

P F SMITH-KEARY

**GENETIC
STRUCTURE
AND
FUNCTION**



GENETIC STRUCTURE AND FUNCTION

P. F. SMITH-KEARY

Lecturer in Genetics
Trinity College, Dublin

M

© P. F. Smith-Keary 1975

All rights reserved. No part of this publication
may be reproduced or transmitted, in any form or by any
means, without permission.

First published 1974 by
THE MACMILLAN PRESS LTD
London and Basingstoke
Associated companies in New York Dublin
Melbourne Johannesburg and Madras

SBN 333 15663 3 (hard cover)
333 17282 5 (paper cover)

Printed in Great Britain by
The Whitefriars Press Ltd
London and Tonbridge

Distributed in the United States and Canada by
Halsted Press, a Division of
John Wiley & Sons, Inc., New York

Library of Congress catalog card no. 73-11882

The paperback edition of this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

Preface

This book was written partly because of an inner motivation and partly, and more important, because of an awareness of the lack of a suitable text for second and third-year courses in molecular genetics at Trinity College, Dublin. There are many good textbooks on general genetics but these tend to concentrate on Mendelian and neo-Mendelian genetics and lead on to a study of natural selection and evolution. On the other hand the text books on molecular genetics take a more biochemical approach and, in general, ignore many important aspects of basic genetics and genetic mapping, even though it is not possible, for example, to understand the rationale behind hybrid DNA models for genetic recombination unless one is fully conversant with tetrad analysis, or the relationships between the different control genes in bacteria without a knowledge of their genetic systems. This introduction to molecular genetics, based on lectures given to our second and third year students, is written from the point of view of a geneticist and attempts to achieve a better balance between the genetic and the biochemical evidence.

The size of any book limits the range of material that can be adequately described and so the choice of examples becomes largely a matter of personal preference. Several important aspects of basic genetics (for example inheritance in man, continuous variation and cytogenetics) receive only a brief mention since they are not necessary for a proper understanding of the aspects of molecular genetics described in this book. Likewise, only a few selected examples of genetic control in higher organisms are highlighted and some of the more important problems indicated. This is not because these systems are any the less interesting or less important than the better known control systems in procaryotes, but because they are not only strikingly different but also relatively poorly understood; many of the conclusions in this wide and rapidly expanding field are still tentative and cannot be adequately discussed in a book of this size and at this level.

In general, I have tried to trace the development of each branch of molecular genetics, to discuss each experiment in its correct historical perspective, wherever possible describing experimental procedure and presenting actual data or results, and to present a picture of molecular genetics as it is today.

I have not attempted to cite original references to all the experiments described, but the bibliography, in addition to some suggestions for further general reading, lists a number of now classic papers, which the student should find intelligible, informative and intellectually satisfying.

I am much indebted to the members of the Department of Genetics in Trinity College, Dublin, who have read and commented on many parts of the manuscript, to Dr Keith Jones, who prepared the photographs reproduced in figure 4.4. especially for this book, and to Dr Oscar Miller (jr) for permission to reproduce the electron micrographs shown in figure 18.6.

*Trinity College,
Dublin
April 1974*

P. F. SMITH-KEARY

Abbreviations and Symbols

A	adenosine
Ala	alanine
Ala-tRNA ^{Ala}	alanine transfer RNA charged with alanine
AMP	adenosine 5'-monophosphate
AP	2-aminopurine
Arg	arginine
Asp	aspartic acid
Asn	asparagine
ATP	adenosine 5'-triphosphate
<i>B-B'</i>	prophage recognition sites on the bacterial chromosome
BUdR	5-bromodeoxyuridine
BU	5-bromouracil
C	cytidine
cAMP	adenosine 3':5'-cyclic monophosphate
Cys	cystine
Δ	chromosomal deletion
d	prefix 'deoxy'
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
F ₁	first filial generation
F	the <i>Escherichia coli</i> sex factor
fMet	formyl-methionine
G	guanosine
GTP	guanosine 5'-triphosphate
Gln	glutamine
Glu	glutamic acid
Gly	glycine
HA	hydroxylamine
Hb	haemoglobin
Hfr	high-frequency recombination
HNI	high negative interference
His	histidine
i	initiator site for replication
I	inosine
Ile	isoleucine
J	joule; practical unit of electrical energy
Leu	leucine
Lys	lysine

