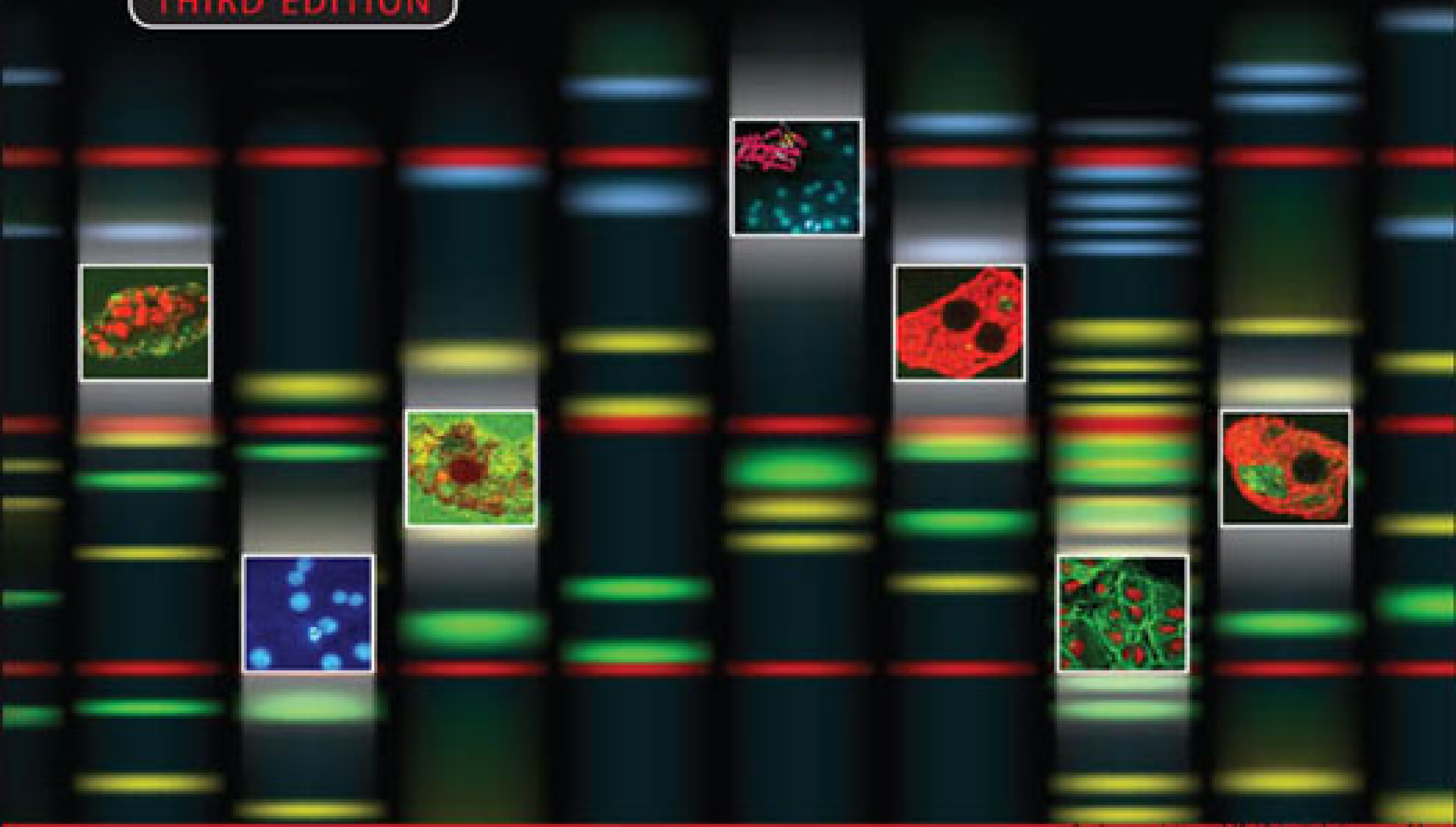
JEREMY W. DALE | MALCOLM VON SCHANTZ | NICK PLANT

FROM GENES TO GENOMES

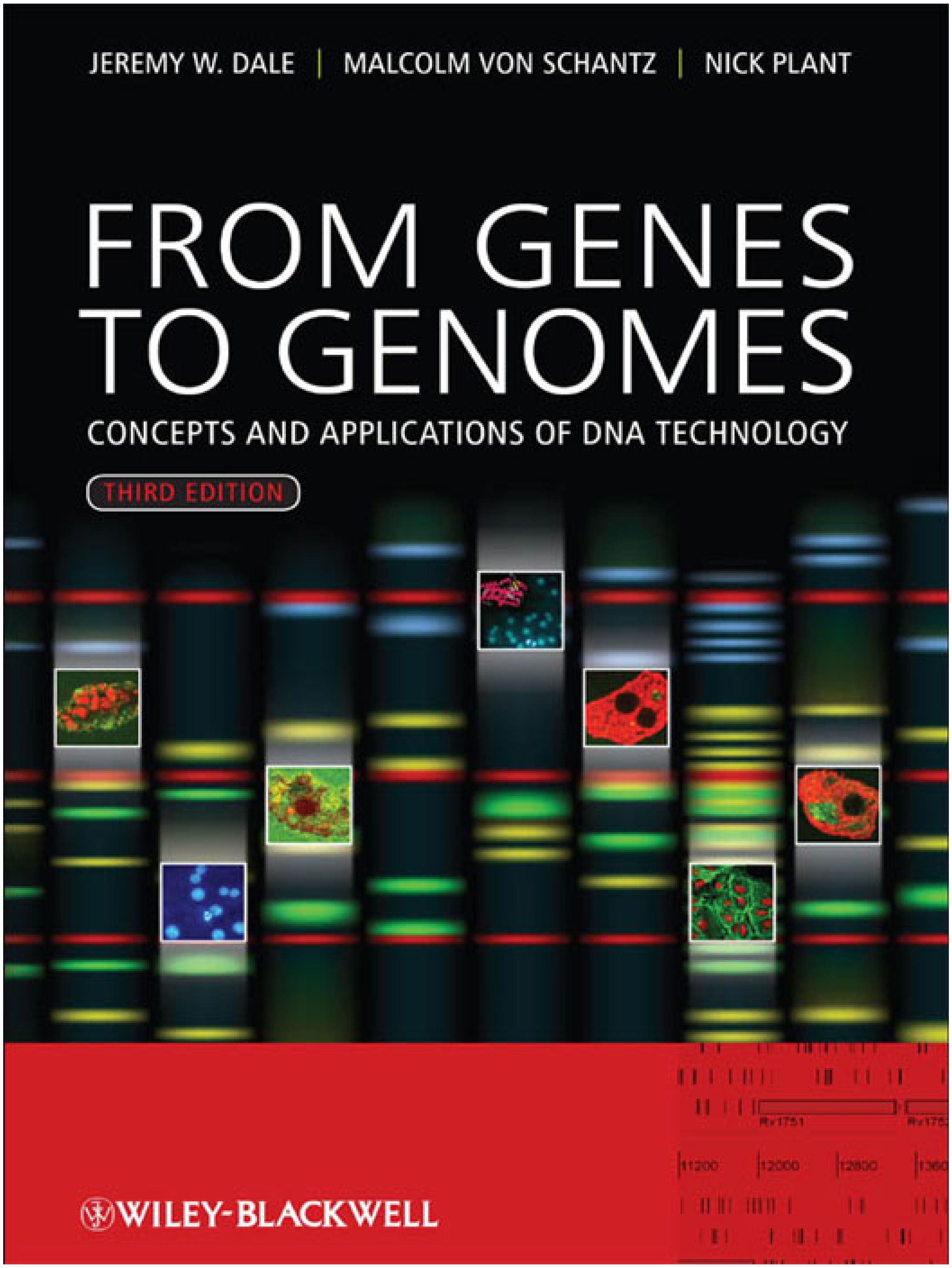
CONCEPTS AND APPLICATIONS OF DNA TECHNOLOGY

