Molecular Genetics of Bacteria

5th edition

Jeremy W. Dale and **Simon F. Park**
University of Surrey, UK

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

Molecular Genetics of Bacteria

5th edition

Molecular Genetics of Bacteria

5th edition

Jeremy W. Dale and **Simon F. Park**
University of Surrey, UK

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2010,

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Other Editorial Offices:
9600 Garsington Road, Oxford, OX4 2DQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell

Library of Congress Cataloguing-in-Publication Data

Record on file

ISBN 9780470741856 (HB) 9780470741849 (PB)

A catalogue record for this book is available from the British Library.

Set in 10.5/12.5 pt TimesTen by Laserwords Private Limited, Chennai, India
Printed in Great Britain by Ashford Colour Press Ltd, Fareham, Hants

First impression 2010

Contents

Preface xi

1 Nucleic Acid Structure and Function 1

1.1 Structure of nucleic acids 1
1.1.1 DNA 1
1.1.2 RNA 2
1.1.3 Hydrophobic interactions 3
1.1.4 Different forms of the double helix 4
1.1.5 Supercoiling 5
1.1.6 Denaturation and hybridization 8
1.1.7 Orientation of nucleic acid strands 10
1.2 Replication of DNA 11
1.2.1 Unwinding and rewinding 12
1.2.2 Fidelity of replication; proofreading 12
1.3 Chromosome replication and cell division 14
1.4 DNA repair 16
1.4.1 Mismatch repair 16
1.4.2 Excision repair 18
1.4.3 Recombination (post-replication) repair 19
1.4.4 SOS repair 21
1.5 Gene expression 21
1.5.1 Transcription 21
1.5.2 Translation 25
1.5.3 Post-translational events 31
1.6 Gene organization 35

2 Mutation and Variation 37

2.1 Variation and evolution 37
2.1.1 Fluctuation test 39
2.1.2 Replica plating 41
2.1.3 Directed mutation in bacteria? 41
2.2 Types of mutation 43
2.2.1 Point mutations 43
2.2.2 Conditional mutants 44
2.2.3 Variation due to larger-scale DNA alterations 45
2.2.4 Extrachromosomal agents and horizontal gene transfer 46

2.3 Recombination 47
2.3.1 A model of the general (homologous) recombination process 48
2.3.2 Enzymes involved in recombination 49
2.4 Phenotypes 50
2.4.1 Restoration of phenotype 52
2.5 Mechanisms of mutation 55
2.5.1 Spontaneous mutation 55
2.5.2 Chemical mutagens 57
2.5.3 Ultraviolet irradiation 60
2.6 Isolation and identification of mutants 63
2.6.1 Mutation and selection 63
2.6.2 Replica plating 64
2.6.3 Isolation of other mutants 65
2.6.4 Molecular methods 66

3 Regulation of Gene Expression 71
3.1 Gene copy number 73
3.2 Transcriptional control 73
3.2.1 Promoters 73
3.2.2 Terminators, attenuators and anti-terminators 84
3.2.3 Induction and repression: regulatory proteins 84
3.2.4 Two-component regulatory systems 93
3.2.5 Global regulatory systems 97
3.2.6 Quorum sensing 98
3.3 Translational control 102
3.3.1 Ribosome binding 102
3.3.2 Codon usage 104
3.3.3 Stringent response 104
3.3.4 Regulatory RNA 105
3.4 Phase variation 114

4 Genetics of Bacteriophages 115
4.1 Bacteriophage structure 116
4.2 Single-strand DNA bacteriophages 118
4.2.1 φX174 118
4.2.2 M13 121
4.3 RNA-containing phages: MS2 121
4.4 Double-stranded DNA phages 121
4.4.1 Bacteriophage T4 122
4.4.2 Bacteriophage λ 124
4.4.3 Lytic and lysogenic regulation of bacteriophage λ 128
4.5 Restriction and modification 136
4.6 Bacterial resistance to phage attack 138
4.7 Complementation and recombination 138
4.8 Why are bacteriophages important? 140

4.8.1 Phage typing 141
4.8.2 Phage therapy 141
4.8.3 Phage display 142
4.8.4 Phages in the natural environment 142
4.8.5 Bacterial virulence and phage conversion 144

5 Plasmids 147
5.1 Some bacterial characteristics are determined by plasmids 148
5.1.1 Antibiotic resistance 148
5.1.2 Colicins and bacteriocins 148
5.1.3 Virulence determinants 149
5.1.4 Plasmids in plant-associated bacteria 149
5.1.5 Metabolic activities 150
5.2 Molecular properties of plasmids 152
5.2.1 Plasmid replication and control 154
5.2.2 Partitioning 164
5.2.3 Host range 164
5.2.4 Plasmid incompatibility 167
5.3 Plasmid stability 168
5.3.1 Plasmid integrity 168
5.3.2 Partitioning 169
5.3.3 Differential growth rate 174
5.4 Associating a plasmid with a phenotype 174

6 Gene Transfer 177
6.1 Transformation 178
6.2 Conjugation 180
6.2.1 Mechanism of conjugation 180
6.2.2 The F plasmid 185
6.2.3 Conjugation in other bacteria 186
6.3 Transduction 190
6.3.1 Specialized transduction 192
6.4 Recombination 192
6.4.1 Consequences of recombination 193
6.4.2 Site-specific and non-homologous (illegitimate) recombination 194
6.5 Mosaic genes and chromosome plasticity 195

7 Genomic Plasticity: Movable Genes and Phase Variation 199
7.1 Insertion sequences 199
7.1.1 Structure of insertion sequences 199
7.1.2 Occurrence of insertion sequences 200
7.2 Transposons 202
7.2.1 Structure of transposons 204
7.2.2 Integrons 206
7.2.3 IS*CR* elements 208

7.3 Mechanisms of transposition 209
7.3.1 Replicative transposition 209
7.3.2 Non-replicative (conservative) transposition 212
7.3.3 Regulation of transposition 213
7.3.4 Activation of genes by transposable elements 214
7.3.5 Mu: A transposable bacteriophage 214
7.3.6 Conjugative transposons 215
7.4 Phase variation 215
7.4.1 Variation mediated by simple DNA inversion 217
7.4.2 Variation mediated by nested DNA inversion 218
7.4.3 Antigenic variation in the gonococcus 218
7.4.4 Phase variation by slipped-strand mispairing 222
7.4.5 Phase variation mediated by differential DNA methylation 223
7.5 Clustered regularly interspersed short palindromic repeats 224