μm	micrometre (10^{-6} metres)
Met	methionine
MR-DNA	middle repetitive DNA
mRNA	messenger RNA
NA	nitrous acid
NG	N-methyl-N'-nitro-N'' nitrosoguanidine
nm	nanometre (10^{-9} metres)
<i>O</i>	operator region
ϕ	bacteriophage
p	phosphate group
<i>P</i>	promoter region
pC, Cp	nucleotides of cytidine ending with a 5' or a 3' phosphate respectively
<i>P-P'</i>	sites on a phage chromosome that recognise <i>B-B'</i>
Phe	phenylalanine
Pro	proline
RG	regulator gene
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal RNA
<i>S</i>	Svedberg unit, the sedimentation coefficient
Ser	serine
SG	structural gene
T	thymidine
<i>t</i>	terminator
Thr	threonine
TMV	tobacco mosaic virus
tRNA	transfer RNA
tRNA ^{Ala}	the transfer RNA for alanine (uncharged)
Trp	tryptophan
Tyr	tyrosine
U	uridine
uv	ultraviolet radiation
Val	valine

Contents

<i>Preface</i>	xiii
<i>Abbreviations and symbols</i>	xv
1. This is Genetics	1
A Mendelian view of genetics	1
A molecular view of genetics	5
The birth of genetics	9
2. The Material Basis of Heredity	13
The Genetic Material is DNA (or sometimes RNA)	13
Transformation	14
Reconstituting a virus	15
The chemical structure of the nucleic acids	18
The biological structure of DNA	19
The Replication of DNA	24
The Watson and Crick model for the replication of DNA	24
Density-gradient centrifugation	25
The Meselson and Stahl experiment	27
Discontinuous replication	28
Summary of Key Points	29
3 The Chromosomes of Bacteria and Viruses	30
The Chromosome of <i>Escherichia coli</i>	31
Autoradiography	31
The Chromosomes of Bacteriophages with Double-stranded DNA	35
DNA/DNA hybridisation	38
Circular permutation and terminal redundancy	39
Cohesive ends	40
The Chromosome of X174—a Phage with Single-stranded DNA	41
The enzymatic synthesis of DNA	41
The rolling-circle model	44
Summary of Key Points	46
4 The Chromosomes of Eucaryotes	48
Cell division	49
Meiosis or Reduction Division	53
The first meiotic division	53
The second meiotic division	54
The Chemical Composition of Chromosomes	56
The Structure of the Eucaryotic Chromosome	58
Polytene chromosomes	61
Summary of Key Points	63

5	Genetic Analysis in Higher Organisms	64
	<i>Drosophila</i>	64
	Maize	64
	Independent segregation	66
	Independent assortment	69
	Linkage	72
	Two-point linkage	74
	Three-point linkage	77
	Genetic Interaction	81
	Pigmentation in the aleurone of maize	81
	Continuous variation	84
	Sex Linkage	85
	Sex determination	86
	Sex linkage	88
	Sex linkage in man	89
	Summary of Key Points	89
6	Genetic Analysis in <i>Neurospora</i> and <i>Saccharomyces</i>	90
	Life cycles	90
	Tetrad Analysis in Yeast	92
	Tetrad Analysis in <i>Neurospora crassa</i>	95
	Independent assortment	98
	Linkage	99
	Extra-chromosomal Inheritance	102
	Summary of Key Points	104
7	Genetic Transfer in Bacteria: General Transduction and the Analysis of Genetic Fine Structure	105
	The nomenclature of bacterial genetics	108
	General Transduction	109
	Abortive transduction	111
	Transduction Analysis in <i>Salmonella typhimurium</i>	113
	Complementation analysis	113
	Two-point transduction	115
	The joint transduction method	116
	The three-point transduction method	118
	Deletion mapping	119
	Summary of Key Points	121
8	Genetic Transfer in Bacteria: Conjugation and Recombination in <i>Escherichia coli</i>	122
	The F factor	123
	F ⁺ × F ⁻ crosses	124
	Hfr × F ⁻ crosses	126
	Mapping by gradient of transmission	127
	Mapping by interrupted mating	129

