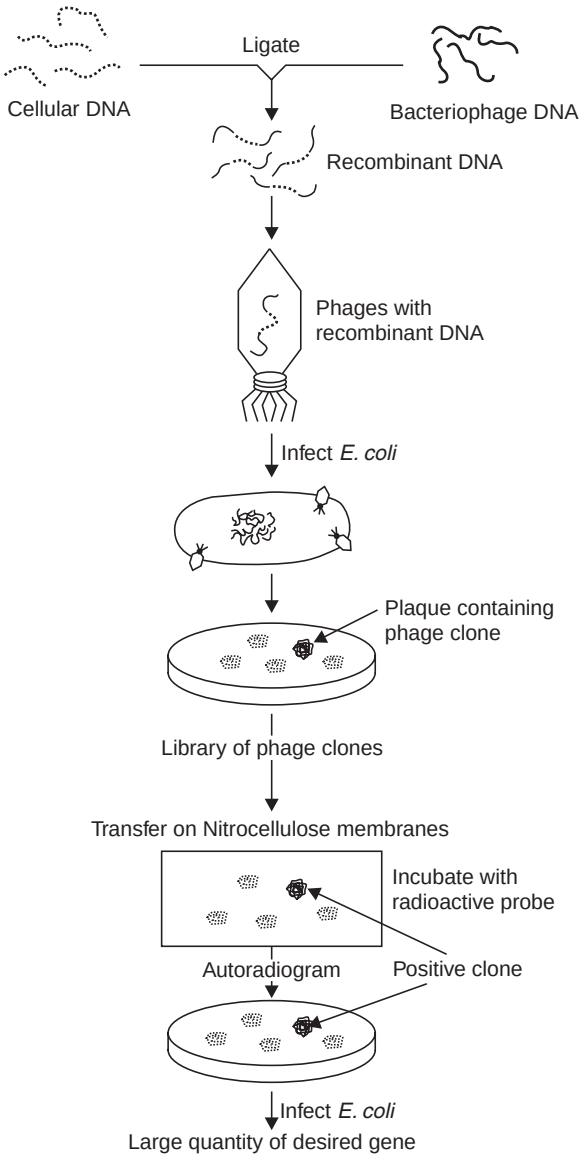


Fig. 23.8 Genome library can be made by cloning genes in phage λ. When a lawn of bacteria on a petri dish is infected with a large number of different hybrid phages, each plaque in the lawn is inhabited by a *single* clone of phages. Each clone carries a different fragment of cellular DNA. The plaque pattern (spatial locations of plaques) is transferred to a filter, and phage protein is dissolved, leaving recombinant DNA, which is denatured and immobilised (fixed) on the filter. The filter is incubated with probe labelled radioactively. This probe is a DNA copy of the mRNA representing the gene of interest. The probe hybridises with that recombinant DNA which has the matching DNA sequence, and its position determined by autoradiography. The desired clone can now be selected from its corresponding position on the petri dish and transferred to a fresh bacterial host, and the gene of interest can be produced on a large scale.

Many bacteria contain restriction systems which can influence the process of transformation through recognition and degradation of foreign DNA. Therefore, restriction-deficient strains of *E. coli* are preferred for use as transformable hosts.

The population of recombinant plasmids that was initially added to bacterial cultures contained *different* foreign DNA fragments from the organism's genome. It is possible to search those very few bacterial cells that carry a particular recombinant plasmid with a specific sequence of interest. *E. coli* cells containing the different recombinant plasmids are grown on petri dishes **at low density**, so that the progeny of each bacterial cell, representing a **clone** of cells, is physically distant from the progeny (clones) of other cells. Because a number of different recombinant plasmids were initially added to the culture medium, the resulting clones on the petri dish will have different foreign DNA fragments. After the various recombinant plasmids have produced separate colonies, it is now required to search the few colonies that contain the sequence of interest.

Screening of colonies for a particular DNA sequence is done by the combined techniques of **replica plating** and **in situ hybridisation**. The basic technique of replica plating is described

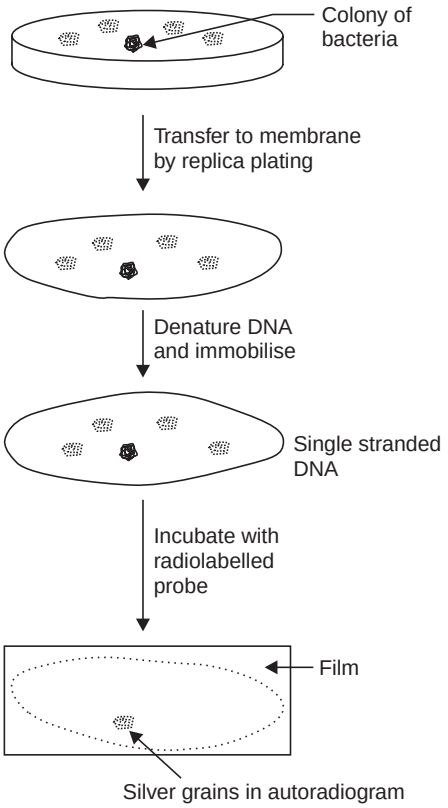


Fig. 23.9 Locating a desired gene in a bacterial colony. In the first step, when the bacterial cells in culture dish are growing into colonies they are replica plated on a filter paper, from where they are inoculated into new culture dishes (replica plating not shown in this figure). The cells in culture dish are then screened for those colonies that contain the recombinant DNA of interest. Such colonies are indentified, their cells removed and grown separately to yield large quantities of the DNA fragment of interest.

in Chapter 20. A large number of petri dishes are prepared reducing the number of colonies on each petri dish, the colonies having the same position on each dish (Fig. 23.9). For localising the specific DNA sequence, the cells on one replica plate are lysed and DNA fixed on to the surface of a filter. The DNA is denatured and subjected to in situ hybridisation. The filter is incubated with a labeled DNA probe whose sequence is complementary to the sequence of interest. After incubation, the unhybridised probe is washed from the filter. The position of the labelled hybrid is determined by autoradiography. The position of the clone carrying hybridised probe would correspond with identical clones on the original plate. Cells from these clones are then subcultured to produce large colonies. This process results in *amplification* of the recombinant DNA plasmid. The cells are harvested and recombinant DNA is separated. The recombinant plasmids are then treated with the same restriction endonuclease that was initially used to insert foreign DNA. The cloned foreign DNA (sequence of interest) is then separated from the vector (plasmid) DNA by centrifugation.

Cloning DNA in Bacteriophage Genomes

Unlike most other DNA viruses that have single-stranded DNA, bacteriophage λ has double-stranded linear DNA about 50 kb in length. Phage has a standard amount of DNA that can be 'packaged' inside the head. Some genes of bacteriophage are absolutely essential for replication of its genome, whereas others are dispensable. A bacteriophage cloning vector must contain the essential DNA sequences. In phage λ, the central part of the genome is not required for replication or packaging of λ DNA molecules in *E. coli*, and can be discarded. A modified strain of phage λ is therefore, prepared for cloning experiments. It contains two cleavage sites for the enzyme *Eco*RI, which produces three large fragments in the genome. The two outer segments at either end of the linear λ genome contain all the information essential for its growth. The central fragment which is dispensable is removed and replaced by DNA up to 25 kb in length. Recombinant DNA molecules can be packaged into phage heads in vitro. Phage particles with recombinant DNA are used to infect *E. coli* cells (Fig. 23.6). The foreign DNA is replicated along with viral DNA and produce a large number of progeny virus particles. The process accomplishes amplification of the foreign DNA segment. Lysis of the bacterial cell which are used to infect new *E. coli* cells. releases the progeny viruses. A clear spot or plaque becomes visible in the petri dish at the site of infection. Each plaque contains numerous phage particles each carrying a single copy of the desired fragment. The fragment of interest is identified by replica plating and in situ hybridisation as described for cloning of recombinant plasmids.

B. Cloning by Polymerase Chain Reaction (PCR)

PCR allows DNA sequences to be selectively amplified millions of times in just a few hours without using bacterial host cells. PCR has had a profound impact upon molecular biology. The technique was developed by Kary Mullis in 1983 using a heat-stable DNA polymerase enzyme that is stable at temperatures higher than 90°C.

The basic PCR procedure involves separation of DNA strands at high temperature (denaturation), and use of synthetic sequences of single-stranded DNA to serve as primers (Fig. 23.10). As described in Chapter 14, a primer (a DNA or RNA primer) with an exposed 3′-OH end, is essentially required for DNA polymerase to start DNA synthesis. Two oligonucleotide primers are used, 17–30 nucleotides in length, which flank the DNA target sequence that is to be amplified. One primer is complementary to one DNA strand at the beginning of the target region, the second primer is complementary to a sequence on the opposite DNA strand at the end of the target region. The primers hybridise to opposite strands of the DNA after it has been denatured, then allowing DNA synthesis by the polymerase to proceed through the stretch between the two

primers. The single-stranded DNA molecules are extended toward each other (extension), forming a double-stranded DNA molecule identical with each starting one. Repeated cycles of heat denaturation, hybridisation with primers, and extension result in exponential accumulation of the PCR amplification product, that was the target DNA sequence. The temperature resistant DNA polymerase used to catalyse extension from DNA primers is called **taq polymerase**, isolated from the bacterium *Thermus aquaticus* growing in hot springs.

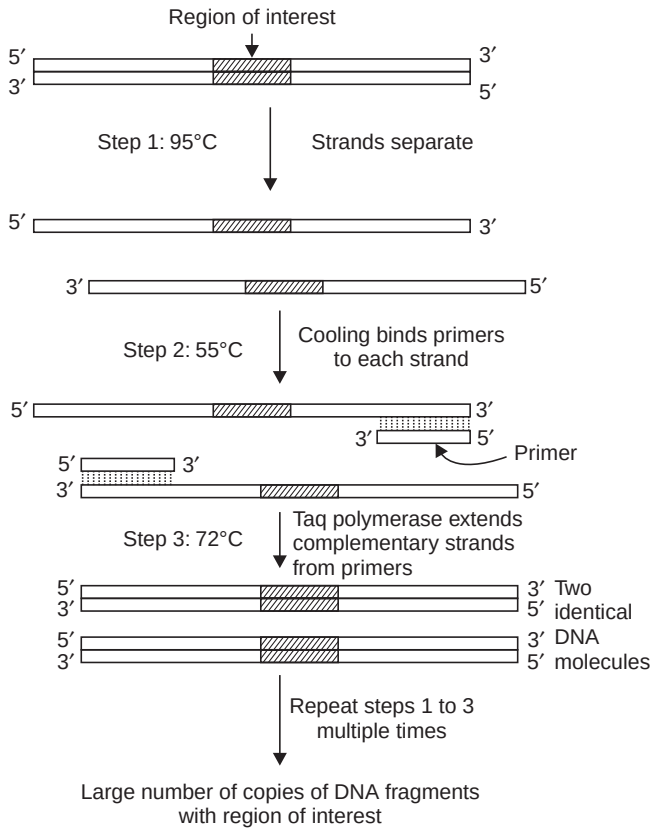


Fig. 23.10 Diagram illustrating procedure followed in polymerase chain reaction (explanation in text).

To start the PCR reaction, a small sample of target DNA is added to the test tube, mixed with all the four deoxyribonucleotides (building blocks of DNA), taq polymerase, two synthetic oligonucleotide primers that are complementary to DNA sequences at the 3′ ends of the region of the DNA to be amplified. The first step consists of heating the mixture to 92 to 94°C to denature duplex DNA into single strands. In the second step the mixture is cooled to 50 to 65°C to allow primers to bind or anneal to the strands of the target DNA. In the third step polymerase adds nucleotides to the 3′ end of the primers at 72°C.

PCR is a very sensitive technique with many applications in molecular biology. Target sequences that are in extremely low copy number in a sample can be amplified, if the primers are designed specific for this sequence. The important point to note is that, amplification is highly selective, only the DNA sequence located between the primers is amplified exponentially. Following amplification, the PCR products are loaded into the wells of an agarose gel for electrophoresis. The amplified products are visualised by staining with ethidium bromide.

Stringency in PCR Experiments

Stringency refers to the specificity with which the target DNA sequence is detected by hybridisation to the nucleic acid probe. It is influenced by temperature and salt conditions prevailing in the hybridisation step, when primers anneal to target DNA strand, as also in the post hybridisation washes. At high stringency, only those sequences that are completely complementary will be bound. At low stringency conditions, even partially matched sequences will show hybridisation. The design of oligonucleotide primers is critical to obtain maximum hybridisation specificity. Oligonucleotide primers hybridise more rapidly than primers with more nucleotides. The longer the oligonucleotide primers, the less chance there is that it will bind to sequences other than the target sequence under conditions of high stringency.

Applications of PCR

There are many applications of PCR including amplification of DNA for cloning purposes, amplifying DNA from genomic DNA preparations for sequencing without cloning, mapping DNA segments, diagnosis of disease, foetal sex determination, forensic science and molecular evolution.

Detection of mutations: PCR is useful for detection of mutations related to genetic disease including point mutations, insertions and deletions. A notable example is that of the human dystrophin gene associated with Duchenne muscular dystrophy. The large-sized dystrophin gene consists of about two million base pairs located on the X chromosome. Male patients affected with Duchenne muscular dystrophy have deletions in the exons (protein-coding regions) of dystrophin gene.

Detection of HIV: PCR can detect bacterial or viral infection from very small samples of cells. In the case of HIV, DNA isolated from peripheral blood cells can be subjected to PCR using primers designed specifically against HIV sequences. PCR amplification followed by gel electrophoresis could reveal presence of the appropriate sized PCR product and HIV infection.

Prenatal Diagnosis: PCR has proved to be of paramount importance for prenatal diagnosis of genetic disease. DNA extracted from small amounts of foetal tissue can be used for PCR. The same is true for forensic sciences. The enormous sensitivity of PCR-based procedures is useful for DNA profiling.

In Forensics: Amplification is carried out from trace amounts of DNA in samples such as hair, blood or semen collected from the crime scene. Amplified DNA can be analysed and compared with DNA from a victim and a suspect and results used in criminal investigation. This analysis or DNA profiling is called **DNA typing** or **DNA fingerprinting.**

Gene Manipulation: Following amplification, there is scope for altering the amplified sequence by incorporating extra sequences at the ends of the amplified DNA. In fact there are many applications of PCR in gene manipulation.

Fossils and Archaeology: PCR has potential in identification of organisms that became extinct a million years ago by amplification of very old (ancient) DNA. The high level of sensitivity of PCR is demonstrated when signals are produced from cell samples and degraded DNA from ancient biological samples stored in museums.

Cloning other than for Recombinant DNA

Besides DNA cloning used in recombinant DNA technology described above, there are two more types of cloning, namely, reproductive cloning and therapeutic cloning.

Reproductive Cloning: The technique used to generate an animal that has genetic material identical to that in another animal, is called **reproductive cloning**. A well known example is that of the sheep named Dolly. She was created in 1996 by a procedure referred to as

"somatic cell nuclear transfer" (SCNT). The method involves transfer of genetic material from the nucleus of an adult donor cell, in Dolly's case an udder cell, into an egg whose nucleus has been removed. The egg has no genetic material of its own. After transfer of donor nucleus from udder cell, the egg is stimulated to undergo cell division by treatment with chemicals or electric current. The multicellular embryo thus consists of a clone of cells whose genetic material, the entire genome had originated from the donor nucleus. The offspring is thus identical to its single 'parent' that provided the nucleus. At the appropriate stage the cloned embryo is transferred to the uterus of a female host where it continues to grow and give birth to an offspring.

Traditionally the term clone has been used for descendents of a single cell that contain genetic material identical to that of the parent cell. It may be noted that the cloned embryo described above has also inherted mitochondrial DNA from the cytoplasm of the enucleated egg. Hence the resulting offspring is not a clone in the strict sense of the term, as it contains genetic material from two sources, the donor nucleus of an udder cell, as well as mitochondrial DNA from the egg.

The **cloned sheep** Dolly died after six years from cancer and arthritis. Her creation however, proved that genetic material from an adult specialised cell such as an udder cell could be reprogrammed to produce an entire new organism. There was clear demonstration to dispel the earlier notion that in a specialised cell from heart, lung, udder or any type of cell, the genetic material was programmed permanently to express only the functions of its tissue of origin. A large number of animals including goats, sheep, dog, cows and cats have been cloned by the nucleus transfer technique.

Therapeutic Cloning: This is a process for production of embryos for use in research by harvesting stem cells to treat human disease, also called "embryo cloning". Stem cells have emerged as very important tools in biomedical research because they are potentially capable of producing any type of specialised cell in the human body.

Stem cells are harvested from the blastocyst stage attained by the dividing egg in about five days. The embryo is destroyed in the process which raises ethical questions and concerns.

Genomic DNA Libraries

Collections of DNA fragments obtained by DNA cloning from the entire genome of an organism constitute a **DNA library**. Basically, the genomic library is a collection of bacteria, typically *E. coli*, each carrying a fragment of DNA from the genome. In one approach for making a DNA library, all the DNA from an organism is cut into fragments with a restriction enzyme. Each segment is inserted into a different copy of the vector, thereby creating a collection of recombinant DNA molecules, which collectively represent the entire genome. These are then used to transform separate recipient bacterial cells, where they are amplified. The resulting collection of recombinant DNA-bearing bacteria is called a **genomic library**. If the cloning vector used could accommodate an average insert size of 10 kb, and if the entire genome size is 100,000 kb, then we can expect 10,000 independent recombinant clones to represent the library of the whole genome.

In another approach, genomic DNA is cleaved by one or two restriction enzymes that recognize very short nucleotide sequences such as *Sau*3A (recognises GATC) and *Hae*III (recognises GGCC). The enzymes are used in *low concentration* so that only a small percentage of target sites are actually cleaved. One can expect that the small-sized tetranucleotide sequence would occur by chance at very high frequency, so that every portion of the DNA could yield fragments. After partial digestion of DNA, the fragments are separated by gel electrophoresis. Fragments about 20 kb in length are incorporated into phage λ heads from which about half a million plaques are generated, to ensure that every portion of the genome is represented (Fig. 23.8). The important

point in this method is that, due to low concentration of the restriction enzyme, the DNA is *randomly* fragmented, and each and every target site is not cleaved. The phage recombinants produced constitute a permanent collection of all the DNA sequences in the genome. Whenever a particular sequence is required for isolation from the library, phage can be grown in bacteria. Each of the plaques produced would have originated from infection of a single recombinant phage, and can be screened for the presence of a particular sequence.

Positional Cloning or Chromosome Walking: When there is no biological information about a gene, but its position can be mapped relative to other genes or markers, it is called **positional cloning**. The approach involves cloning a gene from its known closest markers. It requires only the mapped position of the gene. On the basis of this information, researchers can locate the nearest physical markers. The closest linked marker is used to probe the genomic library. DNA cleaved randomly, as described above, generates overlapping fragments that can be used in the analysis of regions of the chromosome extending out in both directions from a particular sequence, the gene of interest. These extended regions or linked markers serve as starting points for the process of **chromosome walking** (Fig. 23.11).

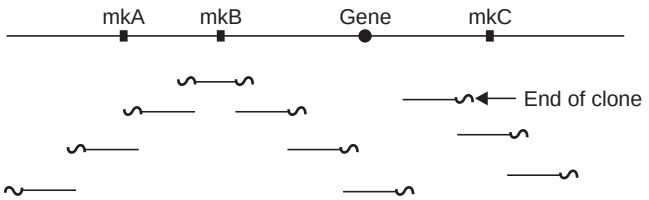


Fig. 23.11 Outline diagram indicating a chromosome walk.

Using a small sequence at the end of the linked marker, let us call it M1, as a labelled probe, find a clone in the library. The positive clone that is found will lead to identification of an adjacent gene segment which is now a second marker. Isolate the end sequence of second marker (M2) and use it to probe library and find another adjacent genomic segment which becomes the third marker. Isolate end of third marker (M3) and use it to again probe for the next adjacent segment, and so on. Using the new fragments as labelled probes the process is repeated in successive screening steps, leading to isolation of more and more of the original DNA molecule. Because this process consists of steps, hence the name chromosome walking. This approach allows study of the organisation of *linked* sequences over a considerable length of the chromosome.

Chromosome jumping is a variation of chromosome walking using larger high capacity vectors to bridge unclonable gaps. Whereas in chromosome walking each step is an overlapping DNA clone, in chromosome jumping each jump is from one chromosome location to another without "touching down" on the intervening DNA.

The gene for **cystic fibrosis (CF)**, a severe autosomal recessive disorder in humans, was identified by chromosome walking and jumping. Cloning of the CF gene was a breakthrough for studying the biochemistry of the disorder (abnormal chloride channel function), for designing probes for prenatal diagnosis, and for potential treatment by somatic gene therapy or other means.

cDNA Libraries: The method described above for making genomic DNA libraries can also be used for generating **cDNA libraries**. A *c*DNA library would contain hundreds of thousands of independent *c*DNA clones, representating collections of *c*DNA inserts. It may be recalled that *c*DNA is produced from the *m*RNA by using reverse transcriptase. If a specific gene that is being actively transcribed in a particular tissue is desired for study, then it is considered useful to convert its *m*RNA into *c*DNA and make a *c*DNA library from that sample. A *c*DNA library represents a subset of the *transcribed* regions of the genome, hence the *c*DNA library would be smaller than a complete genomic library.

Screening DNA Library for a Specific Clone

Any sequence for which a probe is available can be isolated from a recombinant library. Two types of probe can be used, those that recognise a specific DNA sequence and those that recognise part of a specific protein.

Probes for DNA sequences: A probe that consists of a single strand of DNA would be able to find and bind to other complementary denatured (single-stranded) DNAs in the library and specifically hybridise with it. The procedure for identification of a specific clone in a library is carried out in two steps. First, the recombinant phages are plated on *E.coli*, and each phage replicates to produce a plaque on the lawn of bacteria. The pattern of plaques of the library on the petri dish are transferred to an absorbent nitrocellulose membrane by laying the membrane directly on the surface of the medium. When the membrane is peeled off, the plaques remain clinging to its surface, are lysed *in situ* and DNA is denatured. The membrane is incubated in a solution containing radiolabelled probe that is specific for the sequence being searched (Fig. 23.9). Generally, the probe is a cloned piece of DNA that has a sequence homologous to that of the desired sequence. The single-stranded probe will bind to the DNA of the clone being searched. To determine the position of the positive clone, the position of the radioactive label can be found out by placing the membrane on the X-ray film. Emissions from the decay of radioactive label will reduce the grains in X-ray film, seen as a dark spot after developing the film. The procedure is called autoradiography, the exposed film an **autoradiogram**.

The probe used for finding sequence of interest in a library can also be labelled with a fluorescent dye. In that case the membrane is exposed to a particular wavelength of light that would excite fluorescence and a photograph of the membrane is taken. The position of the spot of label (radioactive or fluorescent) indicates the location of the DNA segments containing sequences complementary to the probe.