8 Genetic Modification: Exploiting the Potential of Bacteria 227
8.1 Strain development 228
8.1.1 Generation of variation 228
8.1.2 Selection of desired variants 228
8.2 Overproduction of primary metabolites 229
8.2.1 Simple pathways 229
8.2.2 Branched pathways 230
8.3 Overproduction of secondary metabolites 232
8.4 Gene cloning 233
8.4.1 Cutting and joining DNA 234
8.4.2 Plasmid vectors 234
8.4.3 Bacteriophage λ vectors 238
8.4.4 Cloning larger fragments 240
8.4.5 Bacteriophage M13 vectors 242
8.5 Gene libraries 242
8.5.1 Construction of genomic libraries 242
8.5.2 Screening a gene library 244
8.5.3 Cloning PCR products 246
8.5.4 Construction of a cDNA library 247
8.6 Products from cloned genes 248
8.6.1 Expression vectors 248
8.6.2 Making new genes 250
8.6.3 Other bacterial hosts 253
8.6.4 Novel vaccines 255
8.7 Other uses of gene technology 256

9 Genetic Methods for Investigating Bacteria 257
9.1 Metabolic pathways 257
9.1.1 Complementation 257
9.1.2 Cross-feeding 258

9.2 Microbial physiology 259
9.2.1 Reporter genes 261
9.2.2 Chromatin immunoprecipitation 262
9.2.3 Cell division 264
9.2.4 Motility and chemotaxis 266
9.2.5 Cell differentiation 267
9.3 Bacterial virulence 271
9.3.1 Wide-range mechanisms of bacterial pathogenesis 271
9.3.2 Detection of virulence genes 272
9.4 Specific mutagenesis 280
9.4.1 Gene replacement 280
9.4.2 Antisense RNA 282
9.5 Taxonomy, evolution and epidemiology 282
9.5.1 Molecular taxonomy 282
9.5.2 GC content 283
9.5.3 16 S rRNA 283
9.5.4 Denaturing-gradient gel electrophoresis and temperature-gradient gel electrophoresis 286
9.5.5 Diagnostic use of PCR 286
9.5.6 Molecular epidemiology 287

10 Gene Mapping to Genomics and Beyond 295
10.1 Gene mapping 295
10.1.1 Conjugational analysis 295
10.1.2 Gene libraries 298
10.1.3 Restriction mapping and pulsed-field gel electrophoresis 298
10.2 DNA sequence determination 301
10.2.1 Sanger sequencing 301
10.2.2 Dye terminator sequencing 304
10.2.3 Pyrosequencing 304
10.2.4 Massively parallel sequencing 305
10.3 Genome sequencing 307
10.3.1 Genome-sequencing strategies 308
10.3.2 Relating sequence to function 309
10.3.3 Metagenomics 314
10.4 Comparative genomics 315
10.4.1 Microarrays 318
10.5 Analysis of gene expression 320
10.5.1 Transcriptional analysis 320
10.5.2 Translational analysis 325
10.6 Metabolomics 329
10.7 Systems biology and synthetic genomics 329
10.7.1 Systems biology 329
10.7.2 Synthetic genomics 329
10.8 Conclusion 330

Appendix A Further Reading 333
Appendix B Abbreviations Used 337
Appendix C Glossary 341
Appendix D Enzymes and other Proteins 365
Appendix E Genes 369
Appendix F Standard Genetic Code 373
Appendix G Bacterial Species 375

Index 379

Preface

In the preface to the fourth edition (published in 2004) we referred to the revolution in bacterial genetics that was started by gene cloning and sequencing, coupled with related techniques such as the polymerase chain reaction (PCR) and microarrays, and culminated in knowledge of genome sequences of a rapidly expanding range of bacteria. This posed a dilemma. How could we accommodate these new techniques, and the wealth of exciting new information they provided, while not losing sight of classical bacterial genetics? This dilemma has become even more acute.

True, many of the older methods are now no longer used, and could be relegated to the pages of history. But there is a danger of throwing out the baby with the bathwater. Not only is there a need to maintain some sense of how the subject has got to the stage we are now at, but also a discussion of some of these methods is useful in establishing an understanding of how bacterial genetics operates in natural environments. Genetics is not just about how we find out about bacteria: it is about how bacteria have evolved, and continue to evolve, and continue to adapt to changing environments. Molecular genetics, in isolation, is essentially reductionist. Even genome sequencing, and global analysis of gene expression, by themselves merely provide catalogues of genes. Ultimately, those lists have to be related to the behaviour of the whole organism, and from there to how organisms interact with one another and with their environment.

So we have continued with a compromise approach, slimming down even further the description of classical bacterial genetics to allow space not only for some of the new technological advances but also for some of the advances that these methods have allowed in understanding important aspects of bacterial behaviour.

One further aspect needs a word of explanation. What is a bacterium? It is now clear that there are two distinct types of prokaryote: the bacteria proper and the Archea. Most of this book concerns the bacteria in the proper taxonomic sense, but some aspects are also relevant to Archea, especially where we consider the differences between prokaryotes and eukaryotes. It complicates the text too much to make this distinction, so we apologize if anyone is offended by occasional blurring of the lines between bacteria and prokaryotes.

As with the earlier editions, the choice of what to include and what to leave out is very much a personal one. We hope that the final product will remain accessible for a non-specialist reader, and will succeed in introducing them to both the exciting and rapidly developing field of molecular genetics and also the fascinating world of bacteria.

Jeremy W. Dale
Simon F. Park

1 Nucleic Acid Structure and Function

In this book it is assumed that you will already have a working knowledge of the essentials of molecular biology, especially the structure and synthesis of nucleic acids and proteins. The purpose of this chapter therefore is to serve as a reminder of some of the most relevant points, and to highlight those features that are particularly essential for an understanding of later chapters.

1.1 Structure of nucleic acids

1.1.1 DNA

In bacteria, the genetic material is double-stranded DNA, although bacteriophages (viruses that infect bacteria; see Chapter 4) may have double-stranded or single-stranded DNA, or RNA. The components of DNA (Figure 1.1) are 2′-deoxyribose (forming a backbone in which they are linked by phosphate residues) and four heterocyclic bases: two purines (adenine, A, and guanine, G) and two pyrimidines (thymine, T, and cytosine, C). The sugar residues are linked by phosphodiester bonds between the 5′ position of one deoxyribose and the 3′ position of the next (Figure 1.2), while one of the four bases is attached to the 1′ position of each deoxyribose. It is the sequence of these four bases that carries the genetic information.

The two strands are twisted around each other in the now familiar double helix, with the bases in the centre and the sugar-phosphate backbone on the outside. The two strands are linked by hydrogen bonds between the bases. The only arrangement of these bases that is consistent with maintaining the helix in its correct conformation is when adenine is paired with thymine and guanine with cytosine. One strand therefore consists of an image of the other; the two strands are said to be *complementary*. Note that the purines are larger than the pyrimidines, and that this arrangement involves one purine opposite a pyrimidine at each position, so the distance separating the strands remains constant.