THIRD EDITION



Rv1751
Rv175
11200
12000
12800
1360

WILEY-BLACKWELL



JEREMY W. DALE | MALCOLM VON SCHANTZ | NICK PLANT

FROM GENES TO GENOMES

CONCEPTS AND APPLICATIONS OF DNA TECHNOLOGY

THIRD EDITION

Rv1751
11200
12000
12800

WILEY-BLACKWELL

Contents

Cover

Title Page

Copyright

Preface

1: From Genes to Genomes
- 1.1 Introduction
- 1.2 Basic molecular biology
- 1.3 What is a gene?
- 1.4 Information flow: gene expression
- 1.5 Gene structure and organisation
- 1.6 Refinements of the model

2: How to Clone a Gene
- 2.1 What is cloning?
- 2.2 Overview of the procedures
- 2.3 Extraction and purification of nucleic acids
- 2.4 Detection and quantitation of nucleic acids
- 2.5 Gel electrophoresis
- 2.6 Restriction endonucleases
- 2.7 Ligation

- 2.8 Modification of restriction fragment ends
- 2.9 Plasmid vectors
- 2.10 Vectors based on the lambda bacteriophage
- 2.11 Cosmids
- 2.12 Supervectors: YACs and BACs
- 2.13 Summary

3: Genomic and cDNA Libraries

- 3.1 Genomic libraries
- 3.2 Growing and storing libraries
- 3.3 cDNA libraries
- 3.4 Screening libraries with gene probes
- 3.5 Screening expression libraries with antibodies
- 3.6 Characterization of plasmid clones

4: Polymerase Chain Reaction (PCR)

- 4.1 The PCR reaction
- 4.2 PCR in practice
- 4.3 Cloning PCR products
- 4.4 Long-range PCR
- 4.5 Reverse-transcription PCR
- 4.6 Quantitative and real-time PCR
- 4.7 Applications of PCR

5: Sequencing a Cloned Gene

- 5.1 DNA sequencing

- 5.2 Databank entries and annotation
- 5.3 Sequence analysis
- 5.4 Sequence comparisons
- 5.5 Protein structure
- 5.6 Confirming gene function

6: Analysis of Gene Expression
- 6.1 Analysing transcription
- 6.2 Methods for studying the promoter
- 6.3 Regulatory elements and DNA-binding proteins
- 6.4 Translational analysis

7: Products from Native and Manipulated Cloned Genes
- 7.1 Factors affecting expression of cloned genes
- 7.2 Expression of cloned genes in bacteria
- 7.3 Yeast systems
- 7.4 Expression in insect cells: baculovirus systems
- 7.5 Mammalian cells
- 7.6 Adding tags and signals
- 7.7 In vitro mutagenesis
- 7.8 Vaccines

8: Genomic Analysis
- 8.1 Overview of genome sequencing
- 8.2 Next generation sequencing (NGS)

8.3 De novo sequence assembly
8.4 Analysis and annotation
8.5 Comparing genomes
8.6 Genome browsers
8.7 Relating genes and functions: genetic and physical maps
8.8 Transposon mutagenesis and other screening techniques
8.9 Gene knockouts, gene knockdowns and gene silencing
8.10 Metagenomics
8.11 Conclusion

9: Analysis of Genetic Variation

9.1 Single nucleotide polymorphisms
9.2 Larger scale variations
9.3 Other methods for studying variation
9.4 Human genetic variation: relating phenotype to genotype
9.5 Molecular phylogeny

10: Post-Genomic Analysis

10.1 Analysing transcription: transcriptomes
10.2 Array-based methods
10.3 Transcriptome sequencing
10.4 Translational analysis: proteomics
10.5 Post-translational analysis: protein interactions
10.6 Epigenetics

- 10.7 Integrative studies: systems biology

11: Modifying Organisms: Transgenics

- 11.1 Transgenesis and cloning
- 11.2 Animal transgenesis
- 11.3 Applications of transgenic animals
- 11.4 Disease prevention and treatment
- 11.5 Transgenic plants and their applications
- 11.6 Transgenics: a coda

Glossary

Bibliography

- General books
- Laboratory manuals
- Special topics
- Websites

Index

From Genes to Genomes

Third Edition

Concepts and Applications of DNA Technology

Jeremy W. Dale, Malcolm von Schantz and **Nick Plant**

University of Surrey, UK

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

This edition first published 2012
© 2012 by John Wiley & Sons, Ltd.

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

Editorial offices
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, 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.

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Dale, Jeremy, Professor.
From genes to genomes : concepts and applications of DNA technology / Jeremy W. Dale, Malcolm von Schantz, and Nick Plant. – 3rd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-68386-6 (cloth) – ISBN 978-0-470-68385-9 (pbk.)
I. Schantz, Malcolm von. II. Plant, Nick. III. Title.
[DNLM: 1. Genetic Engineering. 2. Cloning, Molecular. 3. DNA, Recombinant. QU 450]
LC classification not assigned
660.6′5–dc23

2011030219

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

This book is published in the following electronic formats: ePDF 9781119953159; ePub 9781119954279; Mobi 9781119954286

Preface

The first edition of this book was published in 2002. By the time of the second edition (2007) the emphasis had moved away from just cloning genes, to embrace a wider range of technologies, especially genome sequencing, the polymerase chain reaction and microarray technology. The revolution has continued unabated, indeed even accelerating, not least with the advent of high-throughput genome sequencing. In this edition we have tried to introduce readers to the excitement engendered by the latest developments – but this poses a considerable challenge. Our aim has been to keep the book to an accessible size, so including newer technologies inevitably means discarding some of the older ones. Some might maintain that we could have gone further in that direction. Some methods that have been kept are no longer as important as they once were, and maybe there is an element of sentimentality in keeping them – but there is some virtue in retaining a balance so that we can maintain a degree of historical perspective. There is a need to understand, to some extent, how we got to the position we are now in, as well as trying to see where we are going.