Probes for Protein Products of Genes

Expression Cloning: The protein product of a gene can be used to find the clone of its corresponding gene in a library. This can be accomplished if the amino acid sequence of the protein product is known, and the protein can be isolated in a purified form. Antibodies that bind specifically with unique protein molecules are used as probes to screen an expression library. These libraries are special cDNA libraries generated by using **expression vectors**. An expression vector is a cloning vector containing the regulatory sequences necessary to allow transcription and translation of a cloned gene. Expression vector produces the protein encoded by a cloned gene in the transformed host. Expression vectors are essentially derivatives of a phage or plasmid cloning vector that has been modified by addition of a promoter specific to the host. The cloned gene in such an expression vector is placed under the control of the promoter that ensures transcription of the cloned gene in the appropriate host cell. Expression vectors designed in this manner can yield high levels of recombinant protein in host cells that can be purified for structural and functional characterisation.

To make the cDNA library, the *c*DNA to be cloned is inserted into the special phage vector downstream from the bacterial promoter, in the correct triplet reading frame which ensures that the foreign DNA is transcribed and translated during infection. Phages that have incorporated the gene of interest form plaques that contain the protein encoded by that gene. A membrane is laid over the surface of the medium and removed with some cells of each colony attached to the membrane. The locations of these cells are identical to their positions in the original petri dish (**replica plating**). The membrane is dried, and immersed in a solution containing antibody. The antibody will bind only to the protein product of the gene of interest. For detection of positive

clones, a second antibody is prepared that is specifically against the bound antibody and labeled radioactively or with a fluorochrome. The plaque having the gene of interest is thus located on the replica plate through detection of bound antibody.

It is frequently considered useful to express high levels of a cloned gene in eukaryotic cells rather than in bacteria. The reason is that post-translational modifications of the protein (such as addition of carbohydrates or lipids) that take place in eukaryotic cells would take place normally. One system frequently used for protein expression in eukaryotic cells involves infection of insect cells by **baculovirus vectors**, which yield very high levels of protein product of a gene. High levels of protein expression can also be achieved by using appropriate vectors in mammalian cells.

Determination of DNA Sequence

The sequence of bases *A*, *T*, *G*, and *C* in genes and whole genome can provide useful information for understanding gene structure and organisation, encoded protein and cell behaviour. The first complete nucleotide sequence of an entire viral genome (ϕX174) was determined in 1977 in the laboratory of Frederick Sanger. In the 90s two new methodologies were developed for sequencing DNA, one by Sanger and Coulson in Cambridge, England, the other by Maxam and Gilbert at Harvard University. The Sanger-Coulson technique is most widely used and is described. The technique is also called **dideoxy sequencing** or the **chain termination** method because dideoxynucleotide triphosphates (ddNTPs) are the chain terminators of DNA synthesis.

To start the procedure, DNA is cleaved by a restriction enzyme to obtain a fragment which is denatured to get single-stranded DNA. The preparation is divided into four samples for four separate sequencing reactions (Fig. 23.12). Each sample tube gets DNA polymerase, the four nucleotide triphosphates (dATP, dCTP, dGTP, and dTTP) and only one of the four **dideoxyribonucleoside triphosphates** or **ddNTPs** (ddATP, ddCTP, ddGTP, and ddTTP). A radiolabelled oligonucleotide primer for DNA synthesis is used that will hybridise to the 3′ end of the single-stranded fragment and allow addition of nucleotides into the growing complementary chain. The dideoxynucleotide does not contain the deoxyribose 3′ hydroxyl group which is required for the formation of a phosphodiester bond between two nucleotides during DNA strand elongation. Therefore, incorporation of a dideoxynucleotide into the elongating DNA strand leads to termination of DNA chain synthesis, resulting in a series of labeled DNA fragments of varying lengths, all of which end at the base represented by the dideoxynucleotide in each reaction. The size of each fragment is determined by its terminal dideoxynucleotide. The lengths of the fragments increase by one base at a time and gel electrophoresis can separate fragments that differ by only one nucleotide in length. The fragments are detected on X-ray film by subjecting the gel to autoradiography. Since each of the labelled fragments would migrate to a specific location in the gel, the sequence of the fragment can be read directly from the positions of the bands in the gel.

Automated systems are available for performing **large-scale DNA sequencing**. The dideoxynucleotide reactions use fluorescently labelled primers. As the newly synthesised DNA fragments are migrating through the gel during electrophoresis, they pass through a laser beam that excites the fluorescent label. The emitted light of longer wavelength passes through a photomultiplier. The computer collects and analyses the data. This type of automated DNA sequencing allows large-scale analysis for complete genome sequences of organisms including bacteria, yeasts, *Drosophila*, and is also being used for obtaining the complete sequence of the human genome.

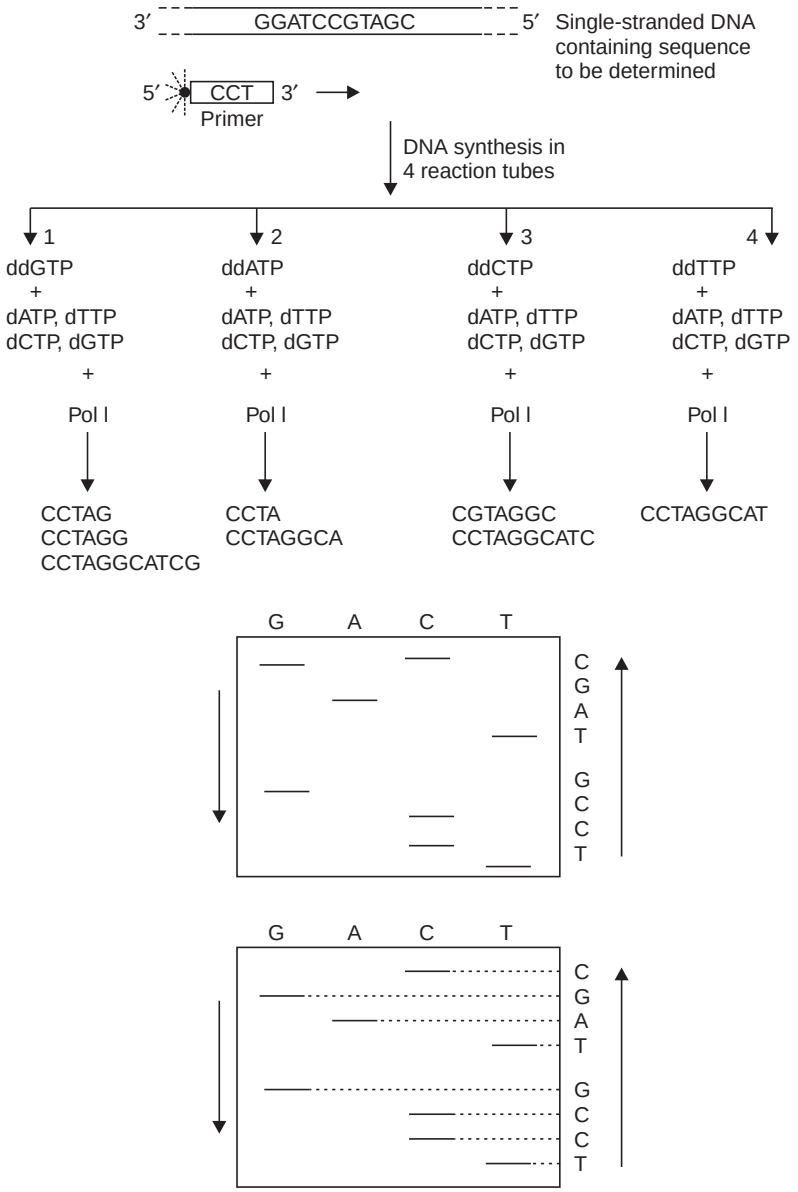


Fig. 23.12 The dideoxy method for determining sequence of a DNA molecule (explanation in text).

Sequences determined by any sequencing method can be entered into databases in the computer. Several computer programs are available that allow further analysis and characterisation of the sequence. Computer programs can be used to analyse DNA sequences for restriction-site location, for comparing a number of sequences, finding homologous regions, transcription regulatory sequences, and more. Programs can also search DNA sequences for possible protein-coding regions by finding a chain initiation codon in a frame with a stop codon. Finding such an open reading frame (ORF) however, does not necessarily mean that the particular DNA sequence encodes a protein *in the cell*, something that can be determined only by doing a number of experiments.

Programs also exist that can translate a cloned DNA sequence into a theoretical amino acid sequence, and to make predictions about the structure and function of the protein. This is possible because the sequences of all *sequenced proteins* have been submitted into the databases, enabling scientists to make rapid comparisons by computer.

QUESTIONS

1. Which technical advances of the last decade were most significant in your opinion for genetic engineering?
2. Comment on the following:
   (*a*) plasmid vector; (*b*) shotgun cloning; (*c*) gene library; (*d*) hybridoma; (*e*) gene machine.
3. Describe the genetic mechanism leading to uptake of recombinant DNA.
4. What are the possible ways of fragmenting eukaryotic DNA, and identifying the fragments?
5. The benefits of genetic engineering are real but the hazards are hypothetical. Explain.
6. Of what use could genetic engineering be to a
   (*a*) pharmaceutical industrialist; (*b*) medical practitioner; (*c*) geneticist?

SELECTED READINGS

Alvarado-Urbina, G. *et al.,* 1981. Automated Synthesis of Gene Fragments. *Science* **214:** 270.

Anderson, W. F. and Diacumakos, E.G. 1981. Genetic Engineering in Mammalian Cells. *Sc. Amer.* **245**: 106.

Beers, R.F. and Bassett, E.G. 1977. Recombinant Molecules. Raven Press, New York.

Boyer, H. W. *et al.,* 1977. The Construction of Molecular Cloning Vehicles. *In* Recombinant Molecules. Eds. Beers, R. F. and Bassett, E.G. Raven Press, New York.

Campbell, A. 1979. Natural Modes of Genetic Exchange and Change. *In* Recombinant DNA and Genetic Experimentation. Eds. Morgan, J. and Whelan, W. J. Pergamon Press, Oxford, New York.

Carbon, J. *et al,* 1977. The Expressison of Yeast DNA in *Escherichia coli. In* Molecular Cloning of Recombinant DNA. Eds. Scott, W. A. and Werner, R. Vol. 13. Academic Press, New York.

Chen, I. and Dubnau, D. 2004. DNA Uptake during Bacterial Transformation. *Nature Revs. Microbiology* **2**: 241–249**.**

Cocking, E. C. *et al.,* 1981. Aspects of Plant Genetic Manipulation. *Nature* **293**: 265.

Cohen, S. N. 1979. Experimental Techniques and Strategies for DNA Cloning. *In* Recombinant DNA and Genetic Experimentation. Eds. Morgan, J. and Whelan, W. J. Pergamon Press, Oxford, New York.

Christensen, A.C. 2001. Bacteriophage Lambda Based Expression Vectors. *Mol. Biotechnol.* **17**: 219–224.

Ehrlich, S. C. and Goze, A. 1979. Expression of Foreign Genes. *In* Recombinant DNA and Genetic Experimentation. Eds. Morgan, J. and Whelan, W.J. Pergamon Press, Oxford, New York.

Gartland, W. J. 1979. Issues in Recombinant DNA Research. *In* Concepts of the Structure and Function of DNA, Chromatin and Chromosomes. Eds. Dion, A. S. Symposia Specialists, Chicago.

Harbers, K. *et al.,* 1981. Microinjection of Cloned Retroviral Genomes into Mouse Zygotes; Integration and Expression in the Animal. *Nature* **293**: 540.

Khorana. H.G. 1979. Total Synthesis of a Gene. *Science* **203**: 614.

McPherson, M.J. and Muller, S.G. 2000. PCR : The Basis. BIOS Scientific Publications, Oxford.

Murray, N.E. and Murray, K. 1979. Application of Recombinant DNA Technology in the Molecular Genetics of Prokaryotes. *In* Recombinant DNA and Genetic Experimentation. Eds. Morgan, J. and Whelan, W.J. Pergamon Press, Oxford, New York.

Pingoud, A. *et al*., 2005. Type II Restriction Endonucleases: Structure and Mechanism. *Cell Mol. Life Sciences* **63**: 685–702.

Riedel, G. *et al,* 1977. The nitrogen fixation (*nif*) operon of *Klebsiella pneumoniae:* cloning *nif* gene and the isolation of *nif* control mutants. *In* Molecular Cloning of Recombinant DNA. Eds. Scott, W. A. and Werner, R. Vol. 13. Academic Press, New York.

Roberts, R.J. 1978. Restriction Endonucleases, A New Role *in vivo. Nature* **271**: 502.

Sambrook, J. and Russel, D. 2001. Molecular Cloning: A Laboratory Manual. 3rd edn. Cold Spring Harbour Laboratory Press, NewYork.

Sherratt, D. 1975. Eukaryotic Genes in *E. coli. Nature* **255**: 523.

Shine, J. *et al.,* 1977. Construction and analysis of recombinant DNA for human chorionic somatomammotropin. *Nature* **270**: 494.

24

Applications of Recombinant DNA Technology

When a gene is cloned it can be used in several different ways. *Site-specific changes* or mutations can be induced in the sequence and the effect of mutations can be studied *in vivo*; a gene can be knocked out of a sequence (*gene knockout*) and phenotypic effects resulting from lack of protein product of the knockout gene can be determined; similar to genes, *polymorphic loci* can be used as markers in mapping experiments and in various other ways; DNA polymorphisms such *as single nucleotide polymorphisms* (SNPs), *short tandem repeats* (STRs) and *variable number of tandem repeats* (VNTRs) are valuable tools in genome mapping; PCR and recombinant DNA technology using *restriction fragment length polymorphisms* (RFLPs) are finding application in human genetic disease, detection of mutations, prenatal diagnosis, newborn screening and detection of heterozygous carriers; DNA *typing* or *fingerprinting* has great importance in distinguishing individuals, determining parentage, based on the concept that no two individuals, with the exception of identical twins, have the same genomic sequence; in molecular biology research this technology is used extensively for determining restriction sites, RNA transcription, translation, regulation, and protein-protein interaction in the cell; treatment of human genetic disorders by replacing an abnormal gene with a normal one, *somatic gene therapy*, is being developed; *genetic engineering* of plants and animals for a wide variety of benefits; the *pharmaceutical industry* is relying heavily on new products based on recombinant DNA technology.

Site-Directed Mutagenesis in DNA

When mutagens are used to induce mutations in an organism, the whole genome becomes a target for the mutagen, and mutagenesis is random. The mutation of interest has to be found by screening and selection methods. If eukaryotic genes that contain mutations in promoter or regulatory sequences are required to be isolated, enormous difficulties would come in the way. *In vitro* techniques have now been developed that allow small portions of a DNA sequence to be deleted, additional nucleotides inserted, and a single specific base can be changed or replaced by another one. The procedure is known as **site-directed mutagenesis** (SDM). The protocol for SDM requires synthesis of a DNA oligonucleotide in which the desired modification is present. This oligonucleotide is allowed to hybridize to normal DNA. The oligonucleotide then becomes the primer for DNA polymerase and normal DNA the template. The polymerase elongates the primer by adding nucleotides that are complementary to the normal DNA. The modified pieces of DNA

can then be cloned and the effect of the mutation can be detected by introducing the DNA into an appropriate host cell. One application of SDM is creation of mutant mice. Mouse models of human mutations serve as valuable tools for investigating gene structure and function in human disease.

Mice have been produced that lack a functional copy of a particular gene. For example, mice that lack a functional copy of the p53 gene develop the malignant phenotype. These are called knockout mice, and they are produced by a series of experimental procedures, details are described later.

DNA Polymorphisms

So far we have been describing genes as markers for genetic analysis. Different alleles of a gene produce different phenotypes which can be detected by making crosses between parents with different alleles of two or more genes. Then by determining recombinants in the progeny, a genetic map can be deduced. These are low resolution genetic maps that contain genes with *observable* phenotypic effects, all mapped to their respective loci. The position of a specific gene, or locus can be found from the map. However, measurements showed that the chromosomal intervals between the mapped genes would contain vast amounts of DNA. These intervals could not be mapped by the recombinant progeny method because there were no markers in those intervening regions. It became necessary to find additional differential markers or genetic differences that fall in the gaps. This need was met by exploitation of various polymorphic DNA markers.

A DNA polymorphism is a DNA sequence variation that is not associated with any observable phenotypic variation, and can exist anywhere in the genome, not necessarily in a gene. Polymorphism means one of two or more alternative forms (alleles) of a chromosomal region that either has a different nucleotide sequence, or it has variable numbers of tandemly repeated nucleotides. Thus, it is a site of heterozygosity (defined in Chapter 1) for any sequence variation. Many DNA polymorphisms are useful for genetic mapping studies, hence they are referred to as **DNA markers**. DNA markers can be detected on Southern blot hybridisation or by PCR. The alleles of DNA markers are codominant, that is they are neither dominant nor recessive as observed in alleles of most genes. DNA polymorphisms constitute molecularly defined differences between individual human beings.

Classes of DNA Polymorphisms: There are three major *classes* of DNA polymorphisms namely, **single nucleotide polymorphisms (SNPs), short tandem repeats (STRs)** and the **variable number tandem repeats (VNTRs).**

Single nucleotide polymorphisms: SNP is a single base pair change, a point mutation, and the site is referred to as SNP locus. SNPs are the most common type of DNA polymorphism, occurring with a frequency of one in 350 base pairs, and accounting for more than 90 per cent of DNA sequence variation. The majority of SNPs are found to be present in the non-coding regions of the genome, known as non-coding SNPs. SNPs in the coding regions, that is within genes, are known as coding SNPs (cSNPs). Detailed studies of cSNPs in humans indicate that each gene has about four cSNPs, half of which resulting in missense mutations in the encoded protein, and half of which produce silent mutations. Whether a cSNP affects a phenotype, depends on the amino acid that is changed by the polymorphism. About one-half of missense mutations that are SNPs are estimated to cause genetic disease in humans. A non-coding SNP can also affect gene function if it is located in the promoter region or in the gene regulatory region. A small number of SNPs can create a restriction site, or eliminate an already existing restriction site. SNP-induced alterations in restriction sites are detected by using the restriction enzyme followed by Southern blot analysis or PCR.

An individual SNP locus can be analysed by using the technique of allele-specific oligonucleotide (ASO) hybridisation. The search for one particular SNP locus in humans is a challenge, because this is one base pair that is polymorphic out of the three billion base pairs in the human genome. In the ASO technique, a short oligonucleotide that is complementary to one SNP allele is synthesised and mixed with the target DNA. Hybridisation is performed under high stringency conditions that would allow only a perfect match between probe and the target DNA. That means, the oligonucleotide will not hybridize with target DNA that has any other SNP allele at that locus. Positive result of hybridisation indicates the SNP locus precisely.

A more recent technique of **DNA Microarrays** can be used for simultaneous typing of hundreds or thousands of SNPs. Details of this technique used for SNPs and genome wide gene expression are described later in this section.

A small number of SNPs can lead to changes in restriction sites either by creating a restriction site or eliminating one. Such SNPs can be detected by using the restriction enzyme for the site, and detection is done by Southern blot analysis or PCR. The different patterns of restriction sites in different genomes yield fragments of different lengths, called **restriction fragment length polymorphisms (RFLPs)** described below.

Restriction Fragment Length Polymorphisms: RFLPs are restriction enzyme *recognition sites* that are present in some genomes and absent in others. Consider an organism heterozygous for an RFLP whose genotype we represent as Rr. This organism is backcrossed with another that is homozygous for the RFLP variation allele (rr). Genomic DNA from the progeny of this cross (Rr x rr gives progeny of which 50% is Rr and 50% is rr) is subjected to restriction enzyme digestion, and fragments separated on Southern blots. The restriction fragments obtained are hybridised with a probe (a cloned DNA fragment) that will distinguish the various genotypes for an RFLP. The probe DNA is unique because it comes from only one DNA segment of the genome and that overlaps the restriction site. A key point of this technique, therefore, is the use of a specific cloned single-copy DNA probe that is specific for an individual marker locus. Crosses between the positive RFLP organism with other RFLP bearing organisms would yield parental combinations and recombinations. From the frequency of recombinants, a detailed RFLP map can be produced. RFLPs were the first DNA markers that were in use for characterisation of plant and animal genomes. They have now been replaced by markers based on variation in the number of short tandem repeats (STRs) described below.

Short tandem Repeats: STRs are also known as **microsatellites** and **simple sequence repeats (SSRs)**. A tandem repeat is a sequence that is repeated end to end in the same orientation. STRs are 2 to 6 base pair DNA sequences tandemly repeated a few times. For example, the sequence TCACATCACATCACATCACATCACA is a five-fold repeat of the sequence TCACA. There are dinucleotide, trinucleotide, four-nucleotide, five-nucleotide and six-nucleotide STRs in the human genome. Microsatellite analysis can be done using a single-copy DNA to serve as a PCR primer pair specific for each marker locus. In contrast with RFLPs that have only one or two alleles in a population, STRs have a much larger number of alleles which can be detected in a population analysis. Consequently, STRs have a higher proportion of heterozygotes which makes them more suitable for mapping purposes. Polymorphisms in STRs is common in populations which makes them valuable tools in genetic mapping.

Variable Number Tandem Repeats: **VNTRs**, also called **minisatellite markers**, the repeat unit is a little larger than in STRs, from seven to a few tens of base pairs long. The VNTR loci in humans are 1 to 5 kilobase sequences containing repeat units about 15 to 100 nucleotides long. VNTR loci also show polymorphisms. Due to the greater length of VNTR repeats that makes PCR unsuitable, analysis of VNTRs relies on restriction digestion and Southern blotting. The entire genomic DNA is cut with a restriction enzyme which cuts on either side of the VNTR locus,

but does not have a target site *within* the VNTR arrays, followed by Southern blotting. The VNTR specific probe against a particular repeat sequence of the VNTR locus, will bind at *all locations* of the repeat sequence *in the genome*, resulting in a large number of different sized fragments. The number of tandem repeats is variable from one individual to the other, therefore Southern blot provides a distinct distinguishing pattern of fragments for a single individual. These patterns are also referred to as **DNA fingerprints**. The technique finds useful application in identification of individuals and in deciding parentage.