Molecular Genetics of Bacteria, Fifth Edition, Jeremy W. Dale and Simon F. Park

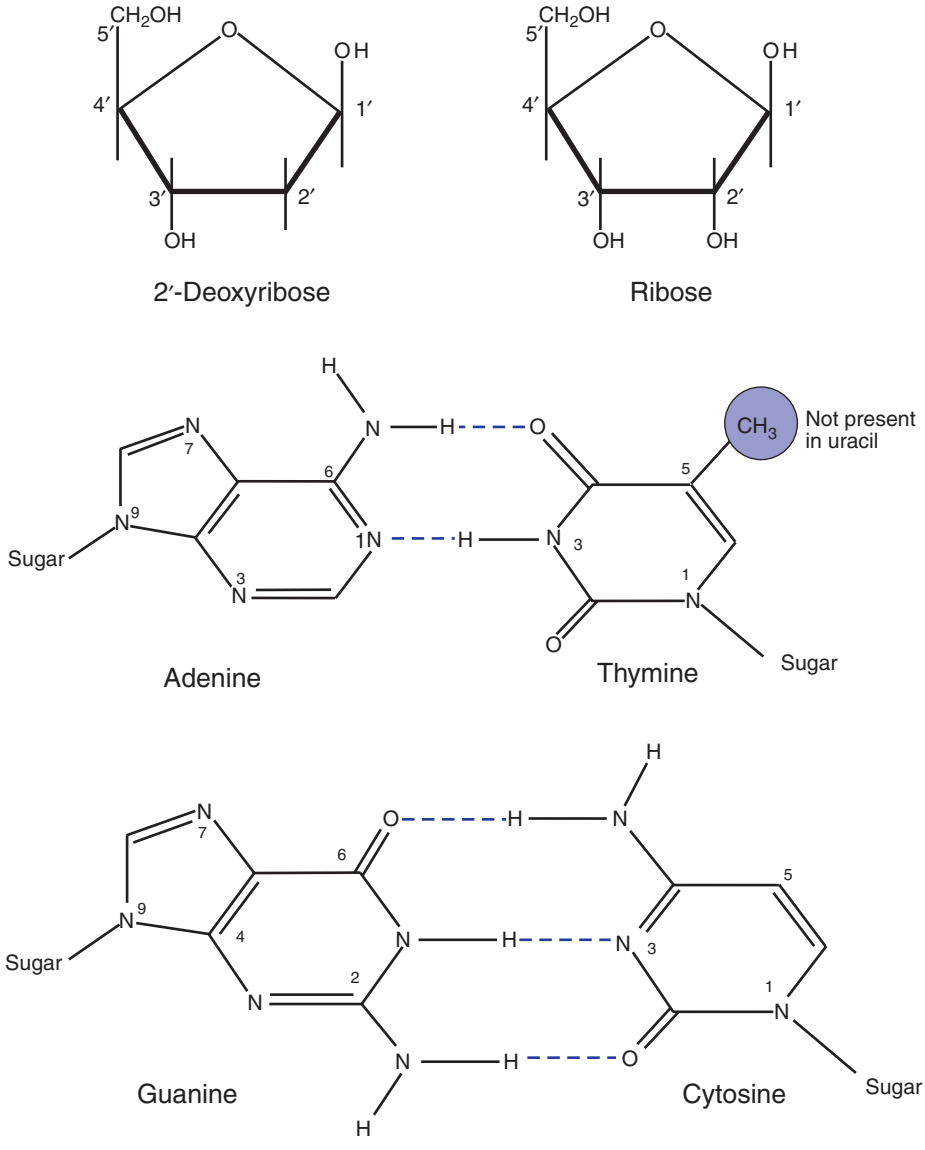

Figure 1.1 Structure of the basic elements of DNA and RNA. RNA contains ribose rather than deoxyribose, and uracil instead of thymine.

1.1.2 RNA

The structure of RNA differs from that of DNA in that it contains the sugar ribose instead of deoxyribose, and uracil instead of thymine (Figure 1.1). It is usually described as single-stranded, but only because the complementary

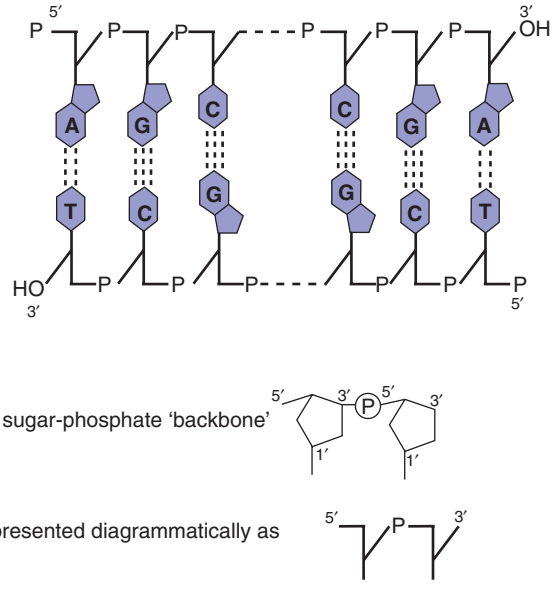

Figure 1.2 Diagrammatic structure of DNA.

strand is not normally made. There is nothing inherent in the structure of RNA that prevents it forming a double-stranded structure: an RNA strand will pair with (hybridize to) a complementary RNA strand, or with a complementary strand of DNA. Even a single strand of RNA will fold back on itself to form double-stranded regions. In particular, transfer RNA (tRNA), and ribosomal RNA (rRNA) both form complex patterns of base-paired regions. The formation of secondary and tertiary structures in RNA via base-pairing can also influence gene expression and this is considered in further detail in Chapter 3.

1.1.3 Hydrophobic interactions

Although geneticists emphasize the importance of the hydrogen bonding between the two DNA strands, these are not the only forces influencing the structure of the DNA. The bases themselves are hydrophobic, and will tend to form structures in which they are removed from the aqueous environment. This is partially achieved by stacking the bases on top of one another (Figure 1.3). The double-stranded structure is stabilized by additional hydrophobic interactions between the bases on the two strands. The hydrogen bonding not only holds the two strands together but also allows the corresponding bases to approach sufficiently closely for the hydrophobic forces to operate. The hydrogen bonding of the bases is, however, of special

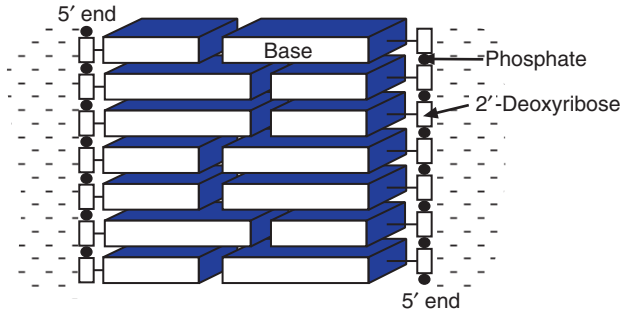

Figure 1.3 Hydrophobic interactions of bases in DNA. The hydrophobic bases stack in the centre of the helix, reducing their contact with water.

importance because it gives rise to the specificity of the base-pairing between the two chains.