F-prime factors	132
Summary of Key Points	134
9 Genetic Systems in Bacteriophages	135
The Virulent Phages	135
Recombination in the <i>rII</i> region of T4	136
Complementation in the <i>rII</i> region	137
Conditional lethal mutants in T4	139
The Temperate Phages	143
Lysogeny (λ)	143
Prophage integration	146
Restricted transduction	148
High-frequency transduction	150
Other temperate phages	151
Summary of Key Points	152
10 Mutation	153
Changes in chromosome number—polyploids	153
Changes in chromosome structure—rearrangements	154
Gene or point mutation	155
Types of mutant available for study	157
Spontaneous Mutation	158
Mutation in higher organisms	158
Mutation in bacteria	160
Biochemical mutations in bacteria	162
Mutation in phage	163
Induced Mutation	165
Radiation and mutation	166
Ultraviolet radiation	168
Chemicals as mutagens	168
The Molecular Basis of Mutation	170
Base analogue induced mutation	170
Mutagens that alter the chemical structure of DNA	173
Compounds that bind to DNA	174
Mutation is not a completely random process	176
Summary of Key Points	176
11 Genes and Proteins	178
The structure of proteins	178
Biochemical Genetics	182
Inborn errors of metabolism	182
Biochemical mutants	183
The one gene-one enzyme hypothesis	185
The one gene-one polypeptide hypothesis	185
The heterocaryon test	186
Sickle cell anaemia	188
The structure of haemoglobin	188

Tryptophan synthetase	190
Intragenic complementation	192
Summary of Key Points	192
12 The Genetic Code	195
Theoretical aspects of the code	195
The genetic code is non-overlapping	196
The General Nature of the Genetic Code	197
The Biochemistry of the Genetic Code	202
<i>In vitro</i> protein synthesis	202
Use of random polynucleotides	203
Triplet-binding experiments	206
Use of regular copolymers	207
Confirmation of the Genetic Code	208
Amino-acid substitution data	209
Tryptophan synthetase A protein	210
Partial nucleotide sequence of a gene in T4	211
Nucleotide sequence of a gene	212
Chain Termination Triplets	213
The head protein mutants of phage T4	214
Summary of Key Points	215
13 Protein Synthesis: Transcription	217
RNA is involved in protein synthesis	217
Proteins are synthesised on the ribosomes	217
Messenger RNA	218
RNA can be synthesised on a DNA template	220
Only one strand of DNA is transcribed <i>in vivo</i>	222
Synthesis is in the 5' to 3' direction	222
The initiation and termination of mRNA synthesis	224
RNA can be processed after transcription	226
RNA can also be copied on RNA templates	227
Interlude—the central dogma reversed	227
Summary of Key Points	228
14 Protein Synthesis: Translation	230
The ribosomes	230
Polyribosomes	232
Transfer RNA	233
tRNA—structural homologies and functions	235
Amino-acid activation	237
Chain elongation	239
Chain initiation	241
Chain termination	244
The intergenic region	244
The wobble hypothesis	247