Microsatellite Markers: Variable numbers of *dinucleotides* repeated in tandem, called **microsatellite markers**, are dispersed in the genome. The most common type are CA and the complementary GT repeats. Probes are designed for detection of DNA regions surrounding *individual* microsatellite repeats by using PCR. The procedure is explained by taking the example of human DNA as follows. Human genomic DNA is subjected to restriction digestion by an enzyme such as *AluI*, that will result in fragments about 400 base pairs in length. The fragments are cloned into a vector and Southern blotting is carried out. To identify genomic inserts that contain CA / GT dinucleotides, probes specific for these dinucleotides are used. Sequence of the positive clones is determined, on the basis of which PCR primers are designed that will hybridise with *single-copy* DNA sequences *flanking* the specific tandemly repeated microsatellite sequences. PCR amplification is carried out using these primer pairs and genomic DNA. Thus, if any size variation exists in the stretch of tandemly repeated microsatellite sequence, it would be detected through gel electrophoresis of the DNAs from different individuals. The size variations may differ among the different individuals, all these variations could be determined. A size variation results in amplification product of a different size and represents a marker allele.

Randomly Amplified Polymorphic DNA: RAPDs are based on random PCR amplification. The procedure is carried out by randomly designing primers for PCR which will amplify several different regions of the genome by chance. Such a primer results in amplification of only those DNA regions that have near them, inverted copies of the primer's own sequence. The PCR products consist of DNA bands representing different sizes of the amplified DNA. The set of amplified DNA fragments is called **randomly amplified polymorphic DNA (RAPD)**. Certain bands may be unique for an individual and can serve as DNA markers in mapping analysis.

DNA Microarrays

The DNA Microarray technology is used to determine the level of expression of many thousands of genes simultaneously. This new approach is used not for individual genetic loci, rather, for the analysis of genome-wide patterns of gene expression. Using DNA microarrays, it is possible to estimate the relative level of gene expression of each gene in the genome.

The **DNA microarray** or **chip** is a high density grid system, consisting of a flat solid substrate about the size of a postage stamp that can be used to detect hybridisation of target DNA under appropriate conditions. The chip contains 10,000 to 100,000 distinct spots, from 75 to 150 μm in diameter. The spacing between spots on an array is usually 100 to 200 μm. Each spot contains a different immobilised DNA sequence that can be hybridised with DNA (or RNA) from a large number of different cells. Two types of chips are currently available: one, in which oligonucleotides have been synthesised *directly* on the chip, one nucleotide at a time, by automated procedures. These chips have hundreds of thousands of spots per array; second, chips in which double-stranded DNA sequences of 500 to 5000 base pairs have been deposited through drops by capillary action from miniaturised devices mounted on the movable head of a robotic workstation. These chips have tens of thousands of spots per array. The surface onto which DNA is spotted is critically important. The ideal surface immobilises the target DNAs, and is compatible with stringent probe hybridisation conditions.

The procedure shown (Fig. 24.1) depicts only 6 spots in a chip, each of which contains a DNA sequence that serves as a probe for a different gene. Experimental cells are used for the

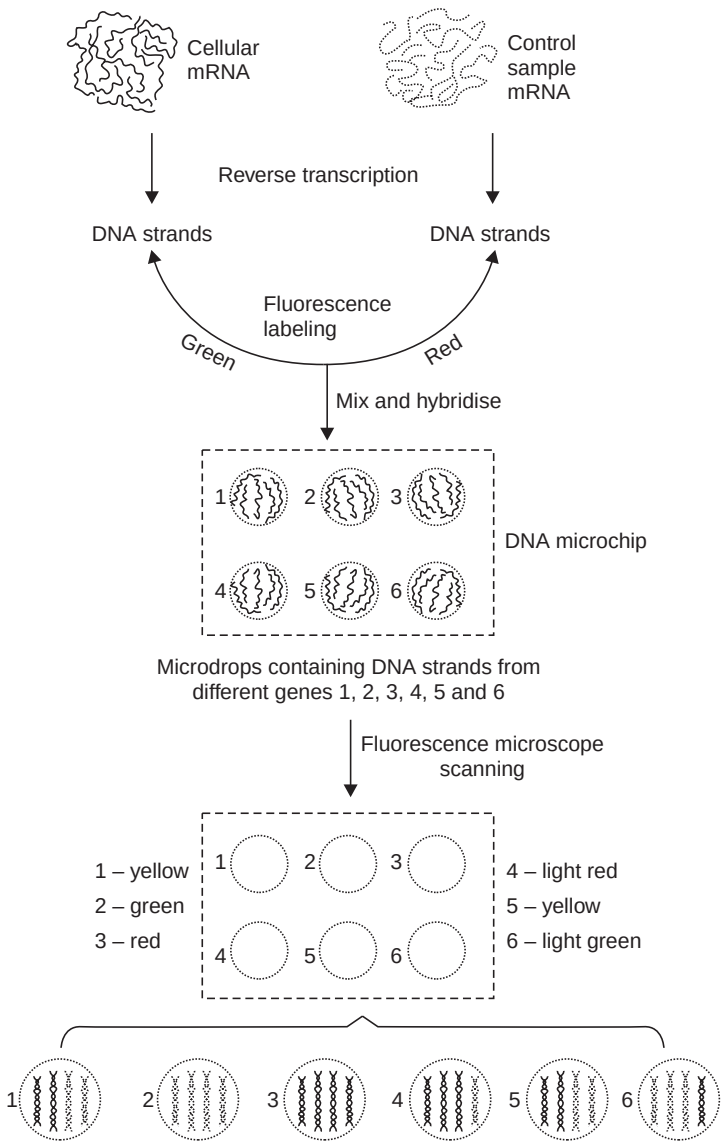


Fig. 24.1 Procedure for DNA microarrays. Six dried microdrops are introduced into a DNA microchip. Each drop contains immobilised DNA strands from a different gene numbered 1 to 6. These are hybridised with fluorescence-labelled DNA samples obtained by reverse transcription of cellular mRNA (green) and red labelled control DNA sample. Competitive hybridisation of green (experimental) and red (control) label is proportional to the *relative* abundance of each mRNA in the sample. The intensity of red and green fluorescence is analysed by microscopy and interpreted as overexpression, underexpression and equal expression of the gene depending on the intensity of red, green, orange, yellow-green and yellow fluorescence.

extraction of cellular *m*RNA, and a control sample of *m*RNA from another source. The samples are subjected to reverse transcription to obtain DNA strands. In the experimental material, the primer for reverse transcription is tagged with a green fluorescent label, while primers of the control material receive red fluorescent label. After the DNA strands have been obtained in sufficient quantity, the fluorescent samples are mixed and hybridised with the DNA in the spots in the chip. The hybridisation is competitive because the two samples were mixed. Therefore, the density of red and green strands bound to the chip is proportional to the concentration of red or green molecules in the mixture. Genes that are overexpressed in the experimental sample relative to the control will have more green strands hybridised to the spot, whereas those that are underexpressed in the experimental sample relative to the control will have more of red strands hybridised to the spot.

The intensity of fluorescence is viewed by placing the chip under a laser scanning microscope or a fluorescence microscope that scans each pixel, which is the smallest discrete unit in a visual image. The intensity of fluorescent label is recorded. The signals are synthesised to produce a signal value for each spot in the microarray. The signals indicate the relative levels of gene expression through colour. Green or yellow green indicate overexpression in experimental sample, while red or orange indicates underexpression in experimental sample. Yellow indicates equal expression in both experimental and control samples.

DNA microarray technology is useful for study of large number of cells growing under different conditions, at different developmental stages, or at different stages of a disease. Besides detection of gene expression, this technology can be used to detect mutations and polymorphisms, to map genomic DNA clones, and to compare the gene expression pattern in normal and diseased tissues.

Gene Transfer into Eukaryotic Cells

Recombinant DNA has opened new avenues for the study of gene function. The function of a cloned gene can be investigated by introducing the cloned DNA into eukaryotic cells. In simple eukaryotes such as yeast*s Saccharomyces cerevisiae*, molecular clones of various mutants have been prepared for genetic analysis. Like bacteria, yeasts offer advantages in having a small genome of about 1.2×10^7 bp and a rapid rate of reproduction. To clone a mutant gene in yeast, first a genomic library of normal yeast DNA is prepared in plasmid vectors that can replicate in yeasts and in bacteria. A mixture of such plasmids is used to transform a **temperature-sensitive mutant** of yeast. Such mutants encode proteins that are functional at one temperature, called the *permissive temperature*, but non-functional at another temperature, called the *nonpermissive temperature*. Transformed yeast cells are easy to select because they grow at the nonpermissive temperature, as they have aquired a normal copy of the gene from plasmid DNA. A number of yeast proteins have been characterised in this manner.

Genetic manipulation techniques that are possible in simple eukaryotic cells as those of yeasts are not applicable to cells of complex eukarytoes. Methods are however available for introduction of cloned DNA into cells of plants and animals, referred to as **gene transfer**. One common approach is to incorporate eukaryotic DNA into the genome of a nonreplicating virus, followed by infection of a host cell. Several viruses can infect a cell and integrate their genome into the DNA of host cell. If the viral genome contains an insert of foreign DNA, that would also be integrated into host cell genome. This technique of gene transfer is called **transduction**.

Methodologies are now available to introduce *naked* DNA into animal cells in culture as a coprecipitate with calcium phosphate. The process is called **transfection** (derived from *transformation* + *infection*). The cells in suspension exposed to a fine calcium precipitate of the DNA take up DNA and transport it to the nucleus, where it can be transcribed for several days, referred to as *transient expression*. A very small fraction of cells, that is less than 0.1%, are competent to be transfected, in which the foreign DNA becomes stably integrated into the cell's genome. Transfected cells generally pick up several fragments of DNA. The stably transformed cells that have integrated foreign DNA can be selected by including a gene, such as a gene for drug resistance that allows transfected cells to grow, and inhibits growth of nontransfected cells. Thus, any cloned gene can be introduced into animal cells by being transferred together with a drug resistance marker that allows stable transformants to be isolated. Transfected cells can be used to address a wide variety of questions concerning cell growth and differentiation, mechanisms that regulate gene expression and protein function.

Depending on the purpose of the experiment, various other methods are available to introduce DNA into mammalian cells. Transfection of both plant and animal cells can be carried out by **electroporation**. The cells are incubated with DNA in vials that contain electrodes that deliver a brief electric shock. The electric pulse opens pores in the plasma membrane, called electroporation, that makes the plasma membrane transiently permeable to DNA molecules, some of which find their way into the nucleus and become integrated into the chromosomes, producing stably transformed cells. Factors that influence the efficiency of electroporation include temperature, some electric field parameters such as voltage, resistance and capacitance, the topological form of DNA and some host cell factors (growth conditions and genetic background).

DNA can also be incorporated into lipid vesicles called liposomes that fuse with the plasma membrane, and deliver DNA inside the cell. Direct microinjection of DNA can also be done into the cell nucleus.

A direct way to introduce foreign genes into a cell is to microinject DNA directly into the nucleus. The eggs of amphibians such as *Xenopus* are particularly suitable, and have been used in classical studies on giant chromosomes called lampbrush chromosomes in which transcriptional activity can be visualised. When foreign DNA is injected into the nucleus it is readily transcribed. The RNA transcripts from the foreign DNA templates are transported to the cytoplasm, and translated into proteins that can be detected by immunological methods using specific antibodies.

Transgenic Animals: Cloned genes can be introduced into the ***germ line*** cells of an animal allowing study of the intact animal rather than in cultured cells. For this purpose, the nucleus in mouse embryonal cells has been used for injecting foreign DNA. The goal in such experiments is to have the foreign DNA integrated into the egg's chromosomes. Cell divisions in the egg will pass on the foreign DNA to all cells of the embryo and subsequently to the adult animal. Mice that have been genetically engineered to carry such foreign genes in their chromosomes are called **transgenic mice**. To generate transgenic mice, cloned DNA is microinjected into a pronucleus of the fertilised egg (Fig. 24.2). The eggs are then transferred to foster mothers and allowed to develop to term. Among the progeny mice, about 10% will have the foreign DNA integrated into the genome of the fertilised egg, and therefore, in all cells of the adult mice. Since the foreign DNA is also present in the germ line cells of the progeny mice, these mice are mated to breed new progeny mice which would inherit the foreign DNA.

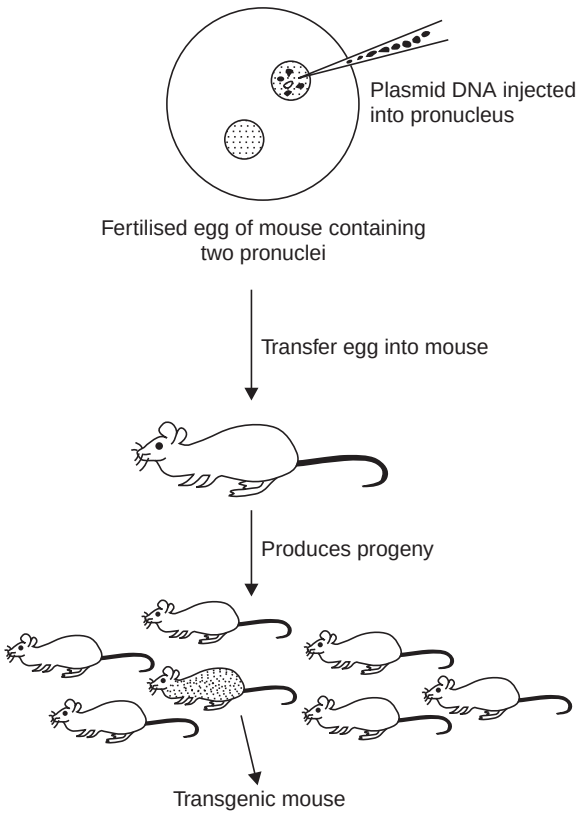


FIg. 24.2 Production of transgenic mice (explanation in text).

Transfection by Embryonal Stem Cells

Cloned genes can be introduced into mice by means of **embryonal stem cells (ES cells)**. Embryonal cells have attracted tremendous attention in recent years, hence a brief introduction on ES cells is warranted. The concept of ES cells arose through studies on mouse and human teratocarcinomas. These tumours were found to contain cells, called *embryonal carcinoma cells*, the stem cells of the tumour that were able to differentiate into a variety of adult cells and tissues. Stem cells can proliferate without changing their phenotype and simultaneously change (differentiate) into one or more new cell types. ES cells were then derived from cells in the inner cell mass of the mouse blastocyst. ES cells can be established in culture. They can also be reintroduced into early embryos where they take part in development and can give rise to cells in all tissues of the mouse, including cells of the germ line. Thus cloned DNA can be introduced into ES cells in culture, the transformed cells can be selected. These cells can be introduced back into mouse embryos. Such embryos produce chimeric offspring in which some cells are derived from the normal embryo cells, and some from the transfected ES cells. If some such mice have transfected cells in their germ line, breeding these mice allows direct inheritance of the transfected gene in the progeny.

Knockout Mice

The ability of embryonic stem cells to change (differentiate) into any cell type of the adult organism clearly demonstrates that they are **totipotent**. In various kinds of studies we come across

phenotypes in mice that lack a functional copy of a particular gene. These animals are called knockout mice and serve as a research tool for gaining insights into mechanisms underlying various cellular activities in which the product of a particular gene is directly involved, as well as in genetic basis of a human disease. Knockout mice in which a particular gene is inactivated (**gene knockout**) can be produced in the laboratory by using **embryonic stem cells** (Fig. 24.3).

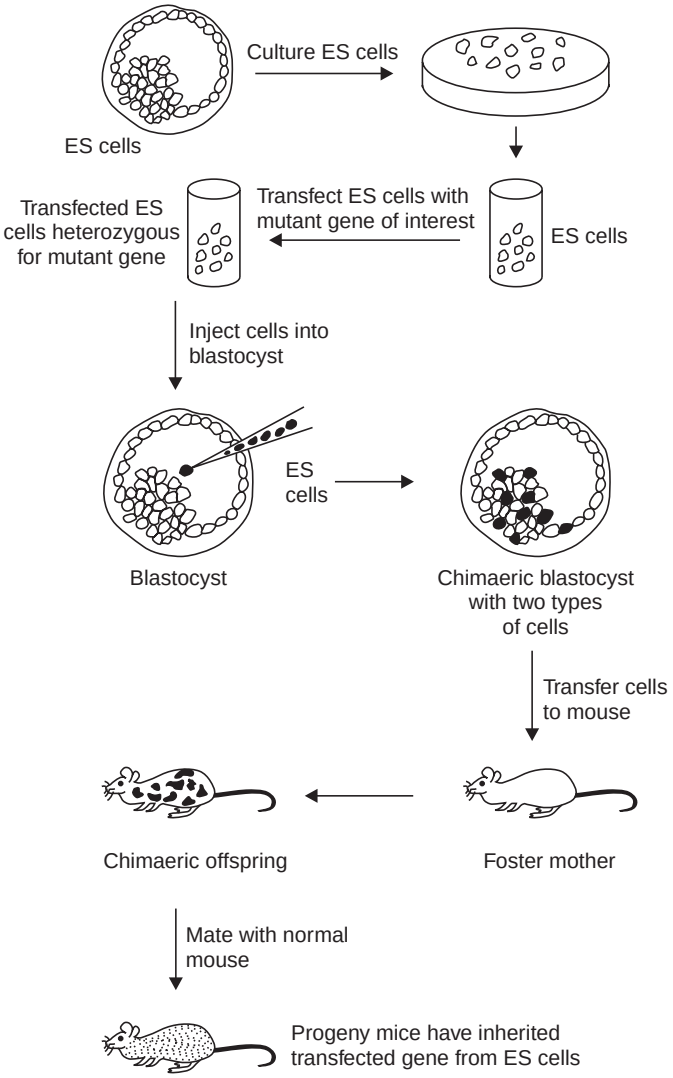


Fig. 24.3 Introduction of genes into mice using embryonal stem cells.

Embryonic stem cells (ES cells) are present in the *early blastocyst* stage of mammalian embryos. The layer of cells lining the blastocyst, called trophoectoderm, contains on its inner surface a cluster of cells referred to as the **inner cell mass** that projects into a cavity called blastocoel. The inner cell mass contains the embryonic stem cells which differentiate into all the cell types present in all tissues of the adult animal. ES cells can be isolated from the blastocyst and established in culture where they proliferate. A DNA fragment that contains a non-functional mutant allele of the gene to be knocked out is transfected into ES cells. A gene for antibiotic

resistance is also incorporated in the fragment to allow for selection of cells that have incorporated the inactivated gene into their genome. Among the ES cells that take up the DNA fragment, about one cell in 10^4 cells undergoes the process of homologous recombination by which the transfecting DNA sequence replaces homologous DNA. In this way, ES cells that have aquired the tranfecting gene become heterozygous for this particular gene. That is because the homologous sequence in identical location in the other chromosome is carrying the normal allele of this gene. ES cells are grown on medium and those with the gene of interest are selected on the basis of antibiotic resistance.

The selected ES cells (donor cells) are then injected into the blastcoel of a recipient mouse embryo. The recipient embryo is implanted into a female mouse, pretreated hormonally to carry the embryo to term. The injected ES cells in the surrogate mother join the embryo's own inner cell mass, and as the embryo develops, these ES cells become part of some embryonic tissues including the reproductive gonads. The progeny mice are found to be **chimaeric** because they display characteristics of both donor and recipient. The word chimaera is derived from the mythological creature *Chimaera* with the head of a lion, body of a goat and tail of a serpent. To find out whether or not the germ cells contain the knockout gene, the chimeric mice are mated to normal mice. If the germ cells contain the knockout gene, the progeny mice will be heterozygous for this gene. Two heterozygous mice are then mated to one another to obtain some mice that are homozygotes, according to mendelian inheritance, and are called knockout mice.

Monoclonal Antibodies

Specific **monoclonal antibodies,** that is, identical descendents of clones of a single cell, can be produced against foreign substances using the **Hybridoma technique.** The antigen is injected into a mouse (Fig. 24.4). After a few days, spleen of the mouse containing lymphocytes which produce antibodies, is removed. These antibody producing cells are fused with cancerous proliferating mouse myeloma cells in the presence of a fusogen, polyethylene glycol. The product of fusion called **hybridoma,** is capable of repeated division, a property which it inherits from the dividing myeloma cells, and produces antibodies. The hybridoma cells are allowed to grow and individual cells are tested for the required antibody and cloned. The selected clone is cultured for making large amounts of the desired antibody. The purified antibodies are monoclonal in origin, hence they are highly specific for particular targets.

Human Gene Therapy

Soon after recombinant DNA techniques became available, scientists turned their attention to their therapeutic use, for correcting defective genes responsible for disease in humans. The following approaches have been tried for correcting abnormal genes.

Replacement of a nonfunctional gene by inserting a normal gene into a *nonspecific* location in the genome; removal and replacement of an abnormal gene itself by a normal gene through homologous recombination; repair of the defective gene through selective reverse mutation, that would revert the gene to its normal function; altering the regulation of a particular gene (control of gene expression).

The most commonly used approach is that of inserting a normal gene to replace a disease-causing gene. This can be achieved by using a vector to deliver the therapeutic gene to the patient's cells. A virus that has been genetically altered to carry normal human DNA is used. This is based on the knowledge that viruses have evolved a way of becoming pathogenic for humans, by encapsulating and delivering their genes to human cells. Scientists have exploited this capability of the virus by manipulating the viral genome to remove the disease-causing genes and inserting therapeutic genes, then using the virus as vector.

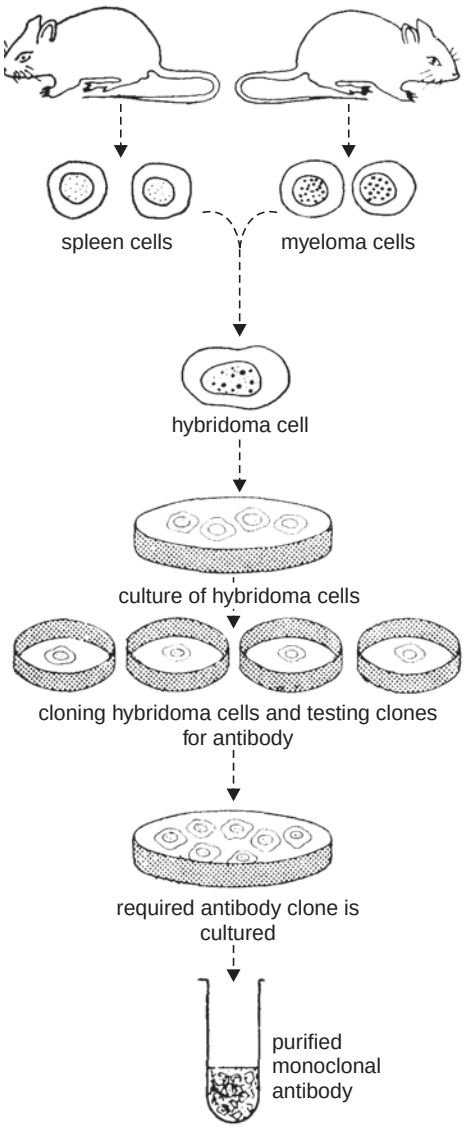


Fig. 24.4 The hybridoma technique for illustrating the formation of monoclonal antibodies.