Although the bases are hydrophobic, and therefore very poorly soluble in water, nucleic acids are quite soluble, due largely to the hydrophilic nature of the backbone, and especially the high concentration of negatively charged phosphate groups. This will also tend to favour a double-helical structure, in which the hydrophobic bases are in the centre, shielded from the water, and the hydrophilic phosphate groups are exposed.

1.1.4 Different forms of the double helix

A full consideration of DNA structure would be extremely complex, and would have to take into account interactions with the surrounding water itself, as well as the influence of other solutes or solvents. The structure of DNA can therefore vary to some extent according to the conditions. *In vitro*, two main forms are found. The Watson and Crick structure refers to the B form, which is a right-handed helix with 10 base-pairs per turn (Figure 1.4). Under certain conditions, isolated DNA can adopt an alternative form known as the A form, which is also a right-handed helix, but more compact, with about 11 base pairs (bp) per turn. Within the cell, DNA resembles the B form more closely, but has about 10.4 bp per turn (it is *underwound*; see below).

Certain DNA sequences, notably those containing alternating G and C residues, tend to form a left-handed helix, known as the Z form (since the sugar-phosphate backbone has a zigzag structure rather than the regular curve shown in the B form). Although Z DNA was originally demonstrated using synthetic oligonucleotides, naturally occurring DNA within the cell can adopt a left-handed structure, at least over a short distance or temporarily. The switch from left- to right-handed can have important influences on the expression of genes in that region.

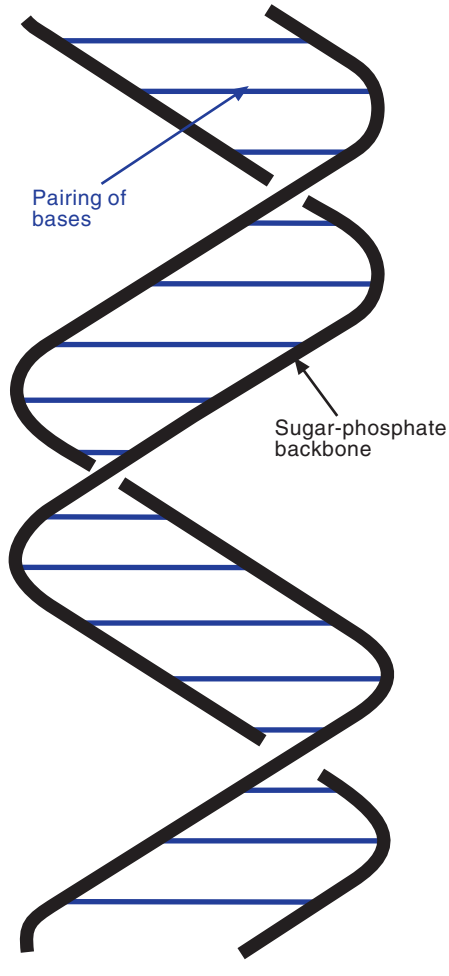

Figure 1.4 Diagrammatic structure of B-form DNA. The two anti-parallel sugar-phosphate chains form a right-handed helix, with the bases in the centre, held together by hydrophobic interactions and hydrogen bonding.

1.1.5 Supercoiling

Within the cell, the DNA helix is wound up into coils; this is known as supercoiling. Figure 1.5 shows a simple demonstration of supercoiling, which you can easily try out for yourself. Take a strip of paper, and twist one end to introduce one complete turn (i.e. the same side of the paper is facing you at each end). It will now look as in Figure 1.5a. Then bring the two ends towards each other; the conformation will change to that shown in Figure 1.5b, which is a simple form of supercoiling. Notice that not only has the strip of paper become supercoiled, but also the degree of twisting seems to have changed

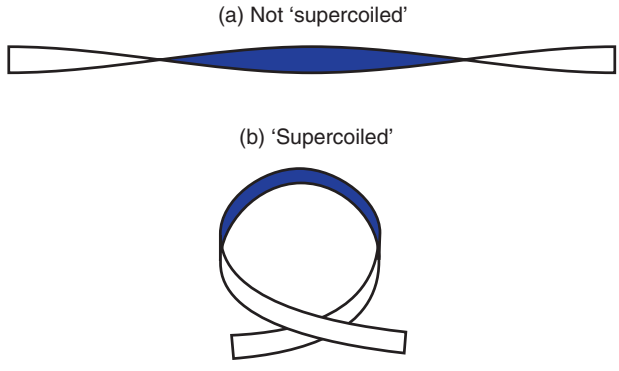

Figure 1.5 Interaction between twisting and supercoiling. (a) A ribbon with a single complete twist, without supercoiling. (b) The same ribbon, allowed to form a supercoil; the ribbon is now not twisted.

(in this example it now appears not to be twisted at all). If you have kept hold of both ends, the twist of the strip cannot have disappeared completely; it has merely changed to a different form. If you pull the ends apart again, it will change back to the form shown in Figure 1.5a.

There are three parameters involved: twist (T), linking number (L) and writhe (W). The twist is the number of turns of the strip whereas writhe (essentially a measure of the degree of supercoiling) can be considered as the number of times the strip crosses over itself in a defined direction. These two parameters vary according to the conformation: in Figure 1.5a there is one twist ($T = 1$) but no supercoiling ($W = 0$), whereas in Figure 1.5b there is no twist ($T = 0$) and the strip crosses itself once ($W = 1$). The linking number, which is a measure of the overall twisting of the strip, is equal to the sum of the other two parameters, i.e. $L = T + W$.

If the ends of the strip are not free to rotate, then the linking number will remain constant. Most of the DNA molecules we will be considering are circular, and therefore do not contain ends that can rotate. Unless there is a break in the DNA, any change in the twist will be balanced by a change in supercoiling, and vice versa. This is illustrated by Figure 1.6.

In Figure 1.6a the strip (or DNA molecule) is not supercoiled ($W = 0$) but contains one complete twist ($T = +1$); the linking number (L) is +1. In Figure 1.6b the overall shape has been changed by rotating one end of the structure (i.e. introducing a degree of supercoiling). The strip crosses itself once, and by convention a crossover in this direction is assigned a negative value, so $W = -1$. At the same time the twist has changed; there are now two complete twists, so $T = +2$. Since $L = T + W$, we see that L remains the same (+1).