The main title of the book, *From Genes to Genomes*, is derived from the progress of this revolution. It also indicates a recurrent theme within the book, in that the earlier chapters deal with analysis and investigation at the level of individual genes, and then later on we move towards genome-wide studies – ending up with a chapter directed at the whole organism.

Dealing only with the techniques, without the applications, would be rather dry. Some of the applications are obvious – recombinant product formation, genetic diagnosis, transgenic plants and animals, and so on – and we have attempted to introduce these to give you a flavour of the advances that continue to be made, but at the same time without burdening you with excessive detail. Equally important, possibly more so, are the contributions made to the advance of fundamental knowledge in areas such as developmental studies and molecular phylogeny.

The purpose of this book is to provide an introduction to the concepts and applications of this rapidly moving and fascinating field. In writing it, we had in mind its usefulness for undergraduate students in the biological and biomedical sciences (who we assume will have a basic grounding in molecular biology). However, it will also be relevant for many others, ranging from research workers and teachers who want to update their knowledge of related areas to anyone who would like to understand rather more of the background to current controversies about the applications of some of these techniques.

Jeremy W. Dale
Malcolm von Schantz
Nick Plant

1

From Genes to Genomes

1.1 Introduction

The classical approach to genetics starts with the identification of variants that have a specific *phenotype*, i.e., they differ from the *wildtype* in some way that can be seen (or detected in other ways) and defined. For Gregor Mendel, the father of modern genetics, this was the appearance of his peas (e.g., green versus yellow, or round versus wrinkled). One of the postulates he arrived at was that these characteristics assorted independently of one another. For example, when crossing one type of pea that produces yellow, wrinkled peas with another that produces green, round peas, the first generation (F_1) are all round and yellow (because round is dominant over wrinkled, and yellow is dominant over green). In the second (F_2) generation, there is a 3 : 1 mixture of round versus wrinkled peas, and independently a 3 : 1 mixture of yellow to green peas.

Of course Mendel did not know why this happened. We now know that if two genes are located on different chromosomes, which will segregate independently during meiosis, the genes will be distributed independently amongst the progeny. Independent assortment can also happen if the two genes are on the same chromosome, but only if they are so far apart that any recombination between the homologous chromosomes will be sufficient to reassort them independently. However, if they are quite close together, recombination is less likely, and they will therefore tend to remain associated during meiosis. They will therefore be inherited together. We refer to genes that do *not* segregate independently as *linked*; the closer they are, the greater the degree of linkage, i.e., the more likely they are to stay together during meiosis. Measuring the degree of linkage (*linkage analysis*) is a central tool in classical genetics, in that it provides a way of mapping genes, i.e., determining their relative position on the chromosome.

Bacteria and yeasts provide much more convenient systems for genetic analysis, because they grow quickly, as unicellular organisms, on defined media. You can therefore use chemical or physical mutagens (such as ultraviolet irradiation) to produce a wide range of mutations, and can select specific mutations from very large pools of organisms – remembering that an overnight culture of *Escherichia coli* will contain some 10^9 bacteria per millilitre. So we can use genetic techniques to investigate detailed aspects of the physiology of such cells, including identifying the relevant genes by mapping the position of the mutations.

For multicellular organisms, the range of phenotypes is even greater, as there are then questions concerning the development of different parts of the organism, and how each individual part influences the development of others. However, animals have much longer generation times than bacteria, and using millions of animals (especially mammals) to identify the mutations you are interested in is logistically impossible, and ethically indefensible. Human genetics is even more difficult as you cannot use selected breeding to map genes; you have to rely on the analysis of real families, who have chosen to breed with no consideration for the needs of science. Nevertheless, classical genetics has contributed extensively to the study of developmental processes, notably in the fruit fly *Drosophila melanogaster*, where it is possible to study quite large numbers of animals, due to their relative ease of housing and short generation times, and to use mutagenic agents to enhance the rate of variation.

However, these methods suffered from a number of limitations. In particular, they could only be applied, in general, to mutations that gave rise to a phenotype that could be defined in some way, including shape, physiology, biochemical properties or behaviour. Furthermore, there was no easy way of characterizing the nature of the mutation. The situation changed radically in the 1970s with the development of techniques that enabled DNA to be cut precisely into specific fragments, and to be joined together, enzymatically –techniques that became known variously as genetic manipulation, genetic modification, genetic engineering or recombinant DNA technology. The term ‘gene cloning’ is also used, since joining a fragment of DNA with a vector such as a plasmid that can replicate in bacterial cells enabled the production of a bacterial strain (a clone) in which all the cells contained a copy of this specific piece of DNA. For the first time, it was possible to isolate and study specific genes. Since such techniques could be applied equally to human genes, the impact on human genetics was particularly marked.

The revolution also depended on the development of a variety of other molecular techniques. The earliest of these (actually predating gene cloning) was *hybridization*, which enabled the identification of specific DNA sequences on the basis of their sequence similarity. Later on came methods for determining the sequence of these DNA fragments, and the polymerase chain reaction (PCR), which provided a powerful way of amplifying specific DNA sequences. Combining those advances with automation, plus the concurrent advance in computer power, led to the determination of the full genome sequence of many organisms, including the human genome, and thence to enormous advances in understanding the roles of genes and their products. In recent years, sequencing technology has advanced to a stage where it is now a routine matter to sequence the full genome of many individuals, and thus attempt to pinpoint the causes of the differences between them, including some genetic diseases.

Furthermore, since these techniques enabled the cloning and expression of genes from any one organism (including humans) into a more amenable host, such as a bacterium, they allowed the use of genetically modified bacteria (or other hosts) for the production of human gene products, such as hormones, for therapeutic use. This principle was subsequently extended to the genetic modification of plants and animals – both by inserting foreign genes and by knocking out existing ones – to produce plants and animals with novel properties.