The virus vector is allowed to infect target cells in the patient. The vector transfers its genetic material that contains the therapeutic human gene, into the target cells. Expression of the therapeutic gene results in a functional protein product, restoring the target cell to its normal state. Generally retroviruses, adenovirus and herpes simplex virus are used as vectors in gene therapy.

Current Status of Gene Therapy Research: Gene therapy is still in its infancy, at the experimental stage. By and large it has not proved to be successful in clinical trials. The first trials began in 1990, and not much progress has been reported. So far, none of the gene therapy products have been approved for sale by the regulating authorities. Factors such as short-lived

nature of gene therapy, the risk of stimulating immune response, toxicity and possible adverse effects of viral vectors have kept gene therapy from becoming an effective form of treatment for genetic diseases.

Transgenic Plants

Genetic engineering of plants is carried out by introducing DNA into a cell in culture that can grow into a mature plant. An efficient vector for introduction of recombinant DNA into plant cells has been developed from plant viruses. Furthermore, the *Ti plasmid* isolated from the bacterium *Agrobacterium tumefaciens* also serves as a vector for inserting foreign DNA.

Agrobacterium tumefaciens is a gram negative, soil bacterium and a plant pathogen that induces tumour-like growths on plants called crown gall tumours. Gene transfer from the bacterium to the plant occurs naturally, resulting in tumours. Tumours can also be induced in gymnosperms and dicotyledonous angiosperms by inoculation of wound sites with *A. tumefaciens*. Evidence suggests that crown gall tissue represents true oncogenic transformation because the undifferentiated cell mass of the tumour (callus) can be cultured *in vitro* even if bacteria are killed by antibiotics, still retaining its tumour-like properties. These properties include unlimited growth as a callus, and synthesis of opines, such as octopine and nopaline which are unusual amino acid derivatives not present in normal plant tissue. The metabolism of opines is a central feature of crown gall disease. Plant cells acquire the property of opine synthesis when they are colonised by *A. tumefaciens*. The bacterium utilises opine as its sole source of nitrogen and carbon.

The virulent strains of *A. tumefaciens* contain a *Ti* plasmid that confers tumour-inducing properties on the bacterium. Earlier investigators on crown gall tumours had observed that continued presence of *Agrobacterium* is not required to maintain plant cells in their transformed state. In 1974 Zaenen and colleagues demonstrated presence of large-sized plasmids, of 140 to 235 kilo base pairs length in virulent strains of *A. tumefaciens*. Further experiments revealed that virulence and ability to use as well as induce synthesis of opines are plasmid-borne traits. These plasmids were known as tumour-inducing plasmids (*Ti* plasmids). *Ti* plasmids also contain information about the specific type of opine that is synthesised in the transformed tissue and utilised by the bacterium.

Furthermore, the complete *Ti* plasmid is not found in the plant tumour cells. A specific segment of the plasmid, about 23 base pair in size is found integrated in plant nuclear DNA at a random site. This segment of DNA transferred from the plasmid is called **T-DNA** (transferred DNA). *T*-DNA carries genes responsible for conferring unlimited growth and ability to synthesise opines upon the transformed plant tissue. The genes responsible for *T*-DNA transfer are located in a separate part of the *Ti* plasmid called *vir* (virulence) segment. Two of these genes (*vir*A and *vir*G) are expressed constitutively at a low level of control. The *vir* gene expresses a protein that forms a conjugative plasmid through which *T*-DNA is transferred to the nucleus. Once inside the nucleus, *T*-DNA is incorporated randomly into nuclear DNA by a process of illegitimate recombination. In addition to plants, *Agrobacterium* can transfer DNA to other bacteria, yeasts and filamentous fungi.

Thus, *Ti* plasmid serves as a natural vector in genetic engineering of plant cells because it can transfer its *T*-DNA from the bacterium to the plant genome. The wild type *Ti* plasmids, however, are not suitable as vectors because of the presence of oncogenes in *T*-DNA which result in disorganised growth in recipient plant cells. To accomplish efficient regeneration in plants, attempts have been made to delete all of the oncogenes from the plasmid. Indeed, when *Agrobacterium* carrying nononcogenic plasmids were allowed to transfer the modified *T*-DNA to plant cells, no tumours were produced.

Transgenic Plants from Transformed Protoplasts

Somewhat parallel to transfection in animal cells, protoplast transformation has been achieved in plants. Plant protoplasts from which the rigid cell wall has been removed, are induced to take up DNA. A number of chemicals, in particular polyethylene glycol promote gene transfer across the protoplast membrane. Alternatively, DNA uptake can be induced by electroporation. In a small proportion of the protoplasts, the DNA is incorporated stably into the genome (transformed protoplasts). The first successful experiments on plant regeneration using protoplasts were carried out on tobacco and petunia, subsequently in monocots such as *Lolium*.

Gene Transfer in Plants by Particle Bombardment

In contrast with animal cells that are bounded by a unit membrane, plant cells are distinguished by presence of a thick, often multilayered cell wall. A technique involving particle bombardment was developed in 1987 for gene transfer in a wide range of plant species. A modified shotgun is used to accelerate small metal particles (1 to 4 μm) into plant cells at a velocity sufficient to penetrate the cell wall. In the first experiments, onion skin cells were bombarded with tungsten particles coated in tabocco mosaic virus (TMV) RNA. About 40% of onion cells were found to contain particles three days after bombardment. These cells also showed evidence of TMV replication. Subsequently, explants from several plant species have been stably transformed by this technique.

SELECTED READINGS

Brinster, R. L. 2002. Germline Stem Cell Transplantation and Transgenesis. *Science* **296**: 2174–2176.

Chee, M. *et al*., 1996. Accessing Genetic Information with High Density Arrays. *Science* **274**: 610–614.

Gelvin, S.B. 2000. *Agrobacterium* and Plant Genes Involved in T-DNA Transfer and Integration. *Annu. Rev. Plant Physiol*. **51** : 223–256.

Maliga, P. 2004. Plastid Transformation in Higher Plants. *Annu. Rev. Plant Biol*. **55** : 289–313.

Solter, D. 2005. What is a Stem Cell? In: Stem Cells. John Wiley, New Jersey.

Taylor, N.J. and Fauquet, C.M. 2002. Microparticle Bombardment as a Tool in Plant Science and Agriculture Biotechnology. *DNA Cell Biol*. **21** : 963–977.

25

Genomics

Recombinant DNA technology has been extended to studies of whole genomes by mapping and sequencing techniques. **Genomics** is, therefore, the molecular mapping and characterisation of whole genomes and their gene products. The small genome of the bacteriophage $\phi X174$ was the first one to be sequenced, followed by plasmid and viral genomes. Subsequently genomes of eukaryotes such as yeast, *Drosophila*, rice plant and chimpanzee were sequenced. The complete sequence of the circular genome in the human mitochondrion (16159 bp) was determined in 1981.

The most recent achievement is the successful completion of the **human genome project** (HGP) in 2003. The project had aimed to identify all of the approximately 20,000 to 25,000 genes in human DNA; determine the sequences of the three billion base pairs that comprise the human genome; information retrieved from the project to be stored in databases; develop improved tools for analysis of data; transfer related technologies to the private sector; and to address the ethical, legal and social issues (ELSI) pertaining to this project. Among the special experimental techniques that were devised include those for handling and sequencing millions of clones. Hi-tech automation for decoding information and computer-driven robotics were employed. Bioinformatics played a central role in data analysis. The project was coordinated jointly by the U.S. Department of Energy (DOE) and National Institutes of Health (NIH) involving cooperative efforts of many scientists and engineers from several countries. Though HGP is finished, analysis of data will continue for many years.

Functional genomics refers to analysis of the function of genes through expression of gene products, as well as the nongene sequences in the entire genome. It includes study of the control of gene regulation, interactions, regulation of RNAs and proteins. Gene expression can be analysed through study of the all the expressed *m*RNA transcripts, known as the **transcriptome**, and through **polypeptides** called the **proteome**.

The earlier attempts in making genetic maps were based on recombination frequencies. Using suitable organisms, such as yeast and *Drosophila*, large-scale crosses were made between different mutant strains. The methodology relied on the concept that the further apart in map distance, two loci are on a chromosome, greater is the chance that a crossover will occur between them during meiosis. Recombinants in the offsprings resulting from crossovers were scored and map distances were recorded (details in Chapter 8). Thus *low resolution* chromosomal maps of

genes producing known mutant phenotypes were constructed. This approach cannot be used for mapping the human genome because of its large size (3000 Mb or 3 billion base pairs), and inability to set up large-scale matings between people carrying different inherited diseases. Instead, human genetic mapping can be carried out through analysis of DNA sequence polymorphisms in the population, that are naturally occurring DNA sequence differences which do not produce visible phenotypes. We describe here the various approaches and technologies which have resulted in mapping the human genome.

Whole genome mapping is approached in two ways: by **mapping** or **shotgun cloning.**

Mapping the whole genome involves the development of high resolution genetic and physical maps in order to generate DNA segments of increasing resolution, and then to determine the sequence of the fragments. The genetic maps can be based on the order of markers by meiotic recombination, or by co-localisation of genes in individual fragments of chromosomes. Physical maps provide a view of how the clones from genomic clone libraries are distributed throughout the genome. The other approach uses whole genome **shotgun cloning** in which the genome is broken up into random overlapping fragments. Then to sequence the fragments and assemble the segments using computational methods.

The Mapping Approach for Sequencing Human Genome

The initial goal of preparing a genetic and physical map is to obtain closely spaced markers throughout the genome. The Human Genome Project (HGP) started out with the aim of constructing a high density genetic map with at least one genetic marker per one Mb (about one map unit) of the genome. The procedure starts by developing low resolution genetic maps using recombinational mapping of inherited differences, or cytogenetic mapping methods. Prepare physical maps indicating positions of individual cloned DNA fragments of each chromosome. These genetic and physical maps are then integrated with molecular maps at higher resolution. A dense array of markers is obtained that can be used directly in gene cloning. Thereafter, conduct large scale genomic DNA sequence analysis to produce a complete sequence map of each chromosome. The idea of *progressively increasing resolution* of analysis is a central feature in mapping the genome.

Meiotic Recombination and Cytogenetic Maps

Meiotic linkage mapping is carried out as already described in Chapter 8, in organisms such as yeasts, fungi (tetrad analysis), *Drosophila* and some plants. These methods are not usable in case of humans owing to lack of information from crosses; small size of progeny does not allow accurate determination of linkage; and very large size of human genome with 24 linkage groups (22 autosomal and 2 sex chromosomes). In humans **pedigree analysis** is one of the approaches used to determine linkage relationships between human genes, including pedigree analysis for determining recombination frequency, as well as molecular methods.

Cytogenetic Maps: High resolution **cytogenetic maps** can be developed by determining locations of DNA markers in relation to visible chromosome bands, puffs and positions of centromeres. One of the methods known as ***in situ*** **hybridisation** was developed in 1970 by Pardue and Gall for locating repeated DNA sequences in eukaryotic chromosomes. In this technique chromosomes are spread on a glass slide and *denatured* by appropriate treatment. A cloned DNA sequence can be used to make a labeled probe for hybridisation to chromosomes. A labelled denatured

probe will hybridise to homologous sequences in single strands of chromosomes *in situ*. Besides landmarks such as positions of bands and centromeres, the short and long arms (denoted *p* and *q* arms) of eukaryotic chromosomes have been subdivided into distinct segments that are numbered consecutively, so that numbers indicate distances from centromeres and telomeres. Hence the positions of binding of labeled probes can be mapped precisely. The resolving power of this technique is not sufficient to distinguish between two genes that are about 5 centiMorgans (5 cM) apart.

The cloned DNA probes used in *in situ* hybridisation are labelled radioactively or by fluorochromes. When the probe is radiolabelled, the positions in denatured chromosomes at which the probe is hybridised are determined by autoradiography. That is, the chromosome preparations on a glass slide are covered with a film or a thin layer of liquid emulsion containing silver bromide (AgBr) crystals, kept in dark for 2 to 3 weeks to allow radioactive emissions from the probe to reduce the AgBr. Developing the film washes out unreduced grains and shows dark spots where the probe has hybridised.

When the probe is labeled using a fluorescent dye, the procedure is called **fluorescent in situ hybridisation (FISH)**. The cloned DNA is labelled with a fluorescent dye and the chromosome preparation is immersed in the dye containing the probe. When the probe has hybridised to the chromosomes, the slide is scanned in a fluorescence or a laser confocal microscope, and images of fluorescent spots are recorded.

Another procedure that is a variation of FISH is called **chromosome painting**. This technique uses a standard control set of probes that are homologous to known locations in order to prepare a cytogenetic map. Sets of cloned DNA sequences that are known to be from specific chromosomes or from specific chromosome regions are used as probes. Each set of cloned DNA is labeled with a *different* fluorescent dye. When the probes have hybridised, each with its homologous region in denatured chromosomes, the fluorescent dyes "paint" specific regions, they are identified microscopically on the basis of colors. This procedure can include a probe consisting of a cloned sequence of unknown location, labelled with another dye. Its position would be indicated by its dye.

Radiation Hybrid Mapping

The technique of **radiation hybrid mapping** is used to develop a high resolution map of molecular markers along a chromosome. It is based on the principle that when cells from two cell lines in culture, one from human the other from rodent, are fused with one another, the resulting hybrid cells retain only a few human chromosomes, the remainder are eliminated during cell divisions. The human chromosomes that are retained are inherited by descendents of the hybrid cell, and represent a *clone* of cells. The selection of human chromosomes that would be retained in the hybrid cell seems to occur in a random manner.

For mapping studies in radiation hybrids, instead of whole chromosomes, each hybrid cell line contains a random set of human chromosome fragments (Fig. 25.1). Fragmentation is accomplished by irradiating human cells in culture with a lethal dose of X-rays (3000 rads). The irradiated cells are then fused with cultured rodent cells to form somatic hybrid cells whose nuclei are also fused (heterokaryons). These are cloned in wells of tissue culture plate, a *series* of clones is obtained, each containing a different random assortment of human chromosome fragments. From these a radiation hybrid mapping panel can be made.

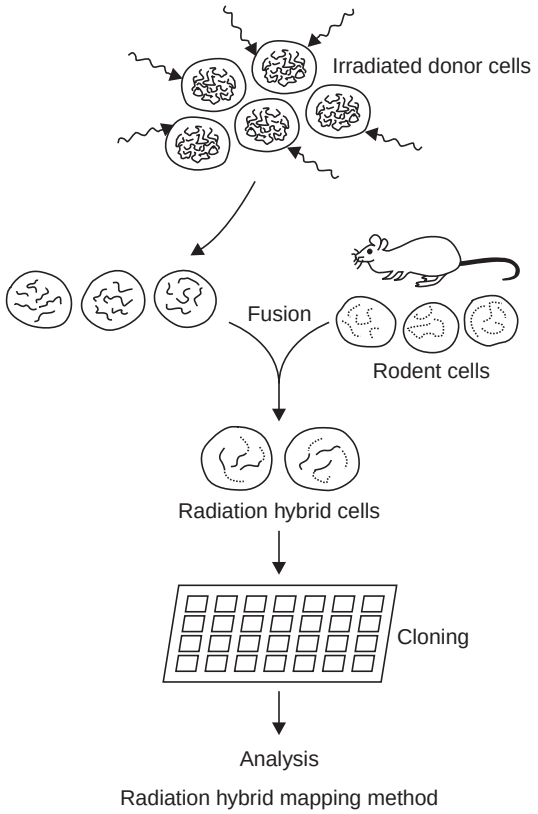


Fig. 25.1 Procedure for radiation hybrid mapping.

The fragments of human chromosomes integrate into rodent chromosomes because of breaks produced by X-rays, followed by joining. Cytological examination of cells shows that some rodent chromosomes have integrated fragments, and some contain whole chromosomes of humans. The integrated fragments are found to display banding patterns that correspond with chromosomes of irradiated human cells. Cell lines with single inserts and identifiable banding patterns are chosen. They are made to overlap one another to make a panel of radiation hybrids representing the whole genome.

DNA is isolated from each cell line in the radiation hybrid mapping panel, placed separately on membranes and denatured. A labeled single copy human DNA probe is hybridised to the membranes. Position of the label identifies the cell line carrying a human chromosome fragment *homologous* to the probe. The next step is to analyse data on probe hybridisation to detect **co-retention** of DNA markers. If two DNA markers map near each other on the same chromosome, it means they have been co-retained. In the mapping panel, higher the frequency of *co-retention* of two human DNA markers, the closer the two markers are, in map distance, on the same human chromosome. The co-retention of different human markers in radiation hybrid mapping panels allows high resolution mapping of the DNA markers on chromosomal loci.

The radiation hybrid mapping method has an advantage. A standard panel of only about 100 to 200 hybrids is sufficient to generate a high resolution map of the human genome.

Physical Mapping of the Genome

The genetic maps developed from meiotic recombination studies, described above, do not have sufficient resolution to allow sequencing of the genome. It is therefore, necessary to generate a detailed physical map. The goal is to have a map of markers based on direct analysis of genomic DNA, instead of analysis of recombinants from pedigree analysis. Physical maps are maps of cloned genomic DNA, that are supplements of a genetic map. There are 24 physical maps, corresponding to the 22 autosomes, an *X* and a *Y* chromosome in humans. ***In situ* hybridisation**, **FISH** and **radiation hybrid mapping** methods provide higher resolution physical maps of human genome. We now describe **restriction mapping** and **clone contig** methods for generating physical maps.

Restriction Mapping: When the entire genome has to be mapped by using restriction enzymes, the commonly used enzymes such as *EcoR*I and *Hind*III present serious limitations. These enzymes have a very large number of restriction sites throughout the genome and yield too many fragments that would produce a smear on the gel and cannot be resolved separately. Therefore, two types of enzymes are used, enzymes that recognise 7 to 8 nucleotide sequences, and enzymes that have recognition sequences that are uncommon in the DNA being mapped. For example, enzyme *Not*I can be used which has recognition sequence

5′-GCGGCCGC-3′

3′-CGCCGGCG-5′

*Not*I cuts human DNA once every 10 Mb on an average. The second type of enzymes recognise sequences that are uncommon, being present in low frequencies in human DNA. For example, the sequence 5′-GC-3′ and its complementary 3′-CG-5′ are rare in the human genome. Therefore, enzymes that have 5′-GC-3′ and 3′-CG-5′ in their recognition sequences can cut human DNA less frequently because it is rare. Restriction maps made with these types of restriction enzymes are able to resolve genes that are several hundred kilobase pairs apart. However, even with the availability of enzymes like *NotI*, generating restriction maps of all the chromosomes is an uphill task. In year 2000, restriction map of one of the smallest human chromosomes, that is chromosome number 21, was the only one to have been completed by this method.

Clone Contig Map: A **clone contig map** (contig is a short form of contiguous) comprises a set of ordered (occurring consecutively) partially overlapping clones comprising all the DNA of an *individual chromosome* without any gaps. It is necessary to construct 24 contig maps, one each for the 22 autosomes, an *X* and a *Y* chromosome. Because of the large size of the human genome, vectors that can accommodate large-sized DNA inserts only can be used. Otherwise the number of clones would become enormous, and difficult to handle. To start with, YAC vectors which can carry several hundred kilobase DNA inserts were used.

The construction of a clone contig map begins with the preparation of a library of partially overlapping DNA fragments (method in Chapter 23). Either whole genomic DNA is used, or individual chromosomes are first separated by flow cytometry before DNA is isolated. Usually the DNA is sheared mechanically by passing through a syringe needle to procure a complete set of randomly overlapping fragments of genomic DNA. This method of shearing produces blunt ends, therefore, to insert these fragments into YAC vector, the restriction enzyme used for cutting must be one that produces blunt ends, such as *Sma*I. By this method a YAC library of either whole genomic DNA is produced, or a library for a particular chromosome.

The YAC clones can yield contig maps by using chromosome mapping techniques such as FISH. For large genomic regions, DNA fingerprinting is used. The clones are assembled in order on the basis of overlaps between them. However, the most commonly used method involves **sequence-tagged site (STS)** method. An STS which serves as a DNA sequence marker is a

short unique sequence in the genome that can be amplified by using defined PCR primers. The technique is based on the principle that clones which share STSs must overlap each other. PCR will screen individually all the clones from a YAC library. All clones containing a particular STS amplified by the primer pair used will yield amplification product. Samples of the DNA in the PCR reaction mixtures for each clone after STS amplification are examined for labelled DNA products using fluorescent or radiolabelling or by techniques that detect complementarity. If the library is complete with no gaps, it would produce a complete contig map for the chromosome or genomic region being assembled, by locating individual STSs on the genome map on the basis of their locations on cloned DNA. YAC mapping has been done successfully for chromosome number 21.

Besides YAC contig maps, BAC cloning vectors have been used and they have provided more accurate clone contig maps. HGP team has also used radiation hybrid mapping of *STS markers* and prepared a map with 15,806 STS markers. These maps were further enhanced by using other methods, one of which is **expressed sequence tag (EST**) method which added an additional 20,104 STS markers. This involves PCR using oligonucleotide primers designed on the basis of the sequence of a complementary DNA (*c*DNA), to obtain an EST marker as a PCR product. As detailed in Chapter 23, a *c*DNA is complementary to the *m*RNA transcript, therefore, an EST marker corresponds to a functional protein coding gene. If the gene is a unique one, then the EST of that gene is also unique. By determining EST markers, the map was enriched with a large number of *protein coding genes*.

Individual Chromosome Libraries: The construction of a physical map of the whole genome presents many challenges even with the use of STS markers. There may be identical sequences, megabase in size, whose duplicate copies may exist in two different chromosomes. Or there may be gaps in the physical map due to certain regions in the genome that do not clone efficiently. To overcome these difficulties, individual chromosomes can be isolated for preparation of chromosome-specific libraries and to develop maps of DNA in each chromosome.

To isolate individual chromosomes if they are small as in yeast, pulse field gel electrophoresis (PFGE, detailed in Chapter 23) can be used. For large chromosomes in the human genome, fragments have to be prepared by using restriction enzyme *Not*I which recognises an eight base pair sequence and cuts only once in every 64,000 base pairs. The large sized fragments are separated by PFGE.

Flow Sorting of Chromosomes: Specific human chromosomes can be separated **(flow-sorted)** by **fluorescence-activated chromosome sorting (FACS)**. Whole metaphase chromosomes are isolated into a suspension by disrupting dividing cells. The metaphase chromosomes are stained with two fluorescent dyes, one which binds to AT-rich regions, the other to GC-rich regions. The technique is based on the principle that every chromosome has a characteristic ratio of AT-rich to GC-rich regions. The suspension of stained chromosomes is diluted to an appropriate concentration, so that when converted into a spray of droplets, each droplet contains *one* chromosome. The droplets flow in a tube under a laser beam that excites fluorescence. The computerized detection system sorts out each chromosome according to the ratio of the AT-rich to GC-rich regions, and collects each chromosome in an individual tube. DNA from individual chromosomes is isolated and used for preparing chromosome-specific libraries containing inserts from these DNAs. Chromosomes are identified by hybridisation with probes known to be complementary to each of the chromosomes. Chromosome libraries are used to map STS or other markers to develop physical maps.