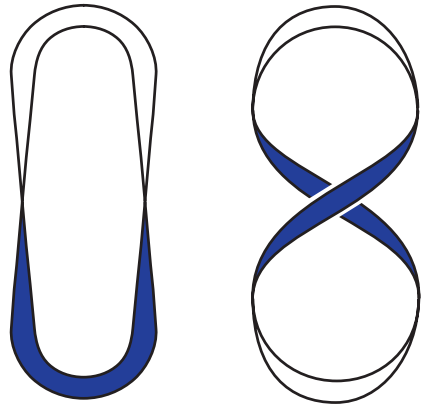

Figure 1.6 Supercoiling of a circular molecule. In (a), the 'molecule' has one twist and no supercoils. Rotating one end of the molecule (b) introduces negative supercoils and increases the amount of twisting. Since the linking number (L) remains the same, the two forms are interchangeable without breaking the circle. See the text for further explanation.

The two structures shown in Figure 1.6 are interchangeable by rotating one end, without opening the circle. With an intact circle, you can change the twist and the writhe jointly but not separately. Any change in supercoiling will involve a compensating change in the twist (and vice versa) so that the linking number remains constant. *It is only possible to alter the linking number in circular DNA by breaking and rejoining DNA strands*, for example through the action of topoisomerases (see below).

Bacterial DNA is normally negatively supercoiled. Another way of putting it is to say that the DNA is underwound so that if the DNA was not supercoiled, the degree of twisting of the helix would be less than that seen in relaxed linear DNA. If the DNA is nicked (i.e. one strand is broken, leaving it free to rotate) it relaxes into an open circular, non-supercoiled form. Chromosomal DNA is usually broken into linear fragments during lysis of the cell, but bacterial plasmids (see Chapter 5) are usually small enough to be isolated intact in a supercoiled form.

The compact supercoiled structure of the DNA is also significant in that the chromosome, in its expanded state, would be a thousand times longer (about 1 mm) than the bacterial cell itself. To put it another way, a bacterial operon of four genes, in its non-supercoiled B form, would stretch from one end of the cell to the other. Supercoiling is only the start of story, as the bacterial chromosome consists of a large number of supercoiled loops arranged on

a core to produce a highly compact and organized structure known as the *nucleoid*. Supercoiling (and other structural features) of the DNA are also important in the regulation of gene expression (see Chapter 3).

Action of topoisomerases

Supercoiling of bacterial DNA is not achieved by physically twisting the circular molecule in the way we illustrated in Figure 1.6. Instead the cell uses enzymes known as *DNA topoisomerases* to introduce (or remove) supercoils from DNA by controlled breaking and rejoining of DNA strands.

DNA topoisomerases can be considered in two classes. Type I topoisomerases act on a segment of DNA by breaking one of the strands and passing the other strand through the gap, followed by resealing the nick. Since this increases the number of times the two strands cross one another, the linking number is increased by 1, which results in an increase in either T or W. The *Escherichia coli* topoisomerase I acts only on negatively supercoiled DNA; the increase in the value of W means that the degree of negative supercoiling is reduced (the DNA becomes relaxed).

Type II topoisomerases break both strands and pass another duplex region through the gap. In Figure 1.7, we start with a non-supercoiled circle (structure A) and move the centres of the right- and left-hand loops across one another to give structure B. Although this may look at first glance to be supercoiled, it is not. The two crossovers are in opposite directions and therefore cancel one another out. (Mathematically $W = -1$ at the upper point, and $W = +1$ at the lower one, so overall $W = 0$.) If both strands of the helix are broken between points L and M, and the lower strands (X – Y) are moved through the gap, followed by resealing the strands between L and M, structure C is formed. Now, both crossovers are in the same direction, so the structure *is* supercoiled ($W = -2$). There is a corresponding reduction in the linking number. The enzyme has introduced a negative supercoil, and at the same time has reduced the winding of the helix. An important example of this type of enzyme is DNA gyrase, which is able to introduce negative supercoils into newly replicated DNA.

1.1.6 Denaturation and hybridization

Since the two strands of DNA are only linked by non-covalent forces, they can easily be separated in the laboratory, for example by increased temperature or high pH. Separation of the two DNA strands, *denaturation*, is readily reversible. Reducing the temperature, or the pH, will allow hydrogen bonds between complementary DNA sequences to reform; this is referred to as *re-annealing* (Figure 1.8). If DNA molecules from different sources are denatured, mixed and allowed to re-anneal, it is possible to form hydrogen bonds between similar DNA sequences (*hybridization*). This forms the basis of the use of DNA probes to detect specific