As is well known, the construction and use of genetically modified organisms (GMOs) is not without controversy. In the early days, there was a lot of concern that the introduction of foreign DNA into *E. coli* would generate bacteria with dangerous properties. Fortunately, this is one fear that has been shown to be unfounded. Due to careful design, genetically modified bacteria are, generally, not well able to cope with life outside the laboratory, and hence any GM bacterium released into the environment (deliberately or accidentally) is unlikely to survive for long. In addition, one must recognize that nature is quite capable of producing pathogenic organisms without our assistance – which history, unfortunately, has repeatedly demonstrated through disease outbreaks.

The debate on GMOs has now largely moved on to issues relating to genetically modified plants and animals. It is important to distinguish the *genetic modification* of plants and animals from *cloning* of plants and animals. The latter simply involves the production of genetically identical individuals; it does not involve any genetic modification whatsoever. (The two technologies can be used in tandem, but that is another matter.) There are ethical issues to be considered, but cloning plants and animals is not the subject of this book.

Currently, the debate on genetic modification can be envisaged as largely revolving around two factors: food safety and environmental impact. The first thing to be clear about is that there is no imaginable reason why genetic modification, per se, should make a foodstuff hazardous in any way. There is no reason to suppose that cheese made with rennet from a genetically modified bacterium is any more dangerous than similar cheese made with ‘natural’ rennet. It is possible to imagine a risk associated with some genetically modified foodstuffs, due to unintended stimulation of the production of natural toxins – remembering, for example, that potatoes are related to deadly nightshade. But this can happen equally well (or perhaps is even more likely) with conventional cross-breeding procedures for developing new strains, which are not always subject to the same degree of rigorous safety testing as GM plants.

The potential environmental impact is more difficult to assess. The main issue here is the use of genetic modification to make plants resistant to herbicides or to insect attack. When such plants are grown on a large scale, it is difficult to be certain that the gene in question will not spread to related wild plants in the vicinity (although measures can be taken to reduce this possibility), or the knock-on effect that such resistance may have on the ecosystem – if all the insects are killed, what will small birds and animals eat? But these concerns may be exaggerated. As with the bacterial example above, these genes will not spread significantly unless there is an evolutionary pressure favouring them. So we would not expect widespread resistance to weedkillers unless the plants are being sprayed with those weedkillers. There might be an advantage in becoming resistant to insect attack, but the insects concerned have been around for a long time, so the wild plants have had plenty of time to develop natural resistance anyway. In addition, targeted resistance in a group of plants may arguably have less environmental impact than the less targeted spraying of insecticides. We have to balance the use of genetically modified plants against the use of chemicals. If genetic modification of the plants means a reduction in the use of environmentally damaging chemicals, then that is a tangible benefit that could outweigh any theoretical risk.

The purpose of this book is to provide an introduction to the exciting developments that have resulted in an explosion of our knowledge of the genetics and molecular biology of all forms of life, from viruses and bacteria to plants and mammals, including of course ourselves – developments that continue as we write. We hope that it will convey some of the wonder and intellectual stimulation that this science brings to its practitioners.

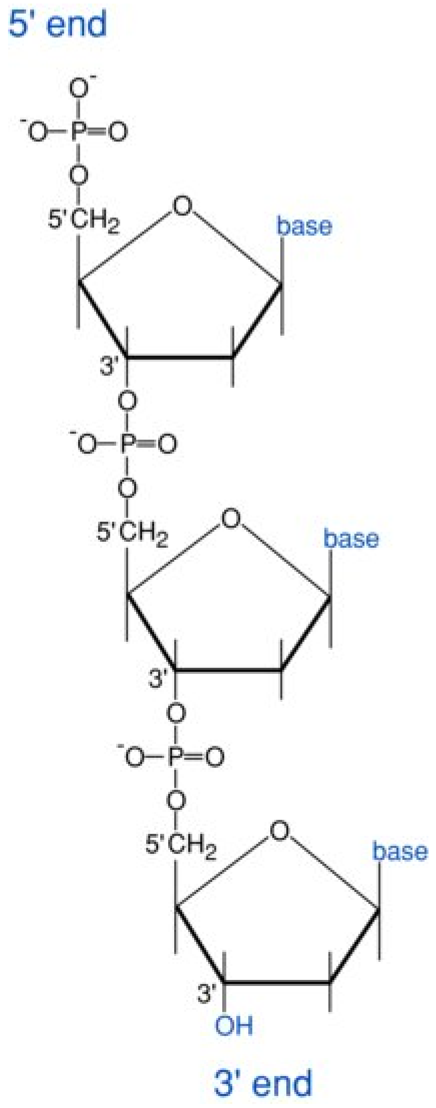
1.2 Basic molecular biology

In this book, we assume you already have a working knowledge of the basic concepts of molecular and cellular biology. This section serves as a reminder of the key aspects that are especially relevant to this book.

1.2.1 The DNA backbone

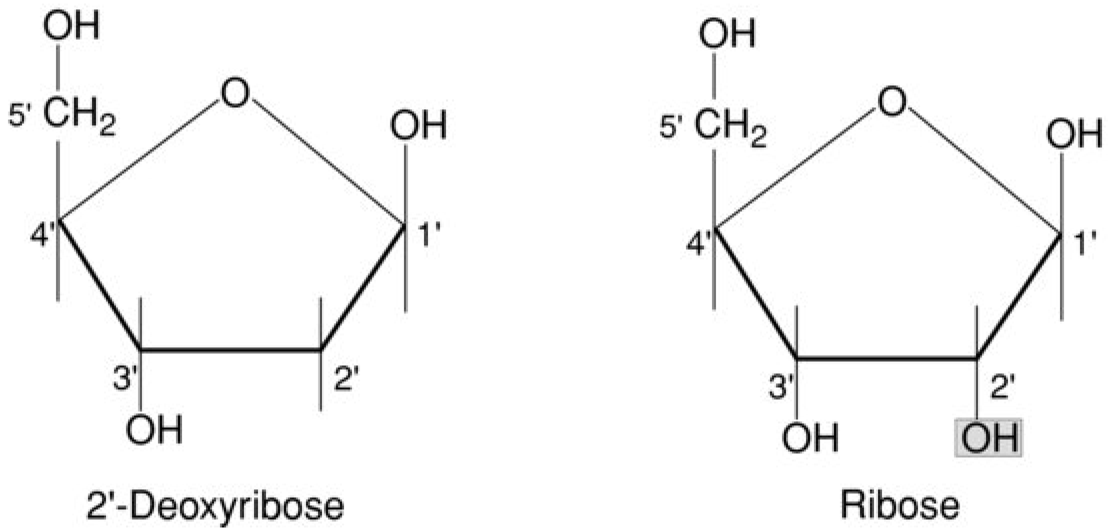
Manipulation of nucleic acids in the laboratory is based on their physical and chemical properties, which in turn are reflected in their biological function. Intrinsically, DNA is a remarkably stable molecule. Indeed, DNA of sufficiently high quality to be analysed has been recovered from frozen mammoths thousands of years old. This stability is provided by the robust phosphate–sugar backbone in each DNA strand, in which the phosphate links the 5′ position of one sugar to the 3′ position of the next (Figure 1.1). The bonds between these phosphorus, oxygen and carbon atoms are all *covalent bonds*, meaning they are strong interactions that require significant energy to break. Hence, the controlled degradation of DNA requires enzymes (nucleases) that catalyse the breaking of these covalent bonds. These enzymes are divided into *endonucleases*, which attack internal sites in a DNA strand, and *exonucleases*, which nibble away at the ends. (We can for the moment ignore other enzymes that attack, for example, the bonds linking the bases to the sugar residues.) Some of these enzymes are non-specific, and lead to a generalized destruction of DNA. It was the discovery of *restriction endonucleases* (or *restriction enzymes*), which cut DNA strands at specific positions, coupled with *DNA ligases*, which can join two double-stranded DNA molecules together, that opened up the possibility of *recombinant DNA technology* (‘*genetic engineering*’).