Sequencing the Genome

The construction of a high resolution map leads to the next step of sequencing the genome. The dideoxy method of sequencing described in Chapter 23 is used. However, for the large-sized inserts

in BAC vectors used for the human genome, a more suitable method of whole genome **shotgun sequencing** is used. In the shotgun approach, each insert is cut out and sheared mechanically into a partially overlapping set of fragments (Fig. 25.2) and cloned into a plasmid vector. These subclones are then sequenced, and the overlapping sequences are assembled into a contiguous sequence using computational methods. Each BAC clone is mapped onto the chromosome. To assemble the complete sequence of one chromosome, sequences for the individual clone inserts are integrated into one contiguous sequence. The sequences of all the chromosomes then constitute the whole genomic sequence.

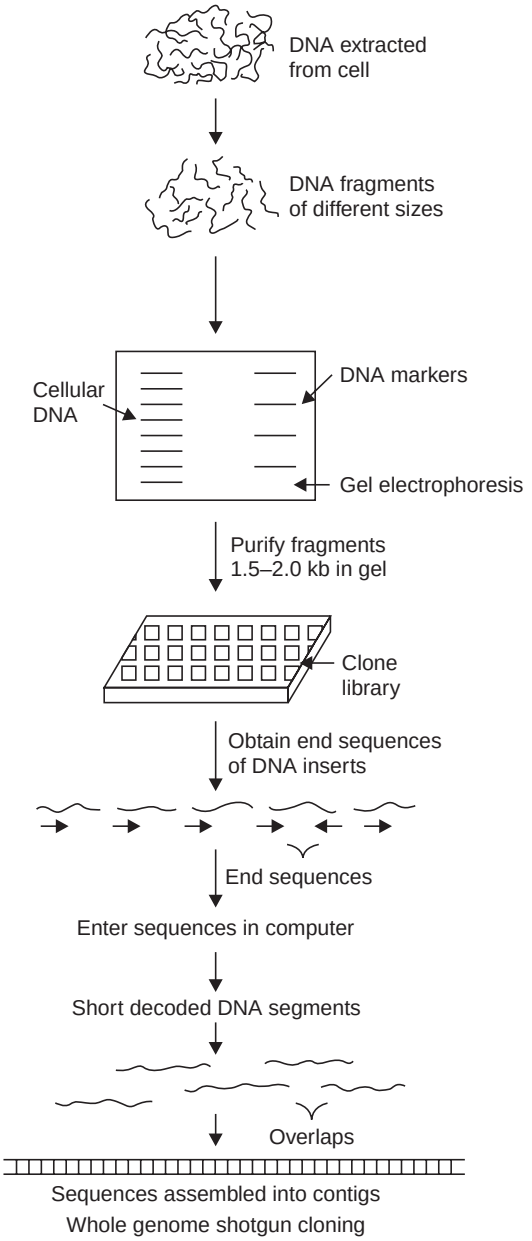


Fig. 25.2 Steps in the process of whole genome shotgun cloning (explanation in text).

OPTICAL MAPPING OF WHOLE GENOME

Optical mapping is a procedure for constructing whole genome restriction endonuclease maps from randomly sheared genomic DNA molecules extracted from cells. A key element of optical mapping system is that the DNA molecules are elongated and *immobilised* on positively charged glass surfaces and subsequently cut with a restriction endonuclease.

The procedure for **optical mapping** begins with extraction of DNA from any source. The extracted DNA is applied to a microfluidic device called **optical chip** that consists of a glass surface bearing a mask of several channels. As the DNA sample flows through the multiple channels of the optical chip, individual DNA molecules elongate and become fixed (immobilised) to the surface. Immobilisation occurs via electrostatic interactions between the negatively charged DNA and positively charged glass surface. The individual DNA molecules range in size from 0.5 to 2.5 million base pairs. Fragments that are 2 kilobase or smaller may be difficult to retain on the glass surface and could be lost, resulting in errors during image acquisition and subsequent processing. The array of all the DNA molecules can be examined for the presence of specific markers. One optical chip is capable of capturing multiple copies of a genome from any source, microbe to man.

Subsequent to immobilisation of the DNA molecules, cutting with a restriction endonuclease is carried out. Typically, restriction enzymes that recognise a 6 base pair sequence are used to cleave the DNA. Since the elongated DNA molecules are under slight tension, their ends retract, resulting in gaps. The gaps are visualised after staining with a fluorescent dye.

The generation of optical images requires a computerised optical mapping station mounted on a vibration-free table. There is a computerised argon ion laser scanning fluorescence microscope in the work station that scans the optical chip and processes the information to generate visual images of each field. The workstations are configured for groups of six stations.

The optical mapping system takes over the images produced by each station. It has a multiprocessing data analysis system called the optical mapping cluster. This system integrates the overlapping visual fields and selects molecules for analysis. Single molecule maps are then constructed from the order and size of the fragments. The mass of the fragments is determined by the automated image analysis software which locates each fragment and measures the amount of fluorescent dye bound to it. The gaps at the ends of the DNA fragments, their presence and spacing, also serve as markers and are typical for each sample.

In optical mapping, any given region of the genome is represented by a number of different DNA molecules which are seen in the form of individual concentric rings in the sample. Overlapping sets of single molecules can be assembled out of the concentric rings by using proprietary softwares that can search similar fragment patterns in different single molecules and develop them into a consensus map representing the whole genome.

Ethical, Legal and Social Issues (ELSIs)

Scientists and general public have expressed their concerns concerning potential use of new genomic information. The following ethical, legal and social issues (**ELSIs**) are under examination. These include privacy and confidentiality of genetic information; fairness in use of genetic information by employers, insurance companies and others; psychological impact of genetic deviations on individuals; reproductive and clinical issues; environmental issues concerning genetically modified microbes.

SYNTHETIC GENOMICS

An emerging new technology involving **genome transplantation** allows one type of bacterium to become transformed into another type. The transformed bacterium then expresses information contained in the transplanted chromosome. Thus, scientists Carole Lartigue and colleagues (2007) have been able to change *Mycoplasma capricolum* into *Mycoplasma mycoides* Large colony (LC). Mycoplasmas are small cells lacking a cell wall.

The procedure for genome transplantation is carried out in several steps. First, a marker gene is incorporated into *M. mycoides* LC chromosome. The marker gene is antibiotic selectable, therefore, it allows selection of living *M. mycoides* LC cells containing the transplanted chromosome. In the next step, DNA (chromosome) is extracted from the selected *M. mycoides* LC cells and purified to eliminate proteins. This chromosome of *M. mycoides* is then transplanted into cells of *M. capricolum*. Repeated proliferation of *M. capricolum* cells resulted in elimination of the recipient *M. capricolum* chromosome, and it was replaced by the *M. mycoides* LC chromosome. The transplanted *M. mycoides* LC chromosome was expressed, and *M. capricolum* cells displayed all the phenotypic characteristics of *M. mycoides* LC cells.

Results of the above experiments were confirmed using gel elecrophoresis and protein sequencing. The expressed proteins were found to be the ones coded for by the *M. mycoides* LC chromosome. Subsequent to the success of the transplant experiments, a future goal of synthetic genomics is to attempt transplantation of *chemically synthesised chromosomes* into viable living cells. The important applications of synthetic genomic research include development of new energy sources, pharmaceuticals and chemicals.

SELECTED READINGS

Anantharaman, T.S., Mishra, B. and Schwartz, D.C. 1997. Genomics Via Optical Mapping II. Ordered Restriction Maps. *J. Computer Biol*. **4** : 91–118.

Barbier, V. and Viovy, J.L. 2003. Advanced Polymers for DNA Separation. *Curr. Opinions Biotechnol.* **14** : 51–57.

Reslewic, S. *et al.,* 2005. Whole Genome Shotgun Optical Mapping of *Rhodospirillum rubrum. Appl. Environ. Biol.* **71** (9) 5511–5522.

This page intentionally left blank

Evolutionary Genetics

This page intentionally left blank

26

Populations and Gene Frequencies

Population genetics concerns inheritance of an interbreeding group of individuals of the *same* species existing as a population. It is the study of the genetic constitution of the group and changes therein from generation to generation. As evolution is also the study of changes that occur through time in the genetic structure of a population, the two branches, population genetics and evolution are together designated "*Evolutionary Genetics*". The field of population genetics is restricted to only one species, and is one of the essential components of the broad field of evolution. The study of evolution includes *all living organisms,* whether single or multicelled, and reproducing sexually or asexually. The primitive precellular aggregates of organic molecules (described in next chapter) that arose over 3 billion years ago are as much a salient feature of evolution as the complex, multicelled, present day organisms including man.

GENE AND GENOTYPE FREQUENCIES

The genotypes of all interbreeding individuals in a population collectively form a *gene pool.* The repository of the human gene pool is in the nearly four billion people existing today in the world. In a population with N individuals there are $2N$ haploid genomes. The haploid gametes combine during fertilisation to create a new set of genotypes which together produce a reconstituted gene pool in each generation. The mating pairs are random associations of genotypes, and the term *deme* is often applied to such populations. Variations in gene pools are expressed in terms of gene frequencies, genotype frequencies or allelic frequencies.

Methods of Measuring Genotype Frequencies

One way of measuring genotype frequency is from phenotype frequency. Consider the case of 3 blood groups *A*, *AB* and *B* determined by two alleles I^A and I^B at a single locus. In a random sample of 1000 humans, the *A* group occurred in 210, *AB* in 450 and *B* in 340 individuals. The frequencies of the blood group phenotypes and their respective genotypes are obtained by dividing the number of individuals for each blood group by the total. Thus, the frequency of blood group *B* for instance would be 340/1000 = 0.34.

Another way of estimating genotype frequency is to first calculate *gene frequency* of genes *A* and *B* in the population. Assume that the above sample contains 210 *AA,* 450 *AB* and 340 *BB* individuals. The gene frequency of *A* in the population is represented by the probability to find *A*

allele at the *AB* locus and is exactly equivalent to the proportion of *A* alleles among all alleles at this locus in the sample or in the population (that is because we cannot determine the frequency of *A* in the whole population). As each individual carries two alleles at the *AB* locus, the total number of alleles in the sample is 1000 × 2 = 2000. Out of these 210 + 210 + 450 = 870 are *A*. Therefore the frequency of the *A* allele is 870/2000 = 0.435. The number of *B* alleles is 450 + 340 + 340 = 1130, and the frequency of *B* allele is 1130/2000 = 0.565. If we represent the gene frequency of *A* by *p,* then *p* represents a value between 0 and 1 (because the proportion of allele *A* must lie between 0 and 100 per cent). In our example $p = 0.435$. Similarly, if we symbolise the frequency of *B* by *q,* then $q = 0.565$. It may be noted that $q = 1 - p$ or $1 - 0.435$. Similarly $p = 1 - q$ or 0.565. Thus $p + q = 1$.

For predicting genotype frequencies some assumptions have to be made, such as random mating in the population. That is to say, with respect to the trait of blood groups, an individual will mate with another without regard to whether the blood group of the mate is *AA, AB* or *BB*. Random mating implies random union of eggs and sperm, which in the example cited is the frequency of *A* and *B* alleles among the eggs and sperm (the gametes). Now the probability for the allele *A* at the *AB* locus in the population is *p,* and this is also the probability for a randomly chosen gamete to carry allele *A*. Similarly the probability for a randomly chosen gemete to carry *B* is *q*. For producing individuals with genotype *AA,* an *A* sperm must fertilise an *A* egg. This occurs with probability $p \times p = p^2$. An *AB* genotype results from fertilisation of *A* sperm with *B* egg or *vice versa,* and the probability is $p \times q$ or $pq + qp = 2pq$ (that is pq for *AB* genotype and qp for *BA* genotype). The frequency of *BB* genotypes depends upon the chance fertilisation of a *B* sperm and *B* egg; this has probability $q \times q = q^2$. Thus the frequencies of *AA, AB* and *BB* genotypes in the population should be expected to be p^2, $2pq$, and q^2 respectively.

If we substitute the values of $p = 0.435$ and $q = 0.565$ we can know that the frequency of *AA* $= (0.435)^2 = 0.189$, *AB* $= 2 \times 0.435 \times 0.565 = 0.246$, and of *BB* $= (0.565)^2 = 0.319$.

THE HARDY-WEINBERG LAW

In 1908, the mathematician G. H. Hardy in England and the physician W. Weinberg in Germany independently developed a quantitative theory for defining the genetic structure of populations. The Hardy-Weinberg Law provides a basic algebraic formula for describing the expected frequencies of various genotypes in a population. The similarity of their work however, remained unnoticed until Stern (1943) drew attention to both papers and recommended that names of both discoverers be attached to the population formula. The Law states that *gene frequencies in a population remain constant from generation to generation* if no evolutionary processes like migration, mutation, selection and drift are operating. Thus if matings are random, and no other factors disturb the reproductive abilities of any genotype, the equilibrium genotypic frequencies are given by the square of the allelic frequencies. If there are only two alleles *A* and *a* with frequencies *p* and *q* respectively, the frequencies of the three possible genotypes are

$$(p + q)^2 = p^2 + 2pq + q^2$$

If there are 3 alleles say *A*1, *A*2 and *A*3 with frequencies *p, q* and *r,* the genotypic frequencies would be

$$(p + q + r)^2 = p^2 + q^2 + r^2 + 2pq + 2pr + 2qr$$

This square expansion can be used to obtain the equilibrium genotypic frequencies for any number of alleles.

It must also be noted that the sum of all the allelic frequencies, and of all the genotypic frequencies must always be 1. If there are only two alleles p and q, then $p + q = 1$, and therefore $p^2 + 2pq + q^2 = (p + q)^2 = 1$. If there are 3 alleles with frequencies p, q, and r, then $p + q + r = 1$, as well as $(p + q + r)^2 = 1$.

The time required for attaining equilibrium frequencies has been determined. If a certain population of individuals with one set of allele frequencies mixes with another set and complete panmixis occurs (that is, random mating), then the genotypes of the next generation will be found in the proportion $p^2 + 2pq + q^2$ where p and q are allele frequencies in the new mixed populations. Thus it takes only *one generation* to reach Hardy-Weinberg equilibrium provided the allelic frequencies are the same in males and females. If the allelic frequencies are different in the two sexes, then they will become the same in *one* generation in the case of alleles on autosomes, and genotypic frequencies will reach equilibrium in *two generations*. In general equilibrium is arrived at within one or at the most a few generations. Once equilibrium is attained it will be repeated in each subsequent generation with the same frequencies of alleles and of genotypes.

The Hardy-Weinberg law is applicable when there is random mating. Random mating occurs in a population when the probability of mating between individuals is independent of their genetic constitution. Such a population is said to be *panmictic* or to undergo *panmixis*. The matings between the genotypes occur according to the proportions in which the genotypes are present. The probability of a given type of mating can be found out by multiplying the frequencies of the two genotypes that are involved in the mating. Matings are not random for instance when a population consists of different races such as blacks and whites in the U.S., or different communities as in India as there are preferred matings between members of the same racial or communal group.

Applications of the Hardy-Weinberg Law

(*a*) *Complete Dominance:* When Hardy-Weinberg equilibrium exists, allele frequencies can even be found out in presence of complete dominance where two genotypes cannot be distinguished. If two genotypes *AA* and *Aa* have the same phenotype due to complete dominance of *A* over *a* the allele frequencies can be determined from the frequencies of individuals showing the recessive phenotype *aa*. The frequency of *aa* individuals must be equal to the square of the frequency of the recessive allele q. Let us suppose $q = 0.5$, then $q^2 = (0.5)^2 = 0.25$. In other words when *aa* phenotype is 0.25 in the population, then it follows that the frequency of the recessive allele *a* is $\sqrt{0.25} = 0.5$. The frequency of the dominant allele *A* would be $1 - q$ or $1 - 0.25 = 0.75$.

(*b*) *Frequencies of Harmful Recessive Alleles:* The Hardy-Weinberg Law can also be used to calculate the frequency of heterozygous carriers of harmful recessive genes. If there are two alleles *A* and *a* at an autosomal locus with frequencies p and q in the population and $p + q = 1$, then the frequency of *AA*, *Aa*, and *aa* genotypes would be $p^2 + 2pq + p^2$. If the *aa* genotype expresses a harmful phenotype such as cystic fibrosis, then the proportion of affected individuals in the population would be q^2, and the frequency of the heterozygous carriers of the recessive allele would be $2pq$. To illustrate with figures, suppose one out of 1,000 children is affected with cystic fibrosis, then the frequency $q^2 = 0.001$, so that $q = \sqrt{0.001}$ which is about 0.032, then $2pq = 2 \times 0.032 \times 0.968 = 0.062$. This means that about 62 individuals out of 1000 or one out of 16 is a carrier of the allele for cystic fibrosis. As already mentioned the number of individuals (*aa*) who are actually affected is one out of 1000. This implies that the frequency of heterozygous carriers is

much higher than that of affected homozygotes. Similar calculation shows that when an allele is very rare in the population the proportion of carriers is still much higher and of affected homozygotes much lower. Thus, lower the frequency of an allele, greater the proportion of that allele that exists in the heterozygotes.

(*c*) *Multiple Alleles:* The Hardy-Weinberg Law permits calculation of genotypic frequencies at loci with more than two alleles, such as the ABO blood groups. There are 3 alleles I^A, I^B and I^O with frequencies p, q and r. Here $p + q + r = 1$. The genotypes of a population with random mating would be $(p + q + r)^2$.

(*d*) *Sex-linked Loci:* It is possible to apply Hardy-Weinberg Law for calculating gene frequencies in case of sex-linked loci in males and females. Red green color blindness is a sex-linked recessive trait. Let *r* denote the recessive allele which produces affected individuals, and *R* the normal allele. The frequency of *R* is *p* and of *r* is *q* where $p + q = 1$. The frequencies of females having *RR*, *Rr*, *rr* genotypes would be p^2, $2pq$, q^2 respectively. Males are different as they are hemizygous, have only one *X* chromosome derived from the mother with a single allele either *R* or *r*. The frequency of affected *r* males would be the same as the frequency of the *r* allele among the eggs, that is *q*. The frequency of normal *R* males would be *p*. Suppose the frequency of *r* alleles is 0.08, then the incidence of affected males would be 0.08 or about 8%. The frequency of affected *rr* females would be $(0.08)^2 = 0.0064$ or 0.64%. Thus the Hardy-Weinberg Law explains that males would be affected a hundred times more frequently than females. This is actually what is observed. Males are more affected by sex-linked recessive traits than females (see Chapter 21).

The difference between the sexes is even more pronounced if the recessive allele is still more rare. The incidence of a common form of haemophilia is one in a thousand males; thus $q = 0.001$. However, only one in 1000,000 females will be affected. Thus males could have haemophilia one thousand times more often than females.

(*e*) *Linkage Disequilibrium:* Consider two or more alleles at one locus and another locus on the *same chromosome* with two or more alleles. Due to genetic exchange by recombination occurring regularly over a period of time, the frequencies of the allelic combinations at the two syntenic loci will reach equilibrium. If equilibrium is not reached, the alleles are said to be in *linkage disequilibrium*. The effect is due to tendency of two or more linked alleles to be inherited together more often than expected. Such groups of genes have also been referred to as *supergenes*.

NONRANDOM MATING

Hardy-Weinberg Law is applicable only when mating is random. When genotypes do not mate at random it is called *nonrandom mating*, that is, individuals with certain genotypes prefer to mate with individuals of certain other genotypes. Consider for example the case of albinos having recessive genotype *aa*; normal individuals are *AA* and *Aa*. The frequency of *a* allele is 0.01, and of the normal *A* allele is 0.99. When the population is at equilibrium, the frequency of *AA* individuals is 980 per thousand, of heterozygous carriers *Aa* is 19.8 in a thousand, and albinos 0.1 per thousand. Obviously there are about 49 times more of heterozygous carriers than albinos in a sample of 1000 members of the population. Now *AA* and *Aa* individuals are both normal in appearance and mate at random. But albinos are less likely to mate with albinos or even perhaps with normals. Thus mainly $Aa \times Aa$ matings are the source of all albinos.

Inbreeding and Assortative Mating

These are two departures from random mating. Inbreeding is a form of nonrandom mating that takes place between relatives having like genotypes. It increases the frequency of homozygotes in the population and decreases the frequency of heterozygotes. The harmful effects of inbreeding in increasing the frequency of recessive disorders in humans is discussed in Chapter 21 under consanguineous marriages. The closest possible degree of inbreeding is *self-fertilisation* which occurs in some plants such as the sweet pea. In *assortative mating* individuals with similar phenotypes mate more often than expected by chance. Assortative mating causes a lesser decrease in heterozygosity than by inbreeding, but produces greater increase in phenotypic variation. Phenotypic characters like height, skin color, I.Q. and others form the basis of assortative mating in humans. The most important effect of assortative mating is to increase the variation of a trait in the population.

Inbreeding Coefficient

The coefficient of inbreeding is the probability that an individual receives at a given locus two alleles that are *identical* by descent. It is a method of measuring the genetic consequences of inbreeding, namely increased frequency of homozygous genotypes. When relatives marry, they are likely to have one or more common ancestors. If an ancestor of both mates is carrier of a harmful recessive gene, in that case, both mates who are direct descendants could also be carriers. Inbreeding gives rise to a homozygous affected child; the two alleles in such an affected child are said to be *identical by descent* because they originate by replication of a single allele carried by an ancestor. The probability of receiving two alleles that are identical by descent is a measure of the inbreeding coefficient and is designated by the symbol *f*.

In a randomly mating population with two alleles having frequency p and q, the frequency of heterozygotes is $2pq$. In a population with coefficient of inbreeding f, the frequency of heterozygotes will be reduced by a fraction f of the total. One aspect of the Hardy-Weinberg Law takes inbreeding into account. Consider 2 alleles A and a in a population with frequencies p and q such that $p + q = 1$. If the inbreeding coefficient of the population is f, then the frequencies of AA, Aa, and aa genotypes will be $p^2(1 - f) + pf$, $2pq(1 - f)$, and $q^2(1 - f) + qf$ respectively. When there is no inbreeding, f is reduced to zero, and the proportions of genotypes attain Hardy-Weinberg equilibrium values of random mating populations. In an imaginary situation of complete inbreeding, $f = 1$. However, inbreeding in humans never reaches such high values. On the average the inbreeding coefficient in humans ranges between $f = 0.003$ in some communities in Japan to its highest value of $f = 0.02$ in some states in South India such as Andhra Pradesh.