Polarity and Translational Control	247
Polarity of the operon	247
Polar mutations	248
Translational control in the small RNA phages	249
Summary of Key Points	251
15 The Regulation of Gene Activity in Prokaryotes	252
Feedback inhibition	252
Enzyme induction and repression	255
The operon model	257
The Lactose Operon of <i>E. coli</i>	259
Structure and function	259
Mutations in the <i>lacI</i> regulator gene	261
Mutations in the operator region	264
The promoter	265
The terminator	265
The <i>lac</i> repressor	266
Other Regulatory Systems	267
The tryptophan operon	267
The arabinose operon in <i>E. coli</i>	268
Catabolite repression	270
Summary of Key Points	271
16 Recombination	273
The chiasmatype theory	273
Breakage and reunion models	274
Copy choice models	276
Recombination as a regular process	277
Aberrant segregations	277
Gene conversion in <i>Drosophila</i>	282
The Molecular Basis of Genetic Recombination	283
The polaron hybrid DNA model for genetic recombination	284
Error correction	286
General and site-specific recombination	291
The Repair of DNA	292
Damage by uv-irradiation	292
Photoreactivation	292
Excision repair	293
Recombination repair	295
DNA repair and mutation	296
Summary of Key Points	296
17 Mutation and Gene Expression	298
Gene Suppression	298
Intragenic suppression	298
Intergenic suppression	299
Alterations in ribosomal structure cause misreading of the code	302

Mutation and Cancer	303
Viruses and cancer	305
Position Effect Phenomena	306
Position effects in <i>Drosophila</i>	306
Controlling elements in maize	308
The <i>Ds-Ac</i> system	309
The <i>Spm</i> system	311
Presetting gene activity	313
X Chromosome Inactivation in Mammals	314
Dosage compensation	315
Summary of Key Points	316
18 Genetic Organisation and Function in Eucaryotes	318
Differential gene activity	318
Differentiation is reversible	321
Programming gene action—transdetermination	323
The Control of Differential Gene Activity	324
Differential gene replication	325
Transcriptional control	325
Translational control	327
The Functional Organisation of Eucaryotic Chromosomes	328
Lampbrush chromosomes	328
The ribosomal RNA genes of <i>Xenopus</i>	330
Pre-mRNA	333
Repetitive DNA	334
Chromomeres and genes	335
Chromosomal RNA	337
Summary of Key Points	337
<i>Bibliography</i>	339
<i>Glossary</i>	345
<i>Author Index</i>	353
<i>Organism Index</i>	356
<i>Subject Index</i>	360

1 This is Genetics

Every schoolboy knows it.
Bishop Jeremy Taylor (1613-1667)

To most students of a few years ago, genetics would have conjured up the idea of ratios in peas or in the fruit fly, and even today a dictionary may define genetics as simply 'the science of the study of heredity'. Although we realise the importance of understanding how characters are inherited, the ramifications of modern genetics are very much greater and spread into all fields of conventional biology. Today, geneticists already know the detailed structure of at least one gene and how it can code for the production of a specific polypeptide molecule, and they are now seeking a finer knowledge of how the action of the genes themselves is controlled; with this knowledge we will be nearer to understanding how the multitude of genes in a higher organism coordinate their actions so as to be able to control the development of a fertilised egg into a complex multicellular adult organism. Genetics is probing at the very nature of life itself and, indeed, 'life' has already been synthesised in the test tube, albeit it by copying the comparatively simple form of an infective viral chromosome. Even more important socially is the possibility that in the not too distant future we may be able to replace defective genes and so alleviate the miseries caused by the many inherited and incurable diseases of man.

Although many who read this book will already have studied some genetics, others will not have done so, and it is to them that the next two sections are addressed; they introduce some of the many technical terms used in the text and outline some of the basic concepts and foundations of genetics. The remaining chapters assume that the reader has understood these terms and concepts.