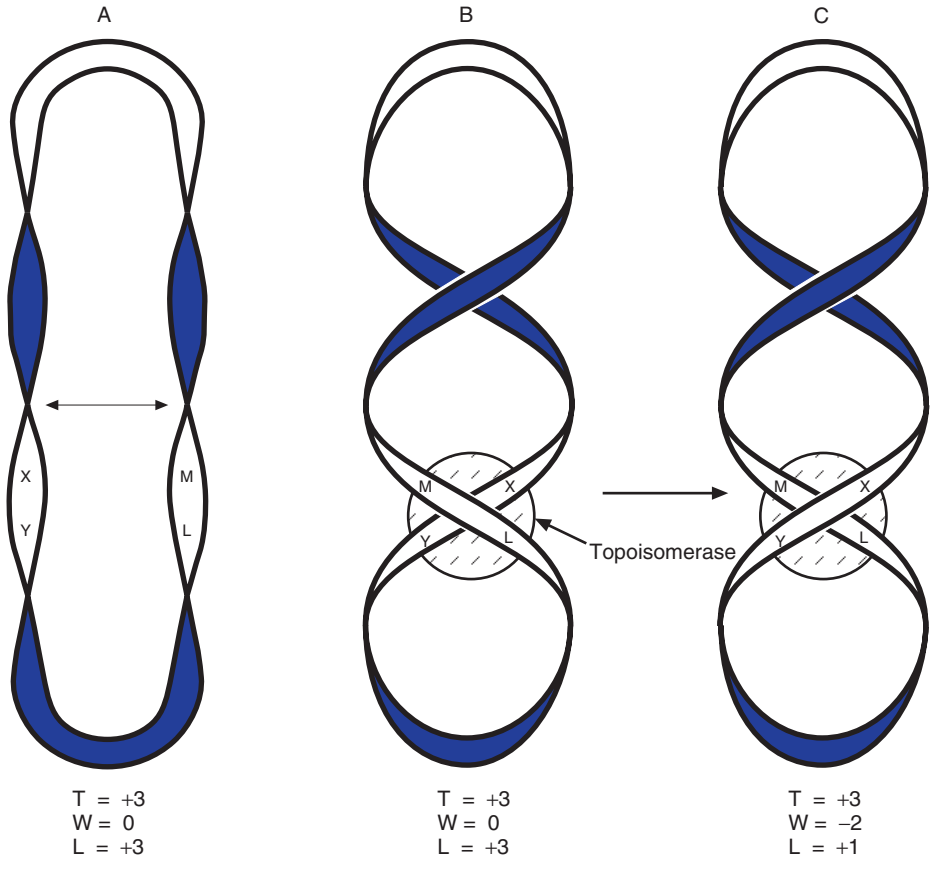

Figure 1.7 Action of Type II topoisomerase. Structure A is not supercoiled, and is converted to B by bending the two sides as shown by the arrow. B is not supercoiled: the two crossing points are of opposite sign and cancel one another. The topoisomerase makes a double-strand break between L and M, passes the X-Y region through the gap, and re-seals the break between L and M. This changes the sign of W at that point, so structure C is now negatively supercoiled.

DNA sequences. The specificity can be adjusted by altering the conditions used for re-annealing (or subsequent washing). Higher temperature, or lower ionic strength, gives greater *stringency* of hybridization. High-stringency hybridization is used to detect closely related sequences, or to distinguish between sequences with only small differences, whereas low-stringency conditions are used to detect sequences that are more remotely related to your probe. This technique forms an important part of modern molecular biology, and we will encounter many applications in subsequent chapters.

Temporary separation of localized regions of the two DNA strands also occurs as an essential part of the processes of replication and transcription.

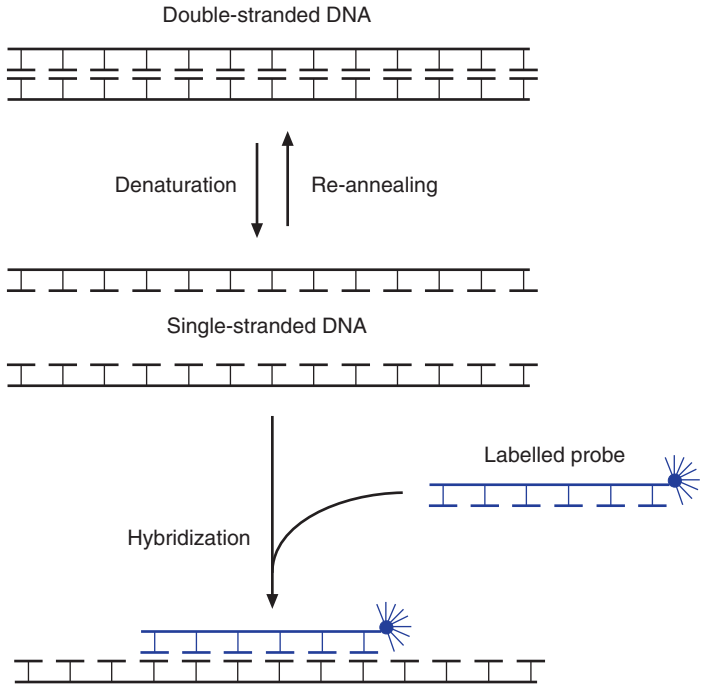

Figure 1.8 Denaturation and hybridization of DNA.

Note that there are three hydrogen bonds linking guanine and cytosine while the adenine–thymine pairing has only two hydrogen bonds. The two DNA strands are therefore more strongly attached in those regions with a high G + C content. Because of this, such regions are more resistant to denaturation and conversely re-anneal more readily. The influence of base composition on the ease of separation of two nucleic acid strands may play an important role in the control of processes such as the initiation of RNA synthesis where an A–T-rich region may facilitate the initial separation of the DNA strands (Chapter 3).

1.1.7 Orientation of nucleic acid strands

A further noteworthy feature of the helix is that each strand can be said to have a direction, based on the orientation of the linkages in the sugar-phosphate backbone. Each phosphate group joins the 5′ position of one sugar residue to the 3′ position of the next deoxyribose. In Figure 1.2, the upper strand has a free 5′ group at the left-hand end and a 3′ OH group at the right-hand end. It is therefore said to run (from left to right) in the 5′ to 3′ direction. Conversely, for the lower strand, the 5′ to 3′ direction runs from right to left.

By convention, if a single DNA (or RNA) strand is shown, it reads in the 5′ to 3′ direction from left to right (unless otherwise stated). If both strands are shown, the upper strand reads (left to right) from the 5′ to 3′ end.

All nucleic acids are synthesized in the 5′ to 3′ direction. That is, the new strand is elongated by the successive addition of nucleotides to the free 3′ OH group of the preceding nucleotide. The phosphate to make the link is provided by the substrate which is the nucleoside 5′-triphosphate (ATP, GTP, CTP and UTP for RNA; dATP, dGTP, dCTP and dTTP for DNA).

1.2 Replication of DNA

A DNA strand can act as a template for synthesis of a new nucleic acid strand in which each base forms a hydrogen-bonded pair with one on the template strand (G with C, A with T, or A with U for RNA molecules). The new sequence is thus complementary to the template strand. The copying of DNA molecules to produce more DNA is known as *replication*; the synthesis of RNA using a DNA template is called *transcription*.

Replication is a much more complicated process than implied by the above statement. Some of the main features are summarized in Figure 1.9. The opposite polarity of the DNA strands is a complicating factor. One of the new strands (the 'leading' strand) can be synthesized continuously in the 5′ to 3′ direction. The enzyme responsible for this synthesis is DNA polymerase III. With the other new strand, however, the overall effect is of growth in the 3′ to 5′ direction. Since nucleic acids can only be synthesized in the 5′ to 3′ direction, the new 3′ to 5′ strand (the 'lagging' strand) has to be made

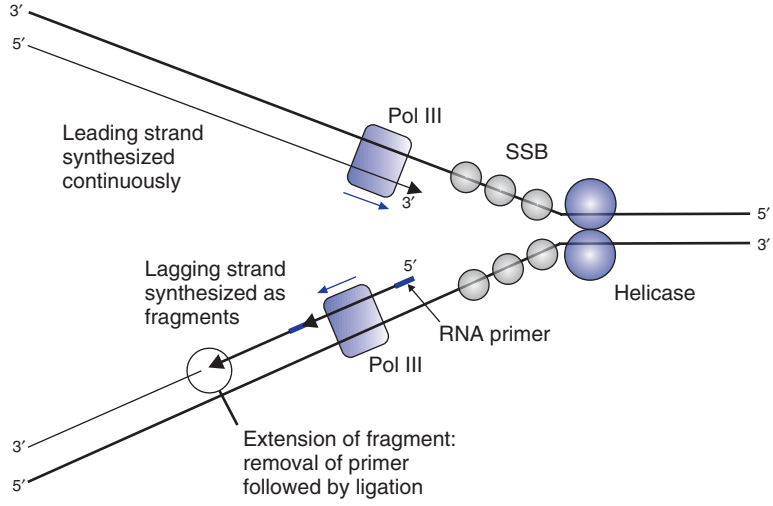

Figure 1.9 Simplified view of the main features of DNA replication. Note that the diagram does not show the helical structure of the DNA.