The harmful effects of inbreeding result from rare, deleterious recessive genes becoming homozygous, and also depend on the frequency of rare harmful alleles in the population, When harmful genes become homozygous in self-fertilising plants, they are rapidly removed from the population. Studies on the effects of inbreeding in humans have shown mental and physical defects and increased risk of death in children of first cousin marriages. Such children may or may not have a lower I.Q., growth rate and lesser capability than noninbred children. It has been estimated that an average human being carries 3–5 genes that could produce severe mental or physical disability. This means that a normal human being is a heterozygous carrier of 6–10 alleles which in the homozygous condition could cause death or severe mental or physical disabilities. The phenomenon is called *mutational load*. As mentioned earlier, only a minority of the total number of the harmful recessive genes in a population are actually present in affected homozygotes. The majority of such genes are hidden in heterozygous, normal and healthy people who are carriers.

Inbreeding Depression and Heterosis

Plant and animal breeders have used inbreeding for increasing homogeneity through homozygosity. It is well known that inbreeding also leads to reduction in vigour, fitness, fertility and other such attributes. This is called *inbreeding depression* and results from deleterious alleles becoming homozygous. However, if independently inbred lines are crossed, the resulting hybrids show increased vigour, fitness and fertility over the parents. This is called *hybrid vigour* or *heterosis* and is much exploited in the improvement of crop plants such as maize, cotton, castor, *Pennisetum* and *Sorghum.* Among animals, the vigour displayed by the mule, the result of a cross between a male horse and a female donkey, is a familiar example. The mule is a better beast of burden than either of its parents.

VARIATION IN POPULATIONS

Enormous diversity is visible in populations. Traits like height, color of skin, hair or eyes, shape of nose, mental ability and others which vary among people may be due to heredity or environment or both. Statistical methods have shown that some of morphologic variation is genetic and comes from the human gene pool.

Evidence for genetic source of variation has been derived from the study of phenotypes that are not influenced by the environment. For example, a single allele controls the development of hairy pinna in males; another allele determines ability to roll the tongue. A person can taste PTC (phenylthiocarbamide) due to a single dominant allele, nontaster being homozygous recessive. The blood group alleles also provide variation. Gene loci that are variable in the population are said to be *polymorphic*.

Changes in Gene Frequencies in Populations

All species of organisms display variation in their populations. There is an enormous range of variation in heritable traits visible at the phenotypic or genotypic level. Genetic variations forms the basis for evolutionary change. Changes in populations may be due to various factors such as migration of individuals from other populations; mutations and recombinations; selection; random fluctuations in the reproductive rates of different genotypes. Genetic variation in populations, also called *polymorphism* relates to gene loci that are variable in the population. A measure of this variation is the amount of **heterozygosity** in the population. There is an immense variety of genetic variation in a species. Some of the more prominent ones are outlined here.

Variation in snail shell morphology

The shell of the land snail *Cepaea memoralis* has two alleles at a single locus that determine color of shell as pink or yellow, with pink dominant over yellow. Another locus linked to the color locus determines banded versus unbanded shells. Unbanded is dominant over banded. Variations of these two loci produce a number of combinations of color (pink or yellow) and banded versus unbanded morphology of the shell.

Chromosomal variation

Several chromosomal variations involving increase in number (supernumerary chromosomes), translocations and inversions are present in many species of plants, insects and mammals.

Polymorphisms in proteins

Polymorphism in proteins can be traced to the structural genes encoding the polypeptide. A change in the triplet codon of a structural gene results in amino acid substitution in the polypeptide chain. The method of detecting polypeptide polymorphism is to isolate a specific purified protein from several individuals and determine the sequence of the protein. This method of sequencing is not practical for detection of genetic variation in a population.

A more practical method is based on assessment of the net charge carried on some of the amino acids of the protein. When an amino acid substitution changes the sequence in a polypeptide chain, it leads to a change in the physical properties of the protein. The amino acids aspartic acid, glutamic acid, arginine, lysine and histidine have side chains that can become ionised and impart a characteristic net charge on the protein, depending on the pH of the surrounding medium. The net charge distribution of the protein could change due to different forms of amino acid substitutions. That is, amino acid substitution could replace a charged amino acid, or affect the degree of ionisation of a charged amino acid, or a substitution at the point of joining of the two α helices could bring about a change in the three dimensional folding of the protein.

Any of these substitutions could alter the net charge on a protein, which could be detected by subjecting the protein to gel electrophoresis. The proteins from a number of different individuals can be loaded in the wells of the gel apparatus. The variant proteins in homozygotes and heterozygotes for one or both alleles can be determined in individuals from the gel. Since the method involves estimation of proteins expressed by the structural genes present at the locus, it allows study of only those loci that are not segregating.

DNA Sequence Polymorphism

The DNA sequences could vary among individuals in a population. These variations can be studied by using restriction enzymes or by DNA sequencing methods.

Restriction Fragment Variation: A restriction enzyme such as *Eco*RI that recognizes a six nucleotide sequence will cut the DNA at a defined number of sites based on probability. If there is variation for one of the six bases at the recognition site, then it will result in restriction fragment length polymorphism (RFLP) in the population. The enzyme will cut one variant at the recognition site, but it will not recognise another variant in the six base sequence. It is however, not possible to determine which particular base is changed in the six nucleotide sequence. The study gives an idea about the amount of nucleotide variation in a particular sequence. The amount of heterozygosity gives a measure of this variation in the population. A detailed study of polymorphism in the X chromosome and the two large autosomes in *Drosophila* using 20 restriction enzymes indicated between 0.1 and 1.0 percent heterozygosity per nucleotide site, and absence of polymorphism in the small autosome.

Repeat DNA Sequence Variation: Restriction enzymes can be used to assess variation in the repeated DNA sequences. The variable number tandem repeats (VNTRs) are short sequences dispersed throughout the genome. If restriction enzymes cut sequences that flank a tandem array of repeats on both sides, then the size of the fragment would be proportional to the number of repeated elements. When fragments from a number of individuals are subjected to gel electrophoresis, the different sized fragments will migrate at different rates during electrophoresis. Based on fragment size, the fragments can be classified and analysed for polymorphism.

CHANGES IN GENE FREQUENCIES IN POPULATIONS

Forces that Change Gene Frequencies in Populations

If the alleles of a population remain constant in frequency, no evolutionary changes will occur. Four major forces are usually listed for changing gene frequencies in populations, namely migration, mutation, selection and random genetic drift. These forces constitute the mechanisms underlying the evolutionary process.

Migration

Migration occurs when a large influx of people moves into another population and interbreeds with the latter. The phenomenon called *gene flow* takes place if one population contributes an allele to the other population. Let us suppose that a migrating population m interbreeds with members of another population. Then the descendants of the next generation will have m genes from the migrants and $1 - m$ genes from members of the original population. Consider an allele A occurring with frequency p in migrant population. In the original population this allele has frequency q. In the next generation the frequency of A in the new population would be

$$r = (1 - m)q + mp$$
$$= q - m\,(q - p)$$

That means the frequency of allele A in the new population now would be the original allelic frequency q multiplied by the genes $(1 - m)$ present in the original population plus the product of reproducing migrant individuals and their gene frequency (mp). Thus there will be a new gene frequency in the next generation.

Migration is a complex phenomenon in humans, influenced by many factors. It seems however, that it leads to make populations genetically more similar than they would be otherwise.

Mutation

The ultimate source of all genetic variation is mutation. Both chromosomal rearrangements and point mutations are implied here as they follow the same rules of population dynamics. However, mutations occur with an extremely low frequency. In humans where there may be from 30–50 successive mitotic divisions in the germ cells in each generation, only one gene in a million or 10 million roughly undergoes a mutation. Consider an allele A that is homozygous in many individuals in a population. Assume that in every generation one A allele in a million mutates to a. This will reduce the frequency of A allele over many generations, while a allele will gradually accumulate in the population. The change in frequencies of A and a occurs at an unimaginably slow rate. The a allele however, can also back mutate to A; this event will take place as the frequency of a alleles increases. After a very long time the number of A alleles lost by forward mutation would be balanced by the number of A alleles arising from back mutation of a to A. When this happens gene frequencies of A and a are said to be in *mutation equilibrium.* Thereafter, no further change in frequency of A and a will occur in subsequent generations. This applies however, only when the other evolutionary forces such as migration, selection and genetic drift are not operating to affect gene frequencies in the population.

Experimental work has shown that the rate of mutation can be affected by environmental factors like radiation and chemicals. In some instances mutations are under genetic control. There is a recessive mutator gene on the second chromosome of *Drosophila melanogaster*. In stocks of flies homozygous for the mutator gene, sex-linked recessive lethals occur spontaneously with high frequency (Demerec, 1937). In maize the recessive mutator gene *Dt* acts on an unlinked

locus *a* which controls synthesis of the purple anthocyanin pigment. The mutator gene *Dt* causes mutation of recessive *a* alleles to the dominant form *A* which leads to synthesis of anthocyanin appearing as purple spots on the stem, leaves and kernels of the maize plant. Mutator genes also occur in micro-organisms including *E. coli*. As the action of mutator genes is directed at a particular locus, they are said to produce *directed* mutations in contrast to *random* mutations which are not specific.

Normally all genes could mutate. In the case of a rare allele its mutation to other alleles is difficult to detect due to low frequency and slow rate of mutation. But when an allele occurs more frequently in the population it leads to higher mutation rate and increases the population's potential for evolutionary change. If a rare deleterious allele accumulates in the population, it is a disadvantage and constitutes what is referred to as the mutational load.

Selection

Selection is one of the forces that change gene frequencies in the population and a fundamental process of evolutionary change. The idea was first conceived by Charles Darwin in his *Origin of Species* published in 1859 and by Alfred Russel Wallace. Selection is defined as differential survival or fertility of different genotypes. If individuals carrying gene *A* are more successful in reproduction than individuals carrying its allele *a,* then the frequency of gene *A* will tend to be greater than that of gene *a*. The wide variety of mechanisms responsible for modifying the reproductive success of a genotype are collectively included under selection. It is the process that determines the contribution that people of different genotypes will make as parents of the next generation. Selection does not act on individual genes, but rather on the organism bearing the genes.

The reproductive efficiency of a genotype is measured in terms of the average number of offsprings born to the bearers of the genotype and is called *Darwinian fitness* or *relative fitness*. It is also referred to as the organism's *adaptive value*. The fitness value of 1 is usually assigned to the genotype with highest reproductive efficiency. However, fitness does not have an absolute value, and is expressed in relative terms as a ratio. *Relative fitness* (w) is obtained by dividing the fitness of all the genotypes by the fitness of any one genotype. Fitness simply describes the average number of progeny that survive and reproduce. The related term *selection coefficient* $s = 1 - w$. Some aspects in the individual's life are likely to affect the survival, growth and reproduction; consequently they affect the fitness of genotypes and are referred to as *fitness components*. Basically fitness depends upon survival and fertility. Persons affected with Huntington's chorea, a dominant condition may have 25% reproductive efficiency as compared to normal human beings. On the other hand children with Tay Sachs disease usually die before reproductive age. Thus the fitness of a person with Huntington's chorea is 0.25 and of Tay Sachs patient is zero.

When fitness of two alleles at a locus differs then selection favours survival of alleles with greater fitness and elimination of the other alleles. Thus frequency of one allele increases and of the other will decrease in the subsequent generations. However, if a rare allele occurs with low frequency, then selection is not able to cause much change in gene frequency. Specifically selection occurs against a recessive allele, or a dominant allele, resulting in its elimination; it could occur in favour of a heterozygote or against a heterozygote leading to polymorphism in a given trait.

When selection occurs in favour of a heterozygote over both homozygotes it is called *overdominance* or *heterosis*. It occurs when the fitness of the heterozygous genotype is *greater* than the fitness of both homozygotes. Assume that the relative fitness of the genotypes *AA*, *Aa* and *aa* are 0.9, 1 and 0.8 respectively. The greater fitness of the *Aa* genotype will not allow either *A* or *a* alleles from homozygotes to become fixed. Ultimately equilibrium gene frequencies would be attained. In humans overdominance has led to polymorphism in sickle cell trait, thalassaemia and G6PD.

The effect of selection is also counterbalanced by mutation. While selection is eliminating some genes from the population, mutation is creating new ones. The two forces selection and mutation operate in opposite directions, and tend to compensate each other. After a long time, gene frequencies will reach equilibrium.

There could be partial selection against recessives. This is a less complete form of selection against homozygous recessive individuals. In this case selection coefficient s is less than one, and the relative fitness w of the homozygous recessive individual is $1 - s$, having value greater than zero.

A popular example of selection is industrial melanism as exhibited by the pepper moth *Biston betularia*. In the mid 19th century the light coloured forms of the moth were abundant on the pale barks of trees growing in unpolluted, non-industrialised regions of England. The dark form of *Biston* was extremely rare. As industry developed in the area, the environment became polluted and the barks of trees turned dark grey with smoke and dust. The light moths on the dark coloured bark were easily noticed by the predators and were preyed upon. Their number began to decrease. In the following decades, the population of dark moths was observed to gradually increase to more than 95%; the light moths were hardly seen. Industrial melanism is thus a clear cut example of selection disturbing gene frequencies in the population.

Artificial and natural selection: Selection was being practised by humans since antiquity. Plant and animal breeders have been attempting to modify hereditary transmission of traits by selecting most desirable individuals to serve as parents for the next generation. This is called artificial selection. By contrast, when organisms are selected by natural forces instead of by human choice, they are said to be subject to natural selection.

Random Genetic Drift

These are unexpected random changes that occur in gene frequencies from generation to generation in all populations. They are particularly noticeable as sampling variation in small populations. In some generations the frequency of a certain allele will by chance increase, in others it will decrease, in still others it may remain the same· These fluctuations in gene frequency occur at random. In small samples there is greater variation as compared to big samples. Drift however, does not depend upon the total size of the population, rather on the number of breeding individuals who would produce the next generation. It is unlikely that random drift *alone* will affect allelic frequencies at a gene locus over long periods of time. It is more likely that selection, mutation or migration would also take place at one time or another.

SELECTED READINGS

Allison, A.C. 1954. Protection Afforded by Sickle Cell Trait Against Subtertian Malarial Infection. *Brit. Med. J*. **1** : 290.

Demerec, M. 1937. Frequency of Spontaneous Mutations in Certain Stocks of *Drosophila melanogaster*. *Genetics* **22** : 469.

Glass, B. and Li, C.C. 1953. The Dynamics of Racial Intermixture—An Analysis Based on the American Negro. *Amer. J. Human Genetics* **5** : 1.

Hardy, G.H. 1908. Mendelian Proportions in a Mixed Population. *Science* **28** : 49.

Kerr, W. E. and Wright, S. 1954. Experimental Studies of the Distribution of Gene Frequencies in Very Small Populations of *Drosophila melanogaster*. I. Forked. *Evolution* **8** : 172.

Li, C. C. 1955. Population Genetics. Univ. Chicago Press, Chicago.

Reed, T. E. 1969. Caucasian Genes in American Negroes. *Science* **165** : 762.

Stern, C. 1943. The Hardy-Weinberg Law. *Science* **97** : 137.

27

Origin of Life—Evolution of Man

For reconstructing the events that led to origin of life, we must go back to the endless void of eternity when the universe began. To do so we have to orient ourselves to the formation of planetary systems and the solar system to which the earth belongs; we must explain the existence of stars, the behaviour of galaxies and the distribution of elements. The most abundant chemical elements were hydrogen and helium. Agglomerations of atoms gave rise either to gas and dust, or to huge luminous bodies we know as stars. Our sun is one such star. Many millions of billions of stars are grouped together to form a galaxy; the galaxies are also clustered. Our galaxy called Milky Way contains some 100 billion stars (including the solar system) and has a lens shape about 60,000 light years in diameter and about 5,000 light years thick in the center. One light year is the distance that light travels in one year, about 6 billion miles. The total of *all* clusters of galaxies together constitute the universe. Every object in the universe is in motion. The 8 planets and the sun of our solar system are all in constant motion. The earth is moving around the sun at 10 miles per second, the sun has a disordered motion at 13 miles per second. Our universe is about 13 billion years old.

There is convincing evidence that thermonuclear reactions and subsequent explosions inside the stars generate all the chemical elements more massive than hydrogen and helium. These elements were distributed into the medium of space and produced subsequent generation of stars and planets. Thousands of new planetary systems are born into the universe every second. A comparison of the relative abundance of some atoms like hydrogen, helium, carbon, nitrogen, oxygen, magnesium, sulphur, phosphorus and a few others shows similar compositions in the universe, earth and living organisms. Similarities between isotopes of certain elements in the earth and in meteorites were also revealed. Studies show that heavy atoms present in living organisms were created in ancient stars which exploded after completing their lifetime and threw debris into space which again condensed into stars and planetary systems.

Our solar system was formed out of a diffuse cloud of cosmic dust which collapsed into a disc. The central portion or *nebula* had a higher density and condensed to form concentric bands around the sun. The planets differentiated from the bands; the inner smaller planets released volatile compounds like CH_4, NH_3, H_2O, H_2 and He. The earth arose by condensation of cosmic dust and larger objects and was probably in a molten state to start with. Because dust contained large quantities of radioactive elements, their decay in the interior of the earth generated much heat and volcanic action. When the earth cooled down sufficiently a crust was formed. Volatile

compounds like CH_4, CO_2, NH_3, N_2 and H_2O were released on to the surface mostly by volcanoes. This process is likely to have continued for several billion years and built up a primitive atmosphere.

ORIGIN OF THE FIRST ELEMENTS

The last few decades have witnessed much debate regarding origin of the first elements. Several theories, all of which go into the realm of subatomic physics have been advanced. Although mainly of historical interest, two of the theories are mentioned briefly. The first, Equilibrium Theory envisages formation of elements in earliest stage of the universe when matter attained very high temperature. The equilibrium abundances of the nuclei depended only on the mass, temperature and density of the nuclei. The second theory proposed by Mayer and Teller regards prestellar bodies to consist of a cold neutron fluid; the elements were formed by splitting off drops from the edges of these bodies.

The most accepted Big Bang theory originated when G. Gamow showed that one could expect that the earliest stages of the universe contained matter consisting largely of neutrons. The neutrons decayed to protons which captured more neutrons to produce the heavier elements. Based on this, George Gamow and Ralph Alpher proposed that all the elements could have been created by thermonuclear fusion during the explosion. The name Big Bang was given to this theory by the British astronomer F. Hoyle. According to this theory the universe was flooded with elementary particles like quarks and electrons constantly dissolving and reforming. There was abundant radiation. As cooling progressed, first of all gravity appeared, then came the strong force and lastly the weak force (Gravity, strong force, weak force and electromagnetism are the four fundamental forces in the universe today). The quarks then fused to form protons and neutrons; then within a fraction of a second they formed light elements like helium, lithium and deuterium by thermonuclear reactions. Gamow had also predicted that energy of the Big Bang should still be present in the universe in the form of a faint radionoise filling all of space. In 1965 this was indeed picked up as a microwave hiss by a sensitive instrument by A. Penzias and R. Wilson (later awarded Nobel Prize).

The more recent cosmologists are adding further information on the evolution of the universe. According to A. Guth the universe had an explosive beginning out of exotic subatomic particles like quarks, gluons, gravitons and neutrinos which became trapped in separate groups away from each other. The universe grew out of these tiny 'seeds' in a fraction of a second billions of years ago. After the initial trigger there was a hyperexplosive spurt of growth and expansion of the universe. Then there was gradual cooling over billions of years. The grand unified theories (GUT's) explain the existence of matter and antimatter in the beginning of the universe from calculations based on the Big Bang. They also predict the formation of the magnetic monopoles (massive particles with single magnetic poles) in the Big Bang, the details of which are beyond the scope of this text.

CHEMICAL EVOLUTION

The gases on the earth's surface were prevented from escaping into space due to the gravitational field of the earth. The earliest atmosphere contained much nitrogen and it was reducing being rich in CH_4, H_2, NH_3 and H_2O. Oxygen was absent and appeared much later as a product of

photosynthesis. Much evidence suggests that in the early history of the earth organic compounds first arose by reactions between various inorganic components of the atmosphere and the geosphere. The sources of energy were ultraviolet light, electric discharges, heat from volcanoes and other sources. This period of chemical evolution lasted over 1500 million years. The primordial biomolecules and the first organic compound arose during this period. Cooling caused water vapour to condense into water; torrential rains and downpours formed oceans. The biomolecules dissolved in the oceans producing a rich broth called the primordial 'soup'. The biomolecules condensed to form polymers; thus polypeptides, polynucleotides, polysaccharides and lipids were produced in this broth. The first living organisms are believed to have arisen from this broth. A. I. Oparin in the Soviet Union and J. B. S. Haldane in Britain share credit for postulating this sequence of events (in the 1920s) on the origin of life.

GEOLOGICAL EVIDENCES

The age of the earth is divided into geological periods on basis of the fossil record (Table 1). About 85% of the geological time is occupied by the earliest Precambrian period. A number of microfossils (electron micrographs of ancient micro-organisms) and organic material are available from the Precambrain rocks. The Precambrian was followed by the Cambrian period which is considered to have begun 600 million years ago. There are no fossils of hardshelled organisms before the Cambrian.

Table 1

Age in millions of years	Geological era	Event
Recent	Cenozoic	Man
	Mesozoic — Cretaceous, Jurassic, Triassic	Mammals and birds
	—	
	Palaeozoic — Permian, Devonian, Silurian, Ordovician, Cambrian	Terrestrial plants and fishes
600		
1000	Proterozoic	Eukaryotes
2000	Precambrian-Archean	Oldest rocks
3000		Photosynthetic blue green algae
4000		First Prokaryotes
4500		Origin of life
?		Origin of earth

The recent development of isotopic methods of dating rocks have led to the discovery of microfossils in the Precambrian. Some of the rocks examined are Bitter Springs Chert from

Central Australia (0.9 billion years), Gunflint Chert from S. Ontario (1.9 billion years), Bulawayan Limestone from Rhodesia (2.7 billion years), Fig Tree Chert from S. Africa (3.1 billion years), and Onverwacht in S. Africa (3.2 billion years). The oldest fossils are bacteria-like cells from the Fig Tree Chert which are more than 3.1 billion years old. These cells have been named *Eubacterium isolatum*; the finding suggests that life originated over 3 billion years ago. The second oldest fossils are filamentous blue green algae 2.2 billion years old (Nagy, 1974) in the dolomite limestone stromatolite in S. Africa. The oldest fossil of nucleated eukaryotic cell was discovered by Cloud (1969) and is about 1.2 billion years old. Some organisms in the Bitter Springs Chert resemble eukaryotic green algae and are less than a billion years old. The Fig Tree shale deposits have also been found to contain porphyrine, pyrimidines, purines and a few hydrocarbons, some of which are related to those found in present day cells. They provide evidence for existence of photosynthetic organisms.