A Mendelian View of Genetics

The fundamental unit of any higher organism is the cell and it is convenient first to examine the basic structure of a generalised animal cell (figure 1.1). Each cell is surrounded by a *cell membrane* about 7.5 nm thick, and like all the cellular membranes it is made up of a layer of phospholipid molecules sandwiched between two layers of protein molecules. This membrane is semi-permeable and so allows the passage of some macromolecules, but not others, and it is a barrier between the exterior and the interior of the cell. The cell membrane encloses the *cytoplasm* and within this lie a number of cell organelles, such as the mitochondria, lysosomes, centrioles, endoplasmic reticulum and the Golgi apparatus. The *mitochondria* consist of two layers of membrane with extensive internal invaginations; they are rich in enzymes and their function is to provide energy by the oxidation of food substances. *Lysosomes* are also membrane bound and they contain the enzymes concerned with the breakdown of

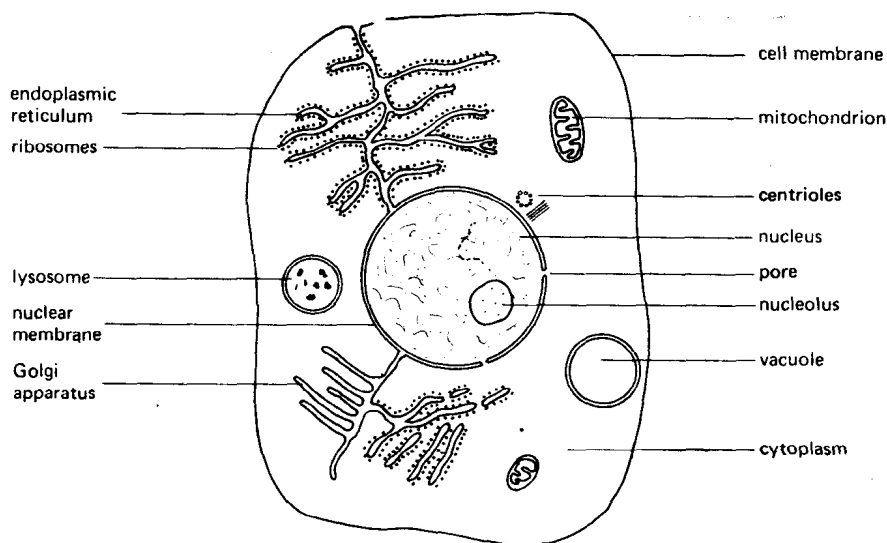


Figure 1.1 The generalised animal cell.

macromolecules. Every animal cell (but not plant cells) contains a pair of *centrioles* and they are responsible for organising the spindle at cell division (chapter 4). The *endoplasmic reticulum* is a system of membranes organised in pairs and it forms an intricate network within the cell. Some of the endoplasmic reticulum, the *rough* endoplasmic reticulum, has one surface lined with *ribosomes*. The ribosomes, made up of RNA and protein, are found either lining the endoplasmic reticulum or free within the cell and they are the factories where the proteins are synthesised; some of these proteins are secreted into the endoplasmic reticulum. Another membrane system, probably continuous with the endoplasmic reticulum, is the *Golgi apparatus*; this has a layered structure and there are no attached ribosomes.

The most important part of the cell is the *nucleus*, the control centre of the cell. The nucleus is also bounded by a double membrane, continuous with the endoplasmic reticulum, and it contains the genetic material (collectively referred to as *chromatin*) and one or more *nucleoli* where the ribosomal RNA is synthesised. The genetic information is stored in discrete bodies found in the cell nucleus, the *chromosomes*. Each chromosome is differentiated along its length into a very large number, perhaps 500 to 2000, of basic genetic units of *genes*. With certain exceptions each gene functions by specifying the biosynthesis of a particular polypeptide, often in the form of an enzyme, or by carrying out a control role in biosynthesis.

In higher organisms each cell usually contains *two* complete sets of chromosomes—in other words the chromosomes occur in pairs of two similar, but not necessarily identical, homologues, one derived from each parent. Each is