'backwards', i.e. in the opposite direction to overall replication. This can be done by making the new strand in short fragments (known as Okazaki fragments) which are subsequently joined together by the action of another enzyme, DNA ligase.

Furthermore, DNA polymerases are incapable of starting a new DNA strand, but can only extend a previously existing molecule. This restriction does not apply to RNA polymerases, which are able to initiate synthesis of new nucleic acids. Each fragment is therefore started with a short piece of RNA, produced by the action of a special RNA polymerase (*primase*). This RNA primer can then be extended by DNA polymerase III. The primer is subsequently removed, and the gap filled in, by a different DNA polymerase (DNA polymerase I); this enzyme can carry out both of these actions since it has exonuclease as well as polymerase activity. After the gap has been filled, the fragments that have been produced are joined together by DNA ligase.

1.2.1 Unwinding and rewinding

Before any of these events can take place, it is necessary for the two strands to be separated, for a short region at least. This is achieved by enzymes known as *helicases* which bind to the template strand and move along it, separating the two strands. The separated strands are prevented from re-associating by the binding of another protein, the single-stranded DNA-binding protein or SSB. A number of copies of the SSB will bind to the DNA strands, maintaining a region of DNA in an extended single-stranded form.

A further complication arises from the twisting of the two DNA strands around each other. DNA molecules within the cell cannot normally rotate freely, and not only because bacterial DNA is usually circular. Therefore it is not possible to produce a pair of daughter molecules by just separating the two strands and synthesizing the complementary strands, as is implied by the simplified representation in Figure 1.9. The strands have to be unwound to be separated. If they are not free to rotate, separating the strands at one point will cause overwinding further along. Unless this problem is overcome, the molecules would quickly get in a hopeless tangle. (If you don't understand this, try it for yourself with some bits of string!) The resolution of the problem requires the action of topoisomerases, as described earlier. By allowing the double helix to unwind ahead of the replication fork, they permit the strands to separate for replication. One topoisomerase, DNA gyrase, has the important role of introducing negative supercoils into the newly replicated DNA.

1.2.2 Fidelity of replication; proofreading

It is essential that the newly synthesized DNA is a precise (complementary) copy of the template strand. This does not arise simply by the nucleotides

aligning themselves in the right position, but involves the specificity of the DNA polymerase in selecting nucleotides that are correctly aligned.

Most DNA polymerases are more complex enzymes than the name suggests, as they also possess exonuclease activity. We have already encountered one such activity: the removal of the RNA primer from the Okazaki fragments is achieved by means of the 5′ to 3′ exonuclease activity of the DNA polymerase (i.e. it can remove bases from the 5′ end of a chain) as it extends the next fragment. The fidelity of replication is enhanced by a second exonuclease function of DNA polymerases: the 3′ to 5′ exonuclease activity, which is able to remove the nucleotide at the growing end (3′ end) of the DNA chain. This is not as perverse as it sounds, since the 3′ to 5′ exonuclease only operates if there is an incorrectly paired base at the 3′ end. The DNA polymerase will only extend the DNA chain, by adding nucleotides to the 3′ end, if the last base at the 3′ end is correctly paired with the template strand. If it is not, polymerization will stop, and the 3′ to 5′ exonuclease function will remove the incorrect nucleotide, allowing a further attempt to be made (Figure 1.10). The reasons for the occurrence of errors in adding bases to the growing DNA chain are dealt with in Chapter 2.

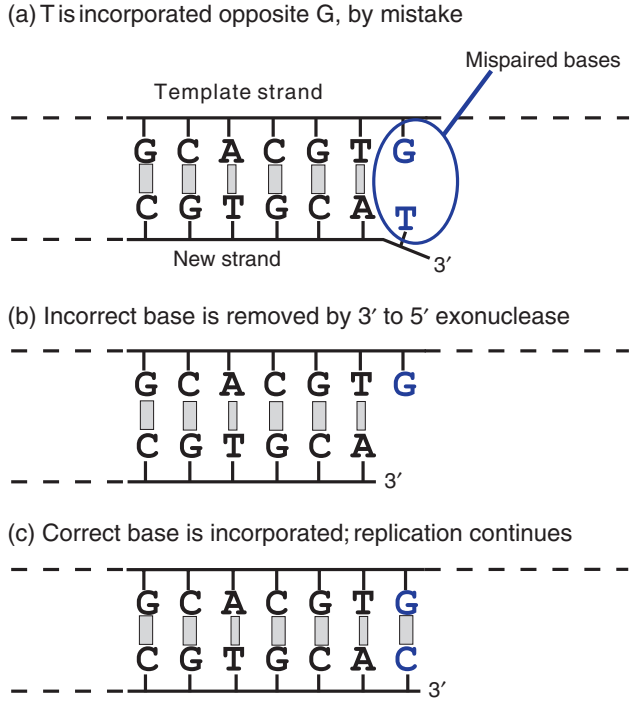

Figure 1.10 Elimination of mispaired bases by proofreading. (a) An incorrect base has been added to the growing DNA strand; this will prevent further extension. (b) The mispaired base is removed by the 3′ to 5′ exonuclease action of DNA polymerase. (c) The correct base is added to the 3′ end; DNA synthesis continues.

This mechanism of correcting errors, known as proofreading or error-checking, adds considerably to the fidelity of replication, thus reducing the rate of spontaneous mutation. There is a price to be paid however, as extensive error-checking will slow down the rate of replication. The balance between the rate of replication and the extent of error-checking will be determined by the nature of the DNA polymerase itself. Some DNA polymerases do not show efficient proofreading and therefore result in a much higher degree of spontaneous errors. The rate of spontaneous mutation shown by an organism is therefore (at least in part) a genetic characteristic that is subject to evolutionary pressure.

The fidelity of replication is further enhanced by DNA repair mechanisms, which are described later in this chapter.