Experimental Evidences

Spark Discharge Experiments. The experiments by Stanley L. Miller performed under simulated primitive earth conditions have yielded amino acids as products. The apparatus claimed to represent a model of the primitive earth is shown in Fig. 27.1. Water is boiled in the small flask to expel air and for bringing vapour to the region of the flask. Then a mixture of CH_4, NH_3 and H_2 is added; these reduced forms of carbon and nitrogen serve to represent the reduced atmosphere on primitive earth. An electric spark is generated across the gap in the large flask. The products condense as they flow through the tube and are collected in a U tube. The spark is operated continuously for a week and the products are analysed by chromatography. The results showed formation of several amino acids and simple compounds like urea, formic acid, succinic acid, lactic acid and a few others; surprisingly, many of these substances occur in living organisms. When the experiment

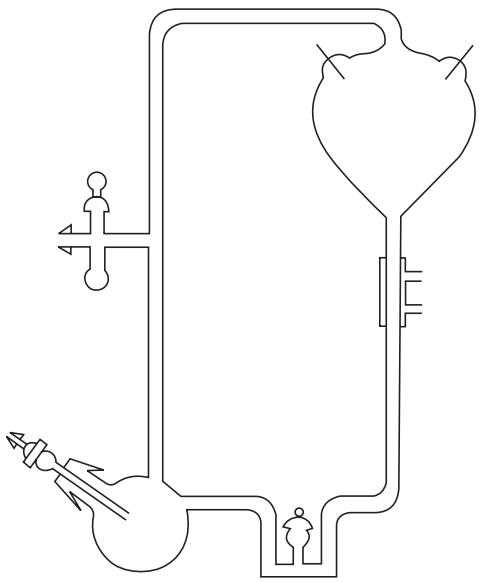


Fig. 27.1 The spark-discharge experiment of Miller.

was modified, large amounts of cyanide and aldehyde were produced during the first 125 hours of sparking. A variety of related experiments were then performed using different energy sources such as silent discharges and ultraviolet light. They yielded amino acids, purines, pyrimidines, sugars and a few other organic compounds out of inorganic precursors likely to have been present in the primordial environment.

Hydrogen cyanide (HCN) seems to be central to most reaction pathways leading to abiogenic synthesis of nitrogen containing organic compounds. Under conditions prevailing in primitive earth HCN is a precursor of compounds like cyanamide and nitriles which in turn are precursors for amino acids, purines and pyrimidines. By heating a simple mixture of an aldehyde, ammonia and HCN, a number of amino acids are formed as follows:

$$RCHO + HCN + NH_3 \rightleftharpoons \underset{\displaystyle NH_2}{\overset{\displaystyle |}{RCH}}—C\equiv N + H_2O$$

$$\underset{\displaystyle NH_2}{\overset{\displaystyle |}{RCH}}—C\equiv N + 2H_2O \longrightarrow \underset{\displaystyle NH_2}{\overset{\displaystyle |}{RCH}}—COOH + NH_3$$

It seems that most of the amino acids in the electric discharge experiments could have formed in this way.

Several energy sources in the primitive earth could have generated organic compounds from inorganic molecules. Direct sources could be sunlight, electric discharges, thermal energy of volcanoes, cosmic rays, radioactivity and shock waves. Although sunlight is the largest source of energy, its wavelength is mostly above 2000 Å which cannot be absorbed by compounds like CH_4, NH_3 and H_2O. Electric discharges are a more important source. Shock waves are produced by thunder after a lightening flash or by collisions between comets and meteorites with the atmosphere; though small in amount they are effective in activating chemical reactions.

Protenoids: Much has been speculated regarding mechanism of synthesis of complex biomolecules such as polynucleotides, polypeptides and polysaccharides from the building block molecules in the primordial soup. According to S. W. Fox prebiotic synthesis of polymers involved thermal methods of condensation. When amino acids are heated to 130 – 180°C for a few hours, or subjected to electric discharges, or warmed with condensing agents like polyphosphoric esters at 50 – 60°C, polypeptide-like polymers are formed. Fox and his colleagues have called them *protenoids.* Their particulate weight is around 20,000 daltons. By chromatography protenoids have been separated into acidic, basic and neutral types. For a particular mixture of starting amino acids used, the resulting protenoids are constant in amino acid composition. Acids hydrolyse protenoids into their component amino acids. Like proteins, protenoids are attacked by proteolytic enzymes.

Microspheres: In the experiments of Fox and his colleagues when hot solutions of protenoids are allowed to cool under appropriate conditions of pH and salt concentration, small cell-like structures called *microspheres* are formed. They are about 2 micron in diameter; under certain conditions of pH their outer boundary shows a double layer structure simulating cells membranes. When placed in hypertonic or hypotonic salt solutions, microspheres shrink or swell suggesting semipermeable nature of this membrane. Inside they may contain some amino acids and salts.

Certain cell-like behaviours of microspheres have aroused much interest. When exposed to $MgCl_2$ or changes in pH they undergo cleavage and divide; if allowed to stand for a long time (1 to 2 weeks) they can reproduce by budding. The detached buds can produce a second generation of microspheres. The formation of microspheres from thermal protenoids points to the way in which evolution of life may have taken place at one stage. It indicates how organic molecules could have become organised into groups, each a unit by itself, separated from the other and the environment by some sort of a membrane.

Further experiments of Fox have demonstrated that when protenoids are rich in lysine residues, they associate with artificially synthesised polycytosine and polyuracil, but not with poly A or poly G. In contrast arginine containing protenoids associate selectively with poly A and

poly G. This suggests that protenoids are informational molecules for selection of polynucleotides. The specificity of interaction between particular protenoids and certain polynucleotides suggests the beginnings of a genetic code.

Coacervates: The formation of coacervates is a physical phenomenon which often occurs in aqueous solutions of highly hydrated polymers. In a coacervate system an aqueous solution containing uniform distribution of a polymer separates into two phases—one with relatively high polymer concentration, other with low. The term coacervate was proposed by a Dutch chemist Bungenberg de Jong (1936). The process can be explained as follows: If in an aqueous solution containing a hydrophilic colloid of protein the concentration of colloid is increased, the colloidal particles come in close proximity. Within the outermost layer or two of these particles coalescence takes place. The coalesced particles form a continuous liquid phase that separates from the dispersion medium. This results in two liquid phases with a sharp boundary between them. The process of separation of a colloid from a medium is called coacervation. The coacervate is rich in colloidal particles, while the liquid has low concentration of colloid.

According to Oparin who has performed several experiments with coacervate systems, the droplets formed by coacervation may have entrapped a catalyst or a substrate molecule in the primitive system and acquired ability for one-reaction metabolism. Under certain conditions of Oparin's experiments, concentrated aqueous solutions of polypeptides, polysaccharides or RNA could form coacervate droplets.

Experiments of Ponnamperuma and of Oró: Using ultraviolet light as energy source, Ponnamperuma *et al* (1965) synthesised under certain conditions, combinations of amino acids and proteins. When HCN was present for example, the amino acids polymerised into polypeptide chains; when phosphoric acid was supplied, nucleotides were formed. On the other hand Oró *et al* (1965) synthesised larger macromolecules using elevated temperatures in a watery environment instead of ultraviolet rays. A number of amino acids were formed when such watery mixtures were allowed to stand for several days at temperatures ranging between 25°C to 150°C. The amino acids polymerised to form peptide chains.

THE PLANETS AND LIFE

There are reasons to believe that prerequisite conditions for birth of life may have existed in the planetary system. The reactions inside stars which produce atoms seem to be closely linked to the creation of life. The stars are continuously adding into space cosmic rays and plasma clouds containing atoms of elements necessary for life. By recent techniques radioastronomers have detected hydrogen atoms in interstellar space at an average density of one atom per cc. Much higher densities around 10 hydrogen atoms per cc as well as dust nuclei are present in clouds in interstellar medium, the temperatures of the clouds being around 10 K. Some fairly complex molecules such as CN, NH_2, H_2CO, HCN, CH_3OH, CH_3CN, HC_3N and CH_3C_2H are reported to exist in interstellar space (Snyder and Buhl, 1970), the most abundant being formaldehyde. A striking feature is the resemblance between these molecules and those synthesised in the laboratory experiments under prebiological conditions.

There is another reason for planetary exploration in connection with origin of terrestrial life. Comets and meteorites are objects emerging from space that are colliding with the earth and releasing cosmic matter. About 800 meteorites have fallen on the earth so far. The presence of organic molecules in these objects suggests a link between origin of life on earth and the interstellar medium. Moreover, the relative abundances of the basic elements are found to be similar on the

earth and on the planets. The presence of ammonia, HCN, formaldehyde, water and cyanoacetyline in dust clouds is also notable for the same reason. In fact the earth itself first originated as an astronomical body, and much later its outer layers solidified into rocks. Lastly, many scientists tend to argue that abiogenic origin of life (that is, origin *de novo* and not from preexisting life) was only possible in a reducing environment. The cosmologists have established that the atmospheres in the planets were predominantly in a reducing state.

The presence of atoms and molecules within extraterrestrial gases is detectable by spectroscopy. H.C. Urey (1952, 1959), a chemist and later astronomer was first to conceive consistent ideas about the history of the solar system. He could combine spectroscopic data on space with thermodynamic equations which must control the existence of atoms and molecules.

Meteorites and Comets

Meteorites are objects that burn up on reaching the earth emitting much heat and light. The type of meteorites classified as carbonaceous chondrites are most useful for study of evolution. They constitute about 2% of all meteorites falling on the earth. All chondrites contain chondrules which are spherical beads about 1 mm in diameter: chondrules are absent from achondrites. The carbonaceous chondrites contain 2–5% of organic carbon along with varying amounts of silicates. $MgSO_4$, sulphur and water. The contents of the famed meteorite Orgeuil which fell in France in 1864 have been analysed by modern analytical techniques. Saturated hydrocarbons have been identified by mass spectrometry, and 'organised' elements resembling bacteria are also reported. However, there is much controversy as to whether the organic compounds are of biological origin or formed abiogenically. They could also result from terrestrial contamination. Another much studied carbonaceous meteorite which fell near Murchison in Australia in 1969 has revealed organic matter in the form of amino acids. That these amino acids are not terrestrial contaminants has been deduced from their optical properties. Amino acids having asymmetric carbon atoms are present as racemic mixtures (conglomerates) of D and L forms (equivalent proportions of the two forms). The terrestrial amino acids are mostly in the L form. This indicates that these amino acids are indigenous to the meteorite. Moreover, out of the 18 amino acids identified from this meteorite, at least 6 (for example α-aminoisobutyric acid, isovaline, norvaline) do not occur or are rare in biological systems. Most of the remaining amino acids are very similar to those synthesized in the spark discharge experiments. Some aliphatic and aromatic hydrocarbons were detected in the Murchison meteorite but no purines and pyrimidines.

Comets are also large-sized space objects that contain organic molecules. The head portion of a comet ranges from 10 to 100 km in diameter and contains "ices" of CH_4, NH_3, and H_2O; the temperature inside is less than 10 K. Near the sun the surface of the comet gets heated by solar radiation and solar protons while the interior remains at low temperature. Due to this a small portion of the comet vaporizes to produce a luminescent tail millions of kilometres long. The tail contains CN, C_2, C_3, C_2H_2, CH_2, CH, NH, NH_2, OH and CO. Whether the organic matter in comets was directly involved in creation of life on earth is not known.

The Planets: Among the eight planets in our solar system, at least 6 (besides earth) have C, H and N in some form. Oxygen occurs only in water or CO_2, but *not* as free oxygen. The terrestrial planets (those nearer the sun) are smaller and contain volatile compounds like CH_4, NH_3, H_2 and H_2O. The major planets contain non-volatile compounds.

The planets Jupiter, Mars and Moon are relevant to origin of life. The atmosphere of Jupiter contains NH_3, CH_4, H_2 and H_2O similar to the primitive earth. The surface of Jupiter has a cold sheath of ammonia crystals while inside are liquid and gas at very high temperatures. A striking feature of Jupiter is the red spot believed to be due to prebiological organic molecules. Mars is the most likely planet to have life besides earth. The atomsphere contains CO_2 and less than 1%

nitrogen as well as water vapour. The Martian ice caps at the poles probably have dry ice of CO_2. A unique changing phenomenon occurs on Mars when seasonal waves of darkening spread from the polar ice caps to the equator. The red color of Mars is probably due to iron-containing minerals. There are also canals, craters and volcanoes. Actually conditions are both favourable and unfavourable for life on Mars. As much as is speculated about life on Mars, it is also uncertain if life could originate on that planet.

A low content of carbon has been shown on the moon. Lunar samples consisting of dust and rocks brought by Apollo astronauts have been thoroughly investigated in search of evidence for life on the moon. There is not much evidence for organic matter from moon samples. The amino acids were studied by sensitive techniques but could not be detected.

PREBIOTIC SYNTHESIS OF MACROMOLECULES

Evidence from mathematical models suggests that the sun and its planets began to form about 4.6 billion years ago. The early earth was covered with a vast planetary ocean. The first biochemical precellular systems, the self-replicating polynucleotides, appeared in this vast body of water.

The first polymeric macromolecules similar to those in living systems, seem to have resulted from purely geochemical processes. How this random collections of biomolecules acquired biochemical properties associated with living forms is an important question. There are no experiments to provide evidence. Since cellular life began about 3.5 billion years ago, what kind of events took place during those one billion years is difficult to envisage. The ideas expressed by scientists are based on speculation, with possible support from mathematical models and computer simulations.

The mid 1980s saw a major breakthrough when it was discovered that RNA can have catalytic activity. Laboratory experiments have shown that ribozymes carry out three types of biochemical activities, namely, self-cleavage as seen in the self-splicing Group I, II and III introns; cleavage of other RNAs; and synthesis of peptide bonds. *Synthetic* RNA molecules can also carry out in the test tube, synthesis of RNA molecules, and transfer of an RNA-bound amino acid to a second amino acid forming a dipeptide that is comparable with the role of *t*RNA in protein synthesis. The discovery of these catalytic properties gave a central role to RNA in the first biochemical reactions. Then began studies on the **RNA world**.

The RNA molecules seem to have provided templates for binding of complemenary nucleotides followed by spontaneous polymerisation. This could have resulted in *replication* of RNA molecules. Among the variety of RNA sequences generated by this process, there may be one or more with nascent ribozyme properties that would direct their own replication. Self replication would have produced a population of RNA molecules. Natural selection would then favor the more efficient replicating systems, making them predominant. This has been shown to occur in experimental systems. A central issue would be accuracy in replication that would allow the RNA population to acquire sequence specificity and potential for further catalytic properties. Having self-replicating RNA molecules capable of performing simple biochemical activities, the next step would be to visualize how the molecules could become enclosed in lipid membranes forming cell-like structures.

Evolution of DNA molecules: The RNA molecules would have been subjected to processes that eventually gave rise to DNA molecules. In the first stage, the development of protein enzymes may have taken place to substitute the catalytic properties of ribozymes. How the transition from RNA to protein could have occurred has raised questions that cannot be answered. Possibly, the

proteins became coding molecules that could specify catalytic proteins. This creates two scenarios for the evolution of the first coding RNA. A ribozyme could have evolved to become the coding molecule, so as to have a dual catalytic and coding function; or coding molecules could have been synthesised by ribozymes. The theories about the origins of translation and the genetic code suggest that ribozymes were probably synthesising the coding molecules.

In the next stage, we have to envisage transfer of the coding function to the DNA molecules. This would involve reduction of ribonucleotides giving deoxyribonucleotides, which could then be polymerised into copies of RNA by a reaction catalysed by reverse transcriptase. Uracil would have been replaced by its methylated derivative thymine, providing additional stability to DNA. The DNA molecules performing coding function would have become double-stranded duplexes in response to the possible requirement for repair of DNA damage by copying the partner strand.

The first DNA molecules can be visualised as specifying a single protein, each therefore representing a single gene. How the single genes became linked to form separate chromosomes is not understood, though several different mechanisms have been proposed.

The first cells which were similar to present day bacteria are believed to have evolved 3.5 billion years ago. The first cells may have had double-stranded DNA molecules and possibly some linked genes in the form of one or two chromosomes. Evidence from fossil records suggests that the first eukaryotic cells arose about 1.4 billion years ago, in the form of structures resembling single-celled algae. The first multicellular algae came into existence about 0.9 billion years ago. Around 640 million years ago, multicellular animals seem to have appeared. During the Cambrian period, about 530 million years ago, invertebrate life seems to have proliferated into many new forms. The mass extinction which occurred 500 million years ago wiped out most of the new forms. Evolution thereafter, is seen to have established insects, animals and plants by 350 million years ago. The dinosaurs are said to have disappeared 650 million years ago, while the hominids appeared 4.5 million years ago.

The evolution of living forms was accompanied by an increase in *complexity of genomes*. Gene number increased from about 1000 in bacteria to 80,000 in vertebrates and humans. Instead of a gradual increase, there seem to have been two dramatic stages of expansion in gene numbers. The first expansion occurred about 1.4 billion years ago with the appearance of eukaryotes, when the few thousand genes present in prokaryotes increased to above 10,000 in the simplest eukaryotes. The second stage of expansion probably occurred when the first vertebrates appeared, with acquisition of about 50,000 genes. There are two possible ways in which new genes could be acquired by a genome: A. By **gene duplication**; B. by **acquisition of genes** from other species.

A. Gene Duplication

The duplication of existing genes could occur in one of two possible ways: by duplication of an entire genome; by duplication of a single gene or more than one gene. **Whole genome duplication** can result from an error during meiosis which produces gametes that are *diploid* instead of haploid. The fusion of two diploid gametes will result in an **autopolyploid** individual containing four copies of each chromosome (details in Chapter 11). Autopolyploids are common in plants and can reproduce successfully. But they cannot be crossed with the original diploids since they would result in triploid offspring in which the third set of chromosomes would not find homologous partners for pairing. Autopolyploidy is a mechanism by which speciation can occur.

An autopolyploid individual has extra *copies* of each gene but no new genes. The extra genes can undergo mutational change without harming the viability of the organism. The resulting nucleotide changes in many genes could produce harmful mutations that would be eliminated;

the other mutations would lead to new gene function that would be useful to the cell. Genome of the yeast *Saccharomyces cerevisiae* seems to be the product of a duplication that occurred about a 100 million years ago. The same may be true of the genomes of some vertebrates.

Gene Amplification

When a gene is duplicated, two new copies of that gene are produced. The selection process then ensures that one of these genes retains the original sequence and hence the function of the original gene. The second copy of this gene can accumulate mutations at random. There is evidence that the mutations acquired by the second copy of the gene are largely deleterious, and lead to inactivation of the gene (*pseudogenes*). Some of these mutations however, are not deleterious and do not inactivate the gene, but instead, lead to new gene function that is useful. Thus we can say that when a gene is duplicated, the new copies tend to evolve different functions. Gene amplification increases the complexity and size of the genome.

Genes related by function evolved by gene duplication. These constitute multigene families that have been revealed as common components of all genomes, by DNA sequencing methods (Chapter 22). It is possible to trace the individual gene duplications involved in the evolution of the gene family from a single original gene that existed in an ancestral genome, by sequencing individual genes in the family. Examples of related genes that evolved by gene duplication include elongation factors of translation, the cytochrome P450 oxidases and the opsins (light sensitive proteins in cone cells of retina). The globin gene family in vertebrates evolved by gene duplication from myoglobin as ancestral globin. Myoglobin is a muscle protein whose main function is storage of oxygen. The haemoglobins (α, ζ, β, δ, γ and ε) evolved by successive gene duplication, starting with myoglobin. These new globin genes are used at different stages in development of humans. The trypsin and chymotrypsin genes are also related by a common ancestor, approximately 1500 million years ago. Now these genes code for proteases involved in protein breakdown in the digestive tract of vertebrates, trypsin breaking the amino acids arginine, while chymotrypsin breaks phenylalanines, tryptophans and tyrosines.

Gene Conversions

There are multigene families made up of genes with almost identical sequences. A good example is the rRNA genes whose copy number could range from two in a *Mycoplasma* species to more than 500 in amphibian *Xenopus*. These multiple copies are required to fulfil the need for rapid synthesis of the gene product at a certain stage in the life cycle, oocyte in the case of *Xenopus*. There could be a mechanism that prevents mutations in these gene copies and acquisition of new function. Members of the rRNA gene families seem to undergo **concerted evolution**, so that their sequences not only remain the same, but also evolve in parallel. If one copy of the gene family acquires a useful mutation, then it is possible for that mutation to spread throughout the family into all members of the family. This could be accomplished by the method of **gene conversion**, which can result in the sequence of one copy of a gene being replaced with all or part of the sequence of another copy. Thus, sequence identity of all members of the gene family would be maintained.

Gene Rearrangements

Most proteins are made up of *structural domains* represented by segments of the polypeptide chain, hence encoded by a contiguous series of nucleotides. It is possible to generate novel protein function by bringing about changes in existing genes. This can be accomplished if gene segments that encode *domains* in the polypeptide chain can be *rearranged.* There are two models of domain duplication: domain duplication and domain shuffling.

Domain duplication occurs when the gene segment coding for a structural domain is duplicated by unequal crossing over or any other method. Thus, a specific structural domain is repeated in the polypeptide chain adding stability to the protein. Domain duplication causes the gene to become longer. The genes of higher eukaryotes are longer, on average, than those of lower organisms. Hence gene elongation could be a general consequence of genome evolution.

Domain shuffling occurs when segments coding for structural domains from completely *different* genes become joined together to form a *new* coding sequence, which will specify a hybrid or mosaic protein. This new protein would have a novel combination of structural features and might provide the cell with entirely new biochemical function.

Mechanisms that could produce gene duplications include *unequal crossing over*; *unequal sister chromatid exchange*; and *DNA amplification* (refers to gene duplication in bacteria).

Unequal Crossing Over : The most common method by which genes are duplicated is by unequal crossing over in meiosis. When homologous chromosomes are misaligned, and there is an exchange at the point of misalignment, the process yields one recombinant with a duplication and another recombinant with a deletion (Chapter 8). The genes encoding red-absorbing opsin gene (OPSR) and green-absorbing opsin gene (OPSG) arose by gene duplication, and lie adjacent on the X chromosome (Xq28). The proteins of these two genes are 96% identical in amino acid sequence. In humans, all normal X chromosomes have one copy of OPSR, but may have one, two or three tandem copies of OPSG. Red-weak and green-weak vision can be caused by mutation or unequal crossing over.

Unequal Sister Chromatid Exchange: Sister chromatid exchanges (SCEs) are induced when both DNA strands in a chromosome undergo breakage, followed by an exchange of whole DNA duplexes (Fig. 27.2). Recall that a metaphase chromosome has two chromatids, and there is a single duplex DNA molecule present in one chromatid, hence two duplexes in a chromosome. This occurs during S phase and is usually induced by exposure to some form of toxicity. Unequal sister chromatid exchanges result in gene amplification and have been proposed as one of the several possible mechanisms resulting in tandemly repeated sequences in chromosomes.