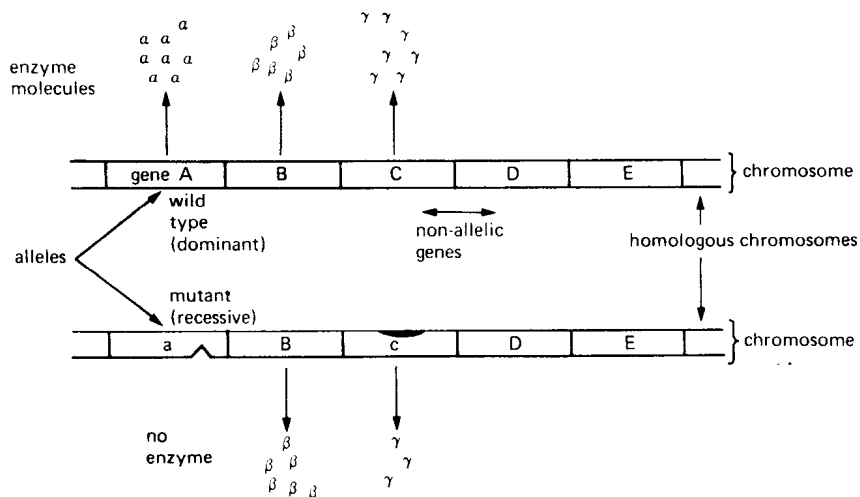


Figure 1.2 The organisation of the genetic material in a diploid eucaryotic cell.

a pair of *homologous* chromosomes. Along each chromosome the sequence of genes is identical although the genes themselves may be slightly different so that they produce qualitatively or quantitatively different products. In figure 1.2 one homologue has a defective gene *a* which has been structurally altered so that it can no longer specify the production of the α protein molecules, while gene *c* has been modified so that it produces fewer molecules of protein γ ; both of the corresponding wild type genes, *A* and *C*, produce normal gene products. Thus *A* and *a* (or *C* and *c*) are different forms of the same gene, or *alleles* of each other, and in this example *A* is the wild type allele and *a* the mutant allele. *Mutation* is the general name given to the processes which can change one allele into another allele (chapter 10).

Since the chromosomes occur in pairs the genes must also occur in pairs, so that for the pair of alleles *A* and *a* there are three possible genic combinations or *genotypes*, *AA*, *Aa* and *aa*. When both homologues carry the same allele (*AA* or *aa*) the cell is said to be *homozygous* (and the organism is a *homozygote*), while if the alleles are different (*Aa*) it is *heterozygous*. In a heterozygous *Aa* organism the *A* allele will enable the production of the wild type protein and so the individuals will have a normal appearance or *phenotype* and be indistinguishable from an *AA* homozygote; the only individuals who will have the abnormal or mutant phenotype will be the *aa* homozygotes. We say that the *A* allele is *dominant* over *a*, or that *a* is *recessive* to *A*. Usually, but not always, the wild type allele is dominant and the mutant allele is recessive; dominant and recessive alleles are frequently denoted by capital and lower case letters respectively. Not all pairs of alleles show dominance and recessiveness and the heterozygotes may have a phenotype intermediate between the phenotypes of the two homozygotes.

It is important to realise that because an organism has a particular gene does not necessarily mean that the corresponding phenotype will be manifest. A seedling is normally green but, because light is necessary for the formation of chlorophyll, it is pale yellow if grown in the dark; and yet these yellow seedlings have all the genes necessary to make chlorophyll and so will rapidly go green when transferred to the light. For the vast majority of inherited characters the phenotype is the result of interaction between the genotype *and* the environment.

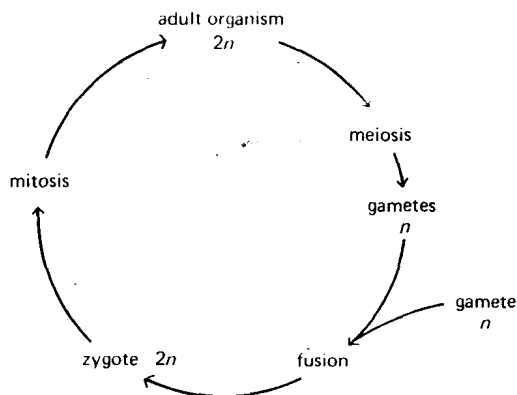


Figure 1.3 A generalised life cycle of a higher animal: n is the number of chromosomes in a haploid set; $2n$ represents a diploid cell or organism.