Figure 1.1 DNA backbone.


5' end
5'CH2
base
3'
5'CH2
base
3'
5'CH2
base
3'
OH
3' end

RNA molecules, which contain the sugar ribose (Figure 1.2), rather than the deoxyribose found in DNA, are less stable than DNA, often surviving only minutes within the cell. They show greater susceptibility to attack by nucleases (*ribonucleases*), and are also more susceptible to chemical degradation, especially by alkaline conditions.

Figure 1.2 Nucleic acid sugars.



1.2.2 The base pairs

In addition to the sugar (2′-deoxyribose) and phosphate, DNA molecules contain four nitrogen-containing bases (Figure 1.3): two pyrimidines, thymine (T) and cytosine (C), and two purines, guanine (G) and adenine (A). It should be noted that other bases can be incorporated into synthetic DNA in the laboratory, and sometimes others occur naturally, but T, C, G and A are the major DNA bases. Because the purines are bigger than the pyrimidines, a regular double helix requires a purine in one strand to be matched by a pyrimidine in the other. Furthermore, the regularity of the double helix requires specific hydrogen bonding between the bases so that they fit together, with an A opposite a T, and a G opposite a C (Figure 1.4). We refer to these pairs of bases as *complementary*, and hence to each strand as the *complement* of the other.

Figure 1.3 Nucleic acid bases.

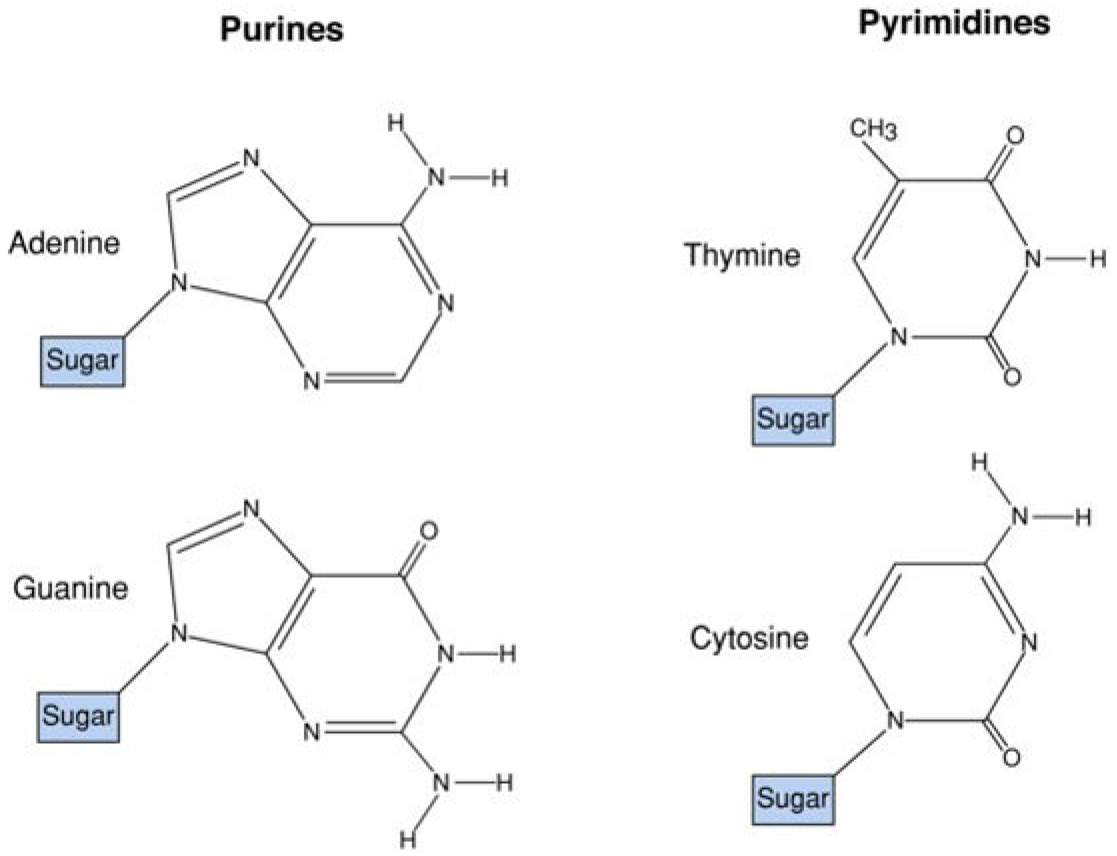
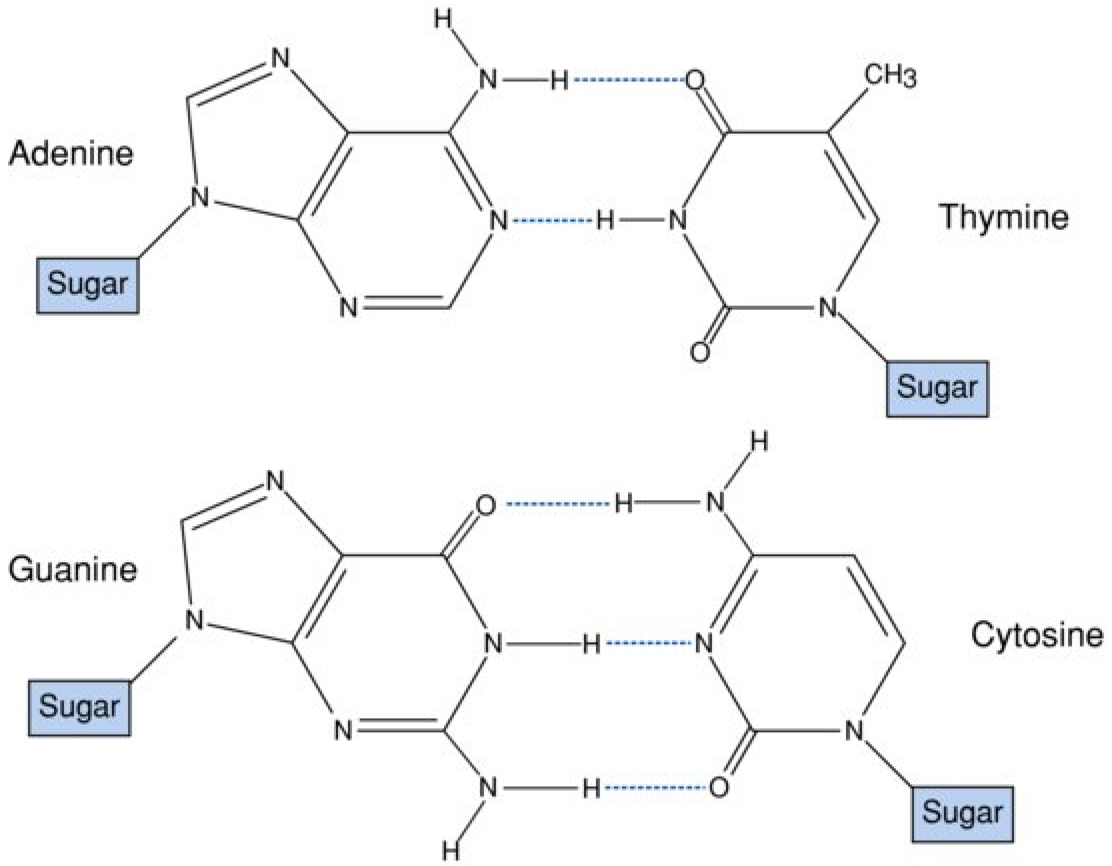