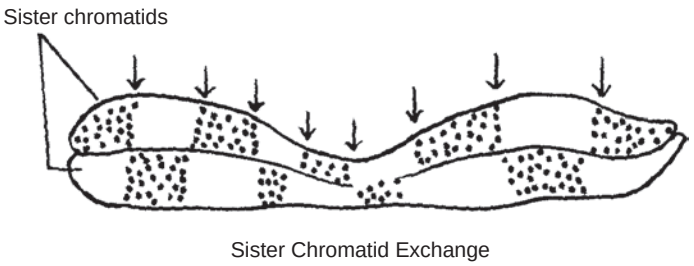


Fig. 27.2 Chromosome from culture of lymphocytes subjected to Bromodeoxyuridine treatment displaying sister chromatid exchanges; arrows indicate points of breakage.

DNA Amplification: In oocytes of amphibia and some insects, the gene for 18S and 28S ribosomal RNA present in the nucleolus organiser region (NOR) of a specific chromosome are increased. The amplification allows the oocyte to synthesise very large quantities of ribosomal RNA required at this stage. Besides oocytes, there is no amplification of these genes in somatic cells.

B. Acquisition of Genes from other Species

In bacteria, the process of DNA transformation apparently had an influence on genome evolution. Only a few bacteria such as *Bacillus*, *Streptococcus* and *Pseudomonas* have efficient mechanisms for the uptake of DNA from the surrounding environment. It has been argued that on an

evolutionary timescale, all species have probably acquired at least a few genes through this route, so that transformation might have been a major factor in genome evolution. Prokaryotic genomes that have been completely sequenced by modern methods provide evidence indicating that widespread exchange of genes has been taking place between species. For example, the bacterium *Aquifex aeolicus* has exchanged genes with *Methanococcus jannaschii*, and *Helicobacter pylori* has exchanged genes with some distantly related bacteria. These gene transfers are thought to have occurred by transformation.

Some plants present clear cases of **allopolyploidy** which results from interbreeding between two different species. The hybrids produced are viable. Usually, two closely related species form allopolyploids. As such they have many genes in common, and each parent will possess a few novel genes, or distinctive alleles of the shared genes. As described in Chapter 11, *Triticum aestivum*, the hexaploid bread wheat arose by hybridisation with the tetraploid cultivated emmer wheat, and a diploid wild grass *Aegilops squarrosa*. The presence of novel and superior alleles for the high molecular weight glutenin genes (glutenin binds the dough well) in *Aegilops* combined with glutenin alleles already present in emmer wheat resulted in superior bread-making properties now found in the hexaploid wheats.

The Origins of Introns

The origins of introns have been debated for last many years. The self-splicing Group I, II and III introns are believed to have evolved in the RNA world, described earlier. They have remained as such over the years without undergoing much change. But the origins of the GU-AG introns that are found in large numbers in the eukaryotic nuclear genomes are still unclear. According to two opposing hypotheses, introns are very ancient and are gradually being lost from eukaryotic genomes (**intron early** hypothesis). The other, **intron late** hypothesis states that introns evolved relatively recently and are gradually accumulating in eukaryotic genomes. There are several different models for each hypothesis , but there is hardly any evidence to support either hypothesis.

The Evolution of Proteins

A few thousand kinds of proteins are required for the functioning of an *E. coli* cell and about a million kinds are required by human beings. Proteins offer a better field for evolutionary study than other macromolecules like nucleic acids because they are more heterogeneous both structurally and functionally, and are easy to isolate for analysis. Recently a number of proteins have been characterized by the method of sequence analysis. Sequence data are now available from diverse biological groups ranging from micro-organisms to mammals. Comparison of sequences allows study of interrelationship between structure and function, and to deduce how proteins have evolved. The laws governing evolution of proteins are similar to those for heritable traits, and interrelatedness of different organisms points toward their descent from a common ancestor. In general the following points have been noted:

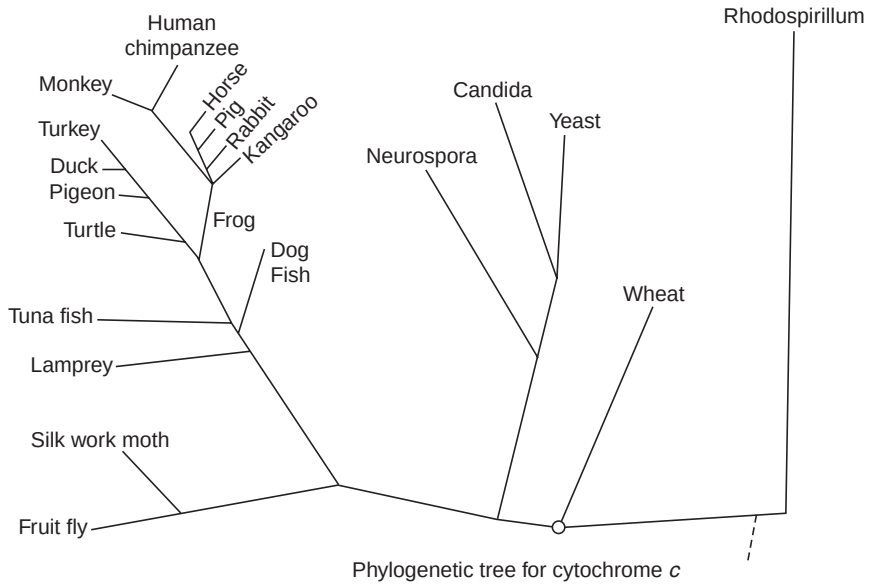
(*a*) Identical proteins are not found among different living species. But homologous proteins with some similarities occur in diverse organisms.

(*b*) Different positions in the amino acid sequence vary with respect to the number of amino acid substitutions that can take place without impairing the function of the polypeptide chain. Thus some positions allow more substitutions than others.

(*c*) Due to the tightly packed three-dimensional conformation of the protein, a change undergone by an amino acid situated in the interior of the chain is likely to affect residues in neighbouring positions. Such evolutionary changes therefore, exist in pairs.

(*d*) Enzymes performing similar functions have similar structure.

The protein cytochrome *c* functions in electron transfer. A related protein occurs in the photosynthesis bacterium *Rhodospirillum rubrum* where it takes part in photosynthetic electron transfer. About 30% of all the amino acids in cytochrome *c* are identical in *R. rubrum* and in humans; 60% of amino acids in human cytochrome *c* are identical with those in the wheat plant (Dayhoff, 1971). In various groups of living organisms certain positions in the polypeptide chain of cytochrome *c* contain the *same* amino acids, while some other positions are filled by amino acids with similar chemical properties. Variations in amino acid sequences have been studied for determining the phylogeny of the species; a phylogenetic tree has been constructed for cytochrome *c* (Dayhoff, 1971).



Phylogenetic tree for cytochrome *c*

The connection of the trunk to the branching structure is estimated to represent earliest time. From here the length of time is proportional to the lengths of the branches. The divergences of the various groups are visible from the manner of branching of the tree.

The haemoglobins are essential respiratory pigments in man and vertebrates. The human haemoglobins have various mutant forms due to altered amino acid sequences in the β or α chains. In some cases it is possible to correlate abnormalities in erythocyte function with specific defects in the chains. Evolution of haemoglobins and myoglobins (present in muscle) has been studied in detail. It seems there was duplication of genes coding for globins in the last billion years (Zuckerkandl and Pauling, 1965). The present myoglobin is the result of gene duplication in very ancient times. The genes for α and β chains of haemoglobin which are on different chromosomes in living organisms must have descended from a single ancestral gene by duplication in the remote past. Another duplication must have occurred later (more recently) producing separate genes coding for the β chain and for the γ chain. Investigations show that divergence of myoglobin and haemoglobin took place about a billion years ago, in the Precambrian. Further evolution then seems to have taken place in the haemoglobin line. Statistical analysis of amino acid sequences of haemoglobin shows much greater differences between haemoglobin of fish and of mammals (human, bovine) than between the mammals themselves.

The evolution of the primitive protein ferredoxin is interesting as it relates to the early evolution of life. Ferredoxin is a hydrogenase present in bacteria, blue green algae as well as higher plants. The bacterial ferredoxin for example, is a short polypeptide containing only 13 of the simple amino acids including 8 essential cysteins in its total of 55 amino acids. The protein is bound to iron and sulfur to form $Fe^{II}S$ which is the reactive center. Ferredoxin is the most electronegative metabolic enzyme known, with a potential close to that of molecular hydrogen. It takes part in a number of biochemical processes of a plant cell like photosynthesis, sulphate reduction, nitrogen fixation and oxidation-reduction reactions. It is surmised that ferredoxin may have performed these functions in the primitive cell. It seems there has been a gene duplication in bacteria and a separate duplication in plants resulting in the present ferredoxin.

FROM MICRO-ORGANISMS TO MAN

That the earliest living organisms were prokaryotes is based on the discovery of microfossils from early precambrian rocks. The first animals were single-celled protozoa. The worms and sponges seem then to have evolved in the seas. Then came the hard-shelled crustaceans which left many fossils of their bony armour. The first land plants (Psilopsids and Lycopsids) appeared in the Silurian of Palaeozoic followed by forests of gymnosperms, seed ferns and bryophytes. The earliest amphibians arose in the age of the forests. The flowering plants came later in the Mesozoic. The earliest jawless vertebrates appeared in early Palaeozoic about 300 million years ago. The fishes, amphibians, and reptiles also arose, in that order, during the Palaeozoic. The reptiles ruled the earth for many millions of years. During the Mesozic giant like reptiles, the *Dinosaurs* were dominating the earth for over a million years, then became extinct. From the reptiles the birds and mammals seem to have diverged along two separate lines, the mammals culminating in man.

The lowermost animals with backbones on the evolutionary scale belong to the chordates possessing either a notochord, or a vertebral column. The first chordates are not preserved as fossils as they lacked a bony armour. The present day sea lancelet *Amphioxus* is perhaps a living representative of the primitive chordates. It has a notochord, a nerve chord and a digestive tube; there is no head, brain or sense organ, except pigment spots sensitive to light. The earliest vertebrates were jawless of which the living form *Lamprey* is a good example. The fossil record shows how amphibians and reptiles evolved. The first mammals seem to have descended from mammal-like reptiles in the Jurassic. These four-footed mammals with uniform body temperature and high basic metabolism are also referred to as warm-blooded animals.

The last to come in evolution of vertebrates are the primates including the lemurs, tarsiers, monkeys, apes and man. The most primitive lemur is *Tupaia*, a tree shrew about the size of a squirrel. Interestingly, it is a placental animal with large-sized brain that shows affinities with both insects and primates. The more advanced primates are comprised of monkeys and apes. The apes have no external tail and have evolved as brachiating animals, swinging on trees by their arms. This type of locomotion is predominant among the apes. The modern apes include the gibbons, orangutan of East Indies and the chimpanzees and the gorillas (the largest in size) of Africa. Even larger than the modern gorilla was the giant ape *Gigantopithecus* that lived in Asia during the Pleistocene and finally became extinct.

The earliest organism resembling human beings was *Australopithecus*, an erect, biped walking, tool making creature discovered from cave deposits of Pleistocene in S. Africa more than 5 million years ago. It had skull, brain and face like an ape while dentition was similar to humans.

After existing for a long time Australopithecines or Ape man evolved into *Homo erectus* with double the size of brain, better bipedalism and skills. In the last 50 to 55 thousand years *Homo erectus* seems to have given way to *Homo sapiens* the present day man. The evolutionary development of man took place mainly along the following lines: (*a*) growth of brain, (*b*) erect posture, (*c*) slow rate of development after birth, (*d*) complexities of social life and behaviour, (*e*) growth in human population, and (*f*) technical advancement.

The size of brain in the modern ape is about 500–600 gram, which is about the same as the cranial capacity of *Australopithecus*. In the primitive men of the Pleistocene, this capacity increased to 900 gram. In modern man the brain size ranges from 1200 gram in the minimum to 2000 gram in the maximum. It seems the size of the brain increased in relation to mental development. In the evolution of man, the brain case has been getting bigger in size compared with the rest of the skull, and the jaws have become proportionately small.

The development of erect posture has allowed man to use two free hands for other skills and has changed the shape of the spinal column. Due to slow rate of growth, the young one of man matures in about two decades after birth while a gorilla matures in less than 10 years. Lastly, with increase in mental powers man has gained ability for perfection in social, behavioural and technical skills.

Evolution of the Human Genome

Evidence suggests that the closest relative of humans among the primates is the chimpanzee. The most recent ancestor that we share among the chimpanzees lived about 4.5 million years ago. The human lineage from the chimpanzee was represented by two genera, namely, *Australopithecus* and *Homo* species. Some species of *Homo* directly descended to *Homo sapiens* that we are. Though we are a novel species with distinctive traits, we do share some common features with the chimpanzee genome. Evidence suggests that we are 1.5% different from the chimpanzee, with respect to the nucleotide sequence dissimilarity between humans and chimpanzees. Many of the coding genes in the two genomes are identical, so that with respect to the coding DNA, we are less than 1.5% dissimilar. Even in the non-coding regions, the dissimilarity is about 3%.

Very few clear differences have been found between human and chimpanzee genomes. Some of these pertain to gene duplications. Some recent gene duplications are believed to have occurred that are present only in one or the other genome. These are described as human-specific and chimpanzee-specific gene duplications. The other set of differences have arisen out of extensive divergence of some components of the noncoding DNA in the two genomes. These include some DNA sequences in the human chromosome centromeic region which have become quite different from those in chimpanzee and gorilla chromosomes. Some *Alu* elements present in the human genome are markedly different from those in the chimpanzee.

A comparison of chromosome banding patterns reveals a few rearrangements in human and chimpanzee genomes. The most remarkable finding is that human chromosome 2 is the product of a fusion between two chimpanzee chromosomes. In other words, human chromosome 2 exists as two *separate* chromosomes in chimpanzee. Therefore, chimpanzees and other apes have 24 pairs of chromosomes, while humans have 23 pairs. Four other humans chromosomes, namely, 5, 6, 9, and 12 show visible differences from corresponding chromosomes in chimpanzee. The remaining 18 chromosomes are almost identical in humans and chimpanzees. Although some of the differences are relevant for genome evolution, they do not seem to throw light on what makes us different biologically from the apes. This question is intriguing molecular biologists, and the answers may be expected when the genome of an ape is sequenced in future.

Phylogeny

All organisms are believed to have arisen from ancestral organisms, implying a single ancestor in the beginning. To understand evolutionary lineages of organisms, a tree-like diagram is constructed. The trunk of the tree represents the original ancestor; the organisms living today can be arranged on branches of the tree. The pattern of branching in the evolutionary tree is **phylogeny**. To get an idea about phylogenetic relationships between organisms, nucleotide sequences of **homologous genes** can be compared.

Genes in different species that descended from a common ancestral gene are said to be homologous. There are two kinds of homology, **orthology** and **paralogy**. Orthologs are copies of a gene, present in two species and inferred to have existed in their most recent common ancestor. For example, the β-haemoglobin genes in mouse and humans are orthologs. Paralogs are copies of a gene that arose by gene duplication, for example, α-hemoglobin and β-hemoglobin. The differences in nucleotide sequences of homologous genes have been attributed to substitutions, deletions, and insertions that occurred within lineages.

Phylogenetic Trees

Evolutionary relationships between species and genera are depicted in a **phylogenetic tree.** It may be noted that a tree attempts to represent *evolutionary history* and *not* a taxonomic classification based on characteristics of the organisms shown in the tree. The branching pattern of the tree is *dichotomous*, and the branches could end in a *pair* of forks, not triplets. Such a branching pattern conveys the idea that speciation is bifurcative. In other words, each speciation event yields exactly two daughter species.

The construction of a **phylogenetic tree** is based on the following assumptions: that one ancestral organism, extinct since long, gave rise to all present day species; a single common ancestor gave rise to each pair of species; organisms become more and more different genetically through the evolutionary time after divergence. DNA sequences of genes, amino acid sequences of proteins, and clear cut phenotypes (morphological traits) are the three widely used types of information for tree construction.

The tree consists of nodes connected by branches; the nodes at the ends of branches are living species (extant species), while internal nodes represent ancestral organisms. The tree is rooted if an inferred ancestral node is included; all paths to extant species branch from the root. The tree is devoid of a root if species are merely connected, but no ancestral node is included.

Building a tree is complex and uses methods that are well beyond the scope of this book. The basic principles are briefly outlined here, using nucleic acid sequences. Species are chosen for phylogenetic analysis, including a species outside the group being studied (an outgroup). Each species must have the set of genes used in the analysis.

There are two approaches to studying the evolutionary relationships of living organisms using DNA sequences, namely, **phenetic** and **cladistic**. Phenetics is based on the degree of similarity among DNA sequences. Cladistic is based on shared nucleotides that differ from the ancestral nucleotides in a sequence.

In the phenetic approach, evolutionary distances between species are calculated by counting the proportion of nucleotides in a sequence that differ between *pairs* of species. From that proportion a distance statistical value is derived. A matrix is then prepared with the species as both rows and columns, and distances are entered in the cells of the matrix. The tree can then be constructed by connecting species so that the distances in the branches of the tree correspond to the distances in the matrix.

The cladistic approach uses one of the two methods, *parsimony* or *maximum likelihood*. The parsimony method finds a tree requiring the minimum number of evolutionary changes necessary to produce the minimum total number of substitutions. The maximum likelihood method makes a statistical model based on trees, and the tree that makes for the model of greatest likelihood is chosen.

It is not possible to detect the existence of multiple substitutions and horizontal gene transfer between species. For example, suppose a GC pair undergoes two substitutions, GC - AT – GC; it will appear that no change has taken place, the two changes will not be detected. If a bacterial gene is transferred from species A to a species B of Archaea (primitive group of prokaryotes) by transformation, then it will appear as if the descendents of Archaea are more closely related to bacteria than they truly are.

DARWIN AND NATURAL SELECTION

Any story of evolution is incomplete without Darwin's theory of *Natural Selection.* During his voyage round the world aboard the Beagle, Darwin made a halt at Galapagos Archipelago on the coast of Ecuador in the Pacific Ocean. Here he studied the flora and fauna on a group of islands isolated from the main continent of South America. He noticed the variation in the finches and the giant tortoises, each island possessing a distinct variety. The variations made the forms distinct enough to be regarded as separate species. Yet Darwin was impressed that the finches for instance, formed a homogeneous group and he thought that they may have descended from a common ancestor. Darwin also listed a large number of plant species on the four islands. He noticed variation in them and could determine that many of them were not present anywhere else in the world; the majority of plants were endemic to a single island in Galapagos.

From his detailed study Darwin concluded that the evolutionary process was producing new varieties of plants and animals. When finches and tortoises came to Galapagos islands they evolved in a different direction to other similar populations. The increase in population of a species is due to reproductive capacity, survival of young ones, low rate of mortality, and absence of competition. Darwin held the view that mortality took place so that the less adapted would not survive. Darwin accounted for evolutionary change through this system. In natural selection the environment determines the efficiency of a species by eliminating the less fit individuals. The best adapted individual survive and their offsprings inherit these capabilities. Thus when a new population reaches a habitat, it will gradually become adapted to its new environment. This is the basic idea of evolution by natural selection. The rediscovery of Mendel's work has revised Darwin's theory of evolution by natural selection (neo-Darwinism) but without altering the basic underlying principles.

SELECTED READINGS

Axelrod, D.E., Baggerly, K.A. and Kimmel, M. 1994. Gene Amplification by Unequal Sister Chromatid Exchange: Probabilistic Modelling and Analysis of Drug Resistance Data. *J. Theoret. Biol.* 168 (**2**) : 151–159.

Bernal, J.D. 1951. The Physical Basis of Life. Routledge and Kegan Paul, London.

Blum, H.F. 1968. Times Arrow and Evolution. Princeton Univ. Press, Princeton. 3rd edition.

Brookfield, J.F. Y. 1997. Genetic Redundancy. *Adv. Genetics* **36** : 137–155.

Buvet, R. and Ponnamperuma, C. (eds). 1971. Molecular Evolution. 1. Chemical Evolution and the Origin of Life. North Holland, Amsterdam.

Calvin, M. 1969. Chemical Evolution. Oxford Univ. Press, New York.

Dodson, E.O. 1960. Evolution Process and Product. Reinhold Publ. Corp., New York.

Doolittle, R.F. 1987. The Evolution of the Vertebrate Plasma Proteins. *Biol. Bull.* **172** : 269–283.

Fitch, W.M. 2000. Homology. *Trends Genet*. **16** : 227–231.

Fox, S.W. 1965. The Origin of Prebiological Systems. Academic Press, New York.

Gibbons, A. 1998. Which of Our Genes Makes us Human? *Science* **281** : 1432–1434.

Gilbert, W. The Exon Theory of Genes. *Cold Spring Harb. Sympos. Quantit. Biol*. **52** : 901–905.

Gupta, R.S. and Golding, G.B. 1996. The Origin of the Eukaryotic Cell. *Trends Biochem. Sci.* **21** : 166–171.

Hartl, D.L., Lohe, A.R. and Lozovskaya, E.R. 1997. Modern Thoughts on An Ancient *Marinere*: Function, Evolution, Regulation. *Annu. Rev. Genetics* **31** : 337–358.

Lynch, M. and Conery, J.S. 2000. The Evolutionary Fate and Consequences of Duplicate Genes. *Science* **290** : 1151–1155.

Miller, S.L. 1960. Formation of Organic Compounds on the Primitive Earth. *In* Pure and Applied Biology. Ed. M. Florkin. Pergamon Press, Oxford, New York.

Miller, S.L. and Orgel, L.E. 1974. The Origins of Life on the Earth. Prentice Hall, Inc. New Jersey.

Oparin, A.I. 1938. The Origin of Life. Macmillan, New York.

Parkin, D.T. 1979. An Introduction to Evolutionary Genetics. Edward Arnold. UK.

Ponnamperuma, C. 1972. The Origins of Life. Thames and Hudson, London.

Rutten, M.G. 1971. The Origin of Life by Natural Causes. Elsevier, Amsterdam.

Scherer, S. 1990. The Protein Molecular Clock. *Evol. Biol*. **24** : 83–106.

Snyder, L.E. and Buhl, D. 1970. *Sky and Telescope* **40** : 267.

Urey, H.C. 1952. The Planets. Yale Univ. Press, New Haven.

Urey, H.C. 1952. The Early Chemical History of the Earth and the Origin of Life. *Proc. Natl. Acad. Sci.* US **38** : 351.

Wald, G. 1954. The Origin of Life. *Sc. Amer.* **191** : 44.

Wolfe, K.H. and Shields, D.C. 1997. Molecular Evidence for An Ancient Duplication of the Entire Yeast Genome. *Nature* **387** : 708–713.

Index

SYMBOLS

A

B

C

D

E

F

G

H

I

K

L

M

N

O

P

R

S

T

U

V

W

X

Y

Z