A cell containing two complete sets of chromosomes is called a *diploid* and a cell with only one set is *haploid*. The terms diploid and haploid are also used to refer to organisms containing predominantly diploid or haploid cells. The term *genome* is used to refer to the total gene content of a cell or organism.

A cell divides by the process of *mitosis* (chapter 4), in such a way that each daughter cell receives a complete and exact copy of all the chromosomes in the mother cell; it is by mitosis that a zygote divides so as to produce, eventually, a multi-cellular adult organism. However, in the life cycle of a sexually reproducing organism (figure 1.3) the diploid zygote is the product of the fusion of a male and a female gamete, and if the number of chromosomes is to remain constant and not to double at each successive generation there must, somewhere in the life cycle, be a special type of cell division which halves the number of chromosomes. This is *meiosis* (chapter 4) and it occurs during gamete formation so that each gamete is haploid and contains only one set of chromosomes; when two gametes fuse to form a fertilised egg cell or *zygote* the diploid number is restored.

A Molecular View of Genetics

The ultimate aim of molecular genetics is to explain all of genetics, and indeed all of biology, in physical and chemical terms, the basis of the explanations being

the standard chemical bonds; it attempts to elucidate the structure and function of the genetic material in molecular terms by using a variety of sophisticated physical, chemical, biochemical and genetical techniques. Molecular genetics is therefore a truly interdisciplinary science, not only ramifying into but extending beyond all the fields of conventional biology, and in the last decade no other science has made such significant advances.

Any cell can be likened to a factory; there is an inward flow of raw materials which, inside the cell, are manufactured into the many different types of organic molecules that are required to enable the cell to grow and to function. The instructions which enable these many intra-cellular reactions to take place, the genetic information, are encoded in the chromosomes, and we are interested in knowing how this information is stored, replicated and processed.

The secret lies in the chemical substance known as deoxyribonucleic acid, or DNA, for this is the genetic material and the most important constituent of the chromosomes (chapter 2). DNA is a very long polymer known as a *polynucleotide*, made up of four types of building block units or *nucleotides*. A molecule of DNA is many thousands of units long and it is the particular sequence of the nucleotide blocks along the molecule that encodes the genetic information. So important is the exact sequence of the blocks in a gene—which may be up to 1500 blocks long—that the replacement of one block by a different block (mutation) may completely destroy the normal activity of the gene. The structure of the four building blocks, the deoxyribonucleotides of adenine (A), thymine (T), guanine (G) and cytosine (C), is such that they occur in pairs, A with T and C with G, so that in fact the molecule of DNA consists of two chains of nucleotides paired off along their length (figure 1.4), and, because of the precise pairing relationships the sequence of nucleotides in one chain is determined by the sequence in the other chain; in other words the two chains are *complementary* to each other. The replication of such a molecule is easy; the two chains can separate and each separate chain can then act as a *template* for the formation of a new complementary chain. Because of the exact pairing each daughter molecule will have exactly the same nucleotide sequence as the parental molecule, so preserving unaltered the genetic information.

In some viruses the genetic material is not DNA but a closely related nucleic acid called ribonucleic acid or RNA. RNA is made up from the ribonucleotides of adenine, guanine and cytosine, uracil (U, which replaces and behaves as thymine in DNA) and it is usually single stranded rather than double stranded.

We must next ask, what is the function of a gene? Except for some genes which produce ribonucleic acid as their end product and others which are involved in control processes, the vast majority direct the synthesis of protein molecules; these proteins may be structural, contributing to the fabric of the cell, or, more usually, *enzymes* which enable the intricate biochemical processes of cell metabolism to take place (chapter 11).

Protein molecules are also long chain polymers, made up of twenty types of repeating units called *amino acids*, and are commonly 200–400 units long. The primary product of gene action is one of these long chains of amino acids, a polypeptide chain, which subsequently folds up according to a highly specific