Purines
Pyrimidines
Adenine
Thymine
Guanine
Cytosine
Sugar
Sugar
Sugar
Sugar
CH3
N
H
O

Figure 1.4 Base-pairing in DNA.



Note that the two DNA strands run in opposite directions. In a conventional representation of a double-stranded sequence the ‘top’ strand has a 5′ hydroxyl group at the left-hand end (and is said to be written in the 5′ to 3′ direction), while the ‘bottom’ strand has its 5′ end at the right-hand end. Because the two strands are complementary, there is no information in one strand that cannot be deduced from the other one. Therefore, to save space, the convention is to represent a double-stranded DNA sequence by showing the sequence of only one strand, in the 5′ to 3′ direction. The sequence of the second strand is inferred from that, and you must remember that the second strand runs in the opposite direction. Thus a single strand sequence written as AGGCTG (or more fully 5′AGGCTG3′) would have as its complement CAGCCT (5′CAGCCT3′) (see Box 1.1).

Box 1.1 Complementary sequences

DNA sequences are often represented as the sequence of just one of the two strands, in the 5′ to 3′ direction, reading from left to right. Thus the double-stranded DNA sequence

$$\begin{array}{l} 5' - \mathrm{AGGCTG} - 3' \\ 3' - \mathrm{TCCGAC} - 5' \end{array}$$

would be shown as AGGCTG, with the orientation (i.e., the position of the 5′ and 3′ ends) being implied.

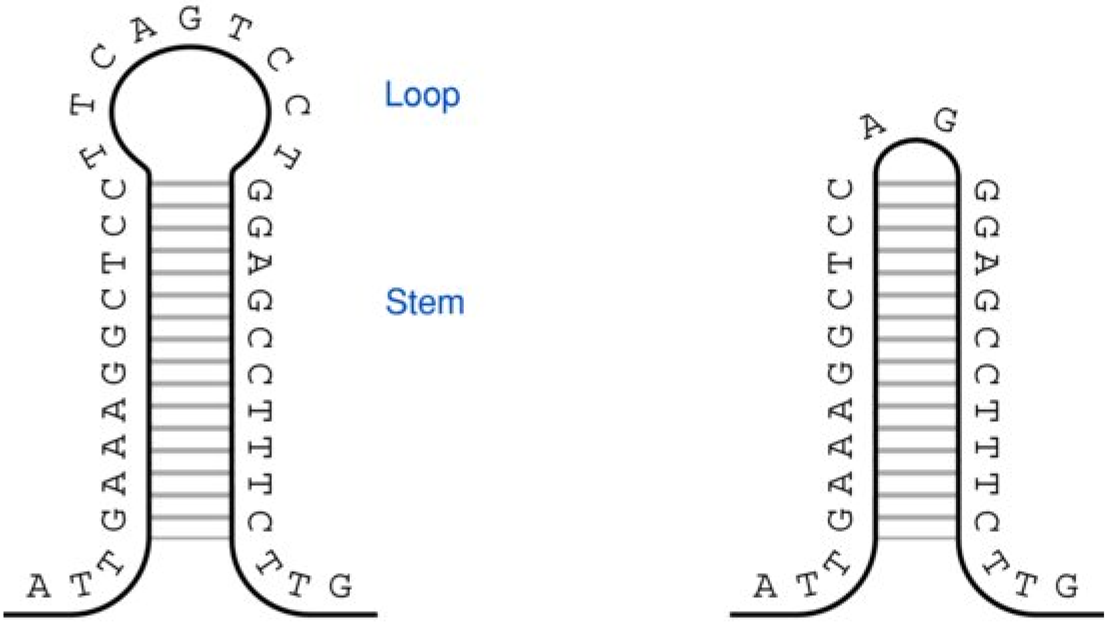
To get the sequence of the other (complementary) strand, you must not only change the A and G residues to T and C (and vice versa), but you must also reverse the order. So in this example, the complement of AGGCTG is CAGCCT, reading the lower strand from right to left (again in the 5′ to 3′ direction).

Thanks to this base-pairing arrangement, the two strands can be separated intact – both in the cell and in the test tube – under conditions that, although disrupting the relatively weak hydrogen bonds that exist between the bases on complementary strands, are much too mild to pose any threat to the covalent bonds that join nucleotides within a single strand. Such a separation is referred to as *denaturation* of DNA, and unlike the denaturation of many proteins it is reversible. Because of the complementarity of the base pairs, the strands will easily join together again and *renature* by reforming their hydrogen bonds. In the test tube, DNA is readily denatured by heating, and the denaturation process is therefore often referred to as *melting*, even when it is accomplished by means other than heat (e.g. by NaOH). Denaturation of a double-stranded DNA molecule occurs over a short, specific temperature range, and the midpoint of that range is defined as the *melting temperature* (T_m). This is influenced by the base composition of the DNA. Since guanine:cytosine (GC) base pairs have three hydrogen bonds, they are stronger (i.e., melt less easily) than adenine:thymine (AT) pairs, which have only two. It is therefore possible to estimate the melting temperature of a DNA fragment if you know the sequence. These considerations are important in understanding the technique known as *hybridization*, in which *gene probes* are used to detect specific nucleic acid sequences. We will look at hybridization in more detail in Chapter 3.