1.3 Chromosome replication and cell division

Bacterial cells are generally regarded as having a single, circular chromosome. This is a simplification in several ways. Firstly, many bacteria often contain additional DNA molecules known as *plasmids*. In most cases these are additional, dispensable, elements, but in some bacterial species all strains carry two or more different DNA molecules, both (or all) of which appear to be essential for normal growth. These can equally well be regarded as essential plasmids or as additional chromosomes. Secondly, not all bacterial DNA is circular. Some bacteria (notably *Streptomyces*) have a linear chromosome and/or linear plasmids. These topics are discussed further in Chapter 5.

More fundamentally, immediately before cell division there must be at least two complete copies of the chromosome, in order to ensure that both daughter cells acquire a copy. Therefore, the chromosome must replicate in tune with the cell-division cycle, which means that at an intermediate time in the cycle part of the chromosome will have been copied, with the consequence that there are at least two copies of this part of the chromosome.

Replication of a bacterial chromosome normally starts at a fixed point (the origin of replication, *oriV*) and proceeds in both directions to a termination point (*ter*) that is approximately opposite to the origin (Figure 1.11). In *E. coli*, this takes about 40 minutes. There is then a period of at least 20 minutes before cell division, making a minimum of 60 minutes between the initiation of replication and cell division. However under favourable conditions *E. coli* will grow much faster than that, dividing perhaps every 20 minutes. How can the cells be dividing faster than the chromosome replicates, and still allow every daughter cell to acquire a complete copy of the chromosome? The answer lies in the timing of the initiation of replication. Initiation is stimulated, not by cell division, but as a function of the size of the cell.

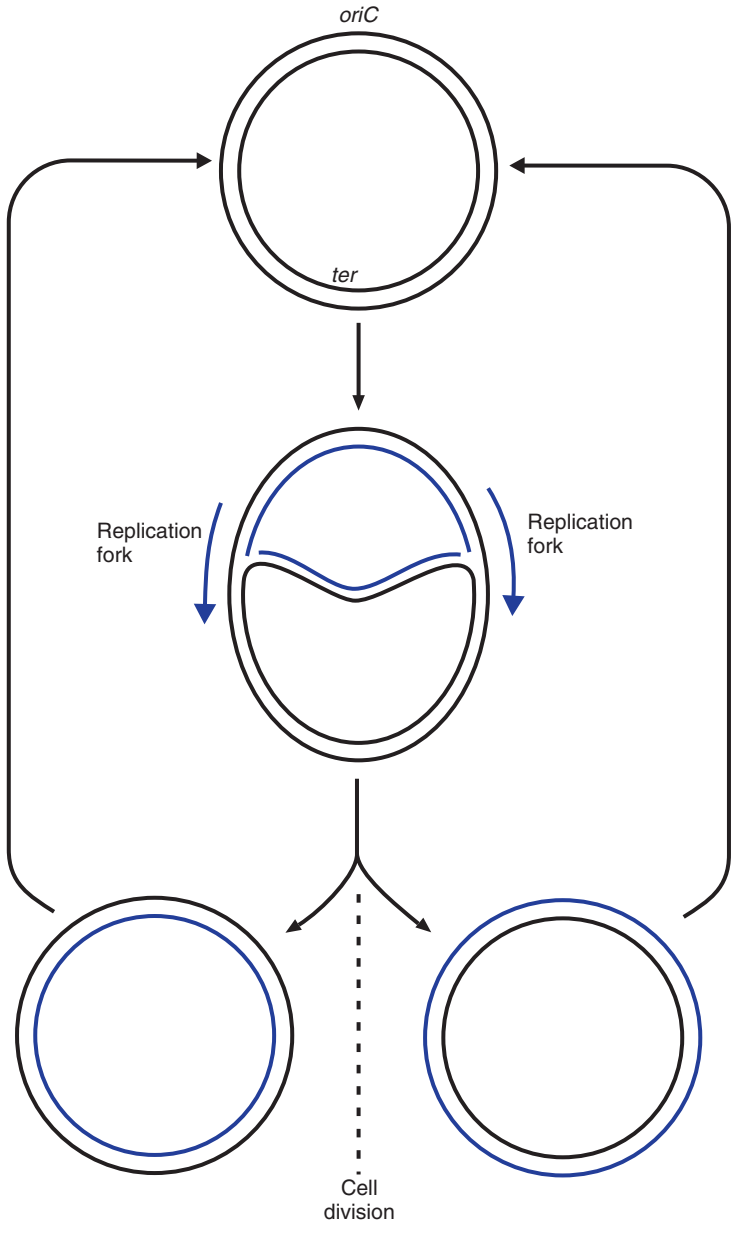

Figure 1.11 Chromosome replication. Bidirectional replication starts at *oriC* and continues to the termination site *ter*, producing two double-stranded molecules.

Consequently, when the cells are growing rapidly, there are several sets of replication forks copying the chromosome, so that when the cell is ready to divide there are not just a pair of completely replicated chromosomes, but each of these has in turn already been partly replicated by a second pair of replication forks (Figure 1.12).

There are two key regulatory points to be considered: the link between the completion of chromosome replication and subsequent cell division, and the control of the initiation of replication. Both are too complex (and still incompletely understood) to be considered fully here, but we can consider some aspects. On the first point, one (simplified) model is that the replicating chromosome occupies a region of the membrane at the midpoint of the cell, which prevents the initiation of cell division at that point. When replication has finished, the two separate molecules can be pulled apart, towards the poles of the cell, thus freeing the site for cell division to start. We will look further at this in Chapter 9.

The second point, the control of initiation, is more difficult. Initiation of new rounds of replication is triggered when the cell reaches a critical mass. It is tempting to think that this means that an inhibitor of replication is diluted out as the cell grows, but it is far from that simple. We know that initiation requires a protein called DnaA which binds to specific DNA sequences known as *DnaA boxes*; the origin of replication contains a number of DnaA boxes. Wrapping the DNA around the aggregated DnaA proteins facilitates the separation of the strands that is necessary for the initiation of replication (Figure 1.13). The availability of active DnaA is a significant component of initiation. However, the full story of the control of initiation is more complex, and still incompletely understood.

1.4 DNA repair

In addition to the measures, described earlier, that enhance the fidelity of replication, the cell also possesses mechanisms to correct damaged DNA, including replication errors that have escaped the proofreading process and damage that may have occurred in non-replicating DNA.

1.4.1 Mismatch repair

The simplest of replication errors is one that leads to the wrong base being incorporated into the new strand. If this occurs, and is not dealt with by the proofreading mechanism, it would lead to mutation. However, the cell has an effective mechanism for removing such mismatches, and replacing them by the correct nucleotide. In order to do this, it has to know which of the two strands contains the correct information. Mature DNA is methylated, i.e. it contains additional methyl groups, especially on the adenine in the sequence GATC, due to the action of deoxyadenosine methylase (Dam methylase). The new