In addition to the hydrogen bonds, the double-stranded DNA structure is maintained by *hydrophobic interactions* between the bases. The hydrophobic nature of the bases means that a single-stranded structure, in which the bases are exposed to the aqueous environment, is unstable. Pairing of the bases enables them to be removed from interaction with the surrounding water, introducing significant stability to the DNA helix. In contrast to hydrogen bonding, hydrophobic interactions are relatively non-specific. Thus, nucleic acid strands will tend to stick together even in the absence of specific base-pairing, although the specific interactions make the association stronger. The specificity of the interaction can therefore be increased by the use of chemicals (such as formamide) that reduce the hydrophobic interactions.

What happens if there is only a single nucleic acid strand? This is normally the case with RNA, but single-stranded forms of DNA (ssDNA) also exist, in some viruses, for example. A single-stranded nucleic acid molecule will tend to fold up on itself to form localized double-stranded regions, producing structures referred to as hairpins or stem-loop structures (Figure 1.5). This has the effect of removing the bases from interaction with the surrounding water, again increasing stability. At room temperature, in the absence of denaturing agents, a single-stranded nucleic acid will normally consist of a complex set of such localized secondary structure elements; this is especially evident with RNA molecules.

Figure 1.5 Stem-loop and hairpin structures.


Loop
Stem
ATTGAAAGGCTCCTTCAGTCCTGGAGCCTTTCTTG
ATTGAAAGGCTCCAGGGAGCCTTTCTTG

Stem-loop structure

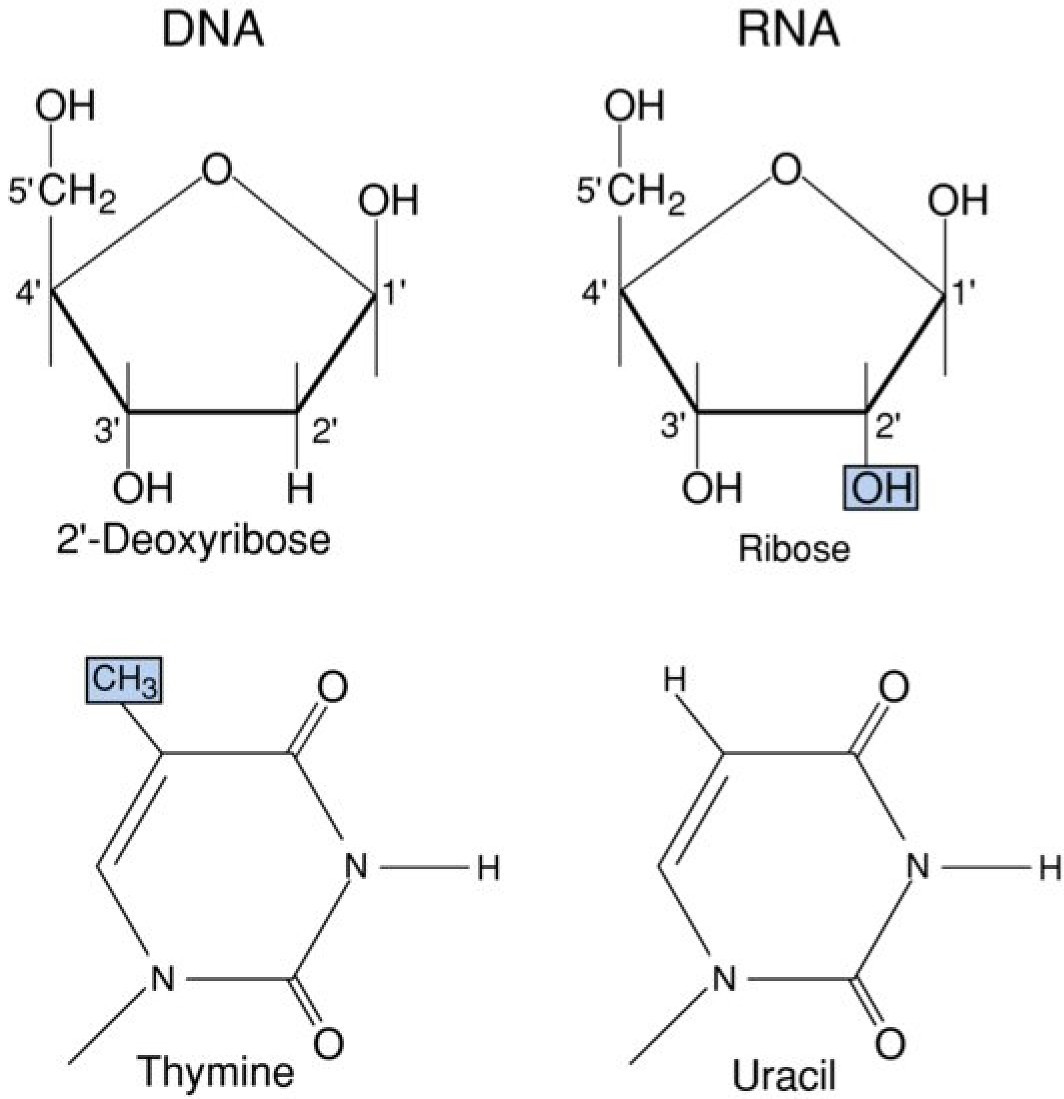
Hairpin

A further factor is the negative charge on the phosphate groups in the nucleic acid backbone. This works in the opposite direction to the hydrogen bonds and hydrophobic interactions; the strong negative charge on the DNA strands causes electrostatic repulsion that tends to make the two strands repel each other. In the presence of salt, this effect is counteracted by a cloud of counterions surrounding the molecule, neutralizing the negative charge on the phosphate groups. However, if you reduce the salt concentration, any weak interactions between the strands will be disrupted by electrostatic repulsion. Hence, at low ionic strength, the strands will only remain together if the hydrogen bonding is strong enough, and therefore we can use low salt conditions to increase the specificity of hybridization (see Chapter 3). Of course, within the cell the salt concentration is such that the double-stranded DNA is quite stable.

1.2.3 RNA structure

Chemically, RNA is very similar to DNA. The fundamental chemical difference is that whereas DNA contains 2′-deoxyribose (i.e., ribose without the hydroxyl group at the 2′ position) in its backbone, the RNA backbone contains ribose (Figure 1.6). This slight difference has a powerful effect on some properties of the RNA molecule, especially on its stability. For example, RNA is destroyed under alkaline conditions while DNA is stable. Although the DNA strands will separate, they will remain intact and capable of renaturation when the pH is lowered again. However, under such conditions, RNA will quickly be destroyed. A further difference between RNA and DNA is that the former contains uracil rather than thymine (Figure 1.6).

Figure 1.6 Differences between DNA and RNA.

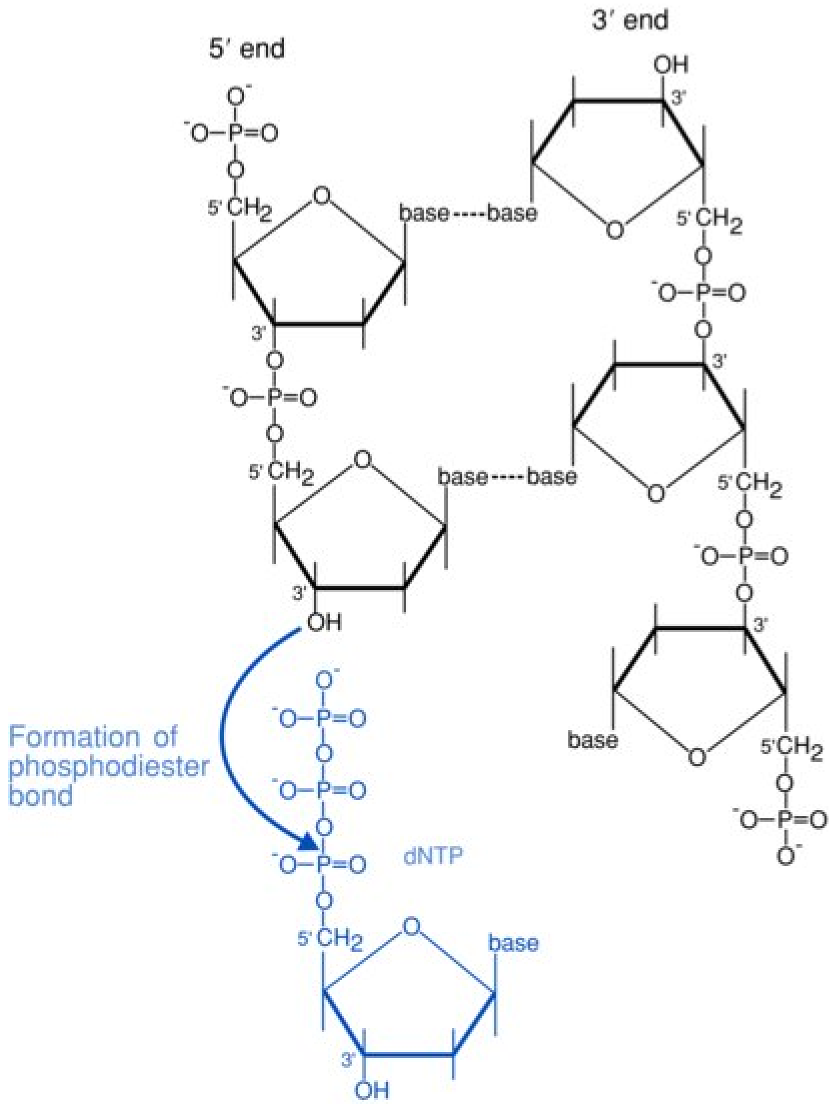

DNA
RNA
OH
5'CH2
O
OH
4'
1'
3'
2'
OH
H
2'-Deoxyribose
OH
5'CH2
O
OH
4'
1'
3'
2'
OH
OH
Ribose
CH3
O
N
H
N
O
Thymine
H
O
N
H
N
O
Uracil

Generally, while most of the DNA we encounter is double-stranded, most of the RNA we meet consists of a single polynucleotide strand. However, DNA can also exist as a single-stranded molecule, and RNA is able to form double-stranded molecules. Thus, this distinction between RNA and DNA is not an inherent property of the nucleic acids themselves, but is a reflection of the natural roles of RNA and DNA in the cell, and of the method of production. In all *cellular* organisms (i.e., excluding viruses), DNA is the inherited material responsible for the genetic composition of the cell, and the replication process that has evolved is based on a double-stranded molecule. By contrast, the roles of RNA in the cell do not require a second strand, and indeed the presence of a second, complementary, strand would preclude its role in protein synthesis. However, there are some viruses that have double-stranded RNA (dsRNA) as their genetic material, as well as some viruses with single-stranded RNA. In addition, some viruses (as well as some plasmids) replicate via single-stranded DNA forms. Double-stranded RNA is also important in the phenomenon known as RNA interference, which we will come to in later chapters.

1.2.4 Nucleic acid synthesis

We do not need to consider here all the details of how nucleic acids are synthesized. The fundamental features that we need to remember are summarized in Figure 1.7, which shows the addition of a nucleotide to the growing end (3′-OH) of a DNA strand. The substrate for this reaction is the relevant deoxynucleotide triphosphate (dNTP), i.e., the one that makes the correct base-pair with the corresponding residue on the template strand. The DNA strand is always extended at the 3′-OH end, thus the nucleotide strand grows in the 5′ to 3′ direction. For this reaction to occur it is essential that the existing residue at the 3′-OH end, to which the new nucleotide is to be added, is accurately base-paired with its partner on the other strand.

Figure 1.7 DNA synthesis.


3′ end
5′ end
OH
3′
O⁻
⁻O–P=O
O
5′ CH2
O
base----base
O
5′ CH2
O
⁻O–P=O
O
3′
3′
O
⁻O–P=O
O
5′ CH2
O
base----base
O
5′ CH2
O
⁻O–P=O
O
3′
3′
OH
Formation of phosphodiester bond
O⁻
⁻O–P=O
O
⁻O–P=O
O
⁻O–P=O
dNTP
base
O
5′ CH2
O
⁻O–P=O
O⁻
O
5′ CH2
O
base
3′
OH

RNA synthesis occurs in much the same way, as far as this simplistic description goes, except that of course the substrates are nucleotide triphosphates (NTPs) rather than the deoxynucleotide triphosphates (dNTPs). There is one very important difference though. DNA synthesis only occurs by extension of an existing strand – it always needs a *primer* to get it started. In contrast, RNA polymerases are capable of starting a new RNA strand, complementary to its template, from scratch, given the appropriate signals.

1.2.5 Coiling and supercoiling