



国外优秀生命科学教学用书

Principles of Gene Manipulation

基因操作原理

影印版

Sixth Edition

- Sandy Primrose
- Richard Twyman
- Bob Old



高等教育出版社
Higher Education Press



国外优秀生命科学教学用书

Principles of Gene Manipulation

基因操作原理

影印版

Sixth Edition

Sandy Primrose

Richard Twyman

Bob Old



高等教育出版社
Higher Education Press

图字: 01-2002-3317 号

Reprinted from

Principles of Gene Manipulation, Sixth Edition

Sandy Primrose, Richard Twyman and Bob Old

© 2001 by Blackwell Science Ltd

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 copyright owner.

This edition is published by arrangement with **Blackwell Publishing**, Oxford.

图书在版编目(CIP)数据

基因操作原理=Principles of Gene Manipulation:

第6版/(英)普里默罗斯(Primrose,S.), (英)特威曼(Twyman,R.),
(英)欧德(Old,B.)著. —影印本. —北京: 高等教育出版社, 2002.11

ISBN7-04-011468-2

I. 基... II. ①普... ②特... ③欧... III. 基因—
遗传工程—英文 IV. Q78

中国版本图书馆 CIP 数据核字 (2002) 第 087053 号

策 划 王 莉 封面设计 王凌波 责任印制 韩 刚

出版发行 高等教育出版社
社 址 北京市东城区沙滩后街 55 号
邮政编码 100009
传 真 010-64014048

购书热线 010-64054588
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>

经 销 新华书店北京发行所
印 刷 高等教育出版社印刷厂

开 本 787×1092 1/16
印 张 25
字 数 600 000

版 次 2002 年 11 月第 1 版
印 次 2002 年 11 月第 1 次印刷
定 价 38.60 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换。

版权所有 侵权必究

出版前言

随着克隆羊的问世和人类基因组计划的完成,生命科学成为 21 世纪名副其实的领头学科,生物高新技术产业逐步成为高科技产业的核心。生物技术和生物产业的发展对世界科技、经济、政治和社会发展等方面产生着深刻的影响,这也是我国赶超世界发达国家生产力水平最有前途和希望的领域。生命科学与技术全方位的发展呼唤高等教育培养更多高水平的复合型科技人才。

为此,教育部在《关于加强高等学校本科教学工作 提高教学质量的若干意见》[教高(2001)4 号文件]中提出,高等学校要大力提倡编写、引进和使用先进教材,其中信息科学、生命科学等发展迅速、国际通用性强、可比性强的学科和专业可以直接引进先进的、能反映学科发展前沿的原版教材。教育部高等教育司还于 2001 年 11 月向全国主要大学和出版社下发了“关于开展‘国外生命科学类优秀教学用书’推荐工作的通知”,有力推动了生命科学类教材的引进工作。

高等教育出版社对国外生命科学教材进行了充分的调研,并委托教育部高等学校生物科学与工程教学指导委员会的专家教授开展了“引进国外优秀生命科学教材及其教学辅助材料专项研究”,并就国内外同类教材进行了比较,提出了具体的引进教材书目。经过版权谈判,目前我社已经购买了 Pearson Education, McGraw-Hill, John Wiley & Sons, Blackwell Science, Thomson Learning, Cambridge University Press, Lippincott Williams & Wilkins 等出版的 13 种教材的影印权,学科领域涉及生物化学、细胞生物学、遗传学、微生物学、生态学、免疫学、神经科学、发育生物学、解剖学与生理学、分子生物学、普通生物学等。这些教材具有以下特点:(1)所选教材基本是近 2 年出版的,及时反映了学科发展的最新进展,在国际上使用广泛,具有权威性和时代感;(2)内容简明,篇幅适中,结构合理,兼具一定的深度和广度,适用范围广;(3)插图精美、丰富,既有很强的艺术性,又不失严谨的科学性,图文并茂,与正文相辅相成;(4)语言简练、流畅,十分适合非英语国家的学生阅读。其中 9 种已入选教育部高等教育司推荐“国外优秀生命科学教学用书”。

考虑到中国国情,为了让学生买得起,同时又能让学生看到原版书彩色精美的插图,我们在引进学生用原版教材时,一方面采用黑白影印,最大限度地降低定价,另一方面随书附赠含有原书彩色插图的光盘,以充分体现原教材的风格、特色,为读者提供方便。

引进国外优秀生命科学教学用书是我社一项长期的重点工作,因此,我们衷心希望广大专家教授和同学提出宝贵的意见和建议,如有更好的教材值得引进,请与高等教育出版社生命科学分社联系,联系电话:010-68344002, E-mail 地址: lifesciences-hep@x263.net。

高等教育出版社
2002 年 11 月

国外优秀生命科学教学用书 (影印教材)

<i>Biochemistry</i> (2nd ed.)	生物化学
<i>Cell and Molecular Biology</i> (3rd ed.)	分子细胞生物学
<i>Essentials of Genetics</i> (4th ed.)	遗传学基础
<i>Microbiology</i> (5th ed.)	微生物学
<i>Ecology: concepts and applications</i> (2nd ed.)	生态学
<i>Roitt's Essential Immunology</i> (10th ed.)	Roitt 免疫学基础
<i>Neuroscience: Exploring the Brain</i> (2nd ed.)	神经科学
<i>Essential Developmental Biology</i>	发育生物学基础
<i>Understanding Human Anatomy and Physiology</i> (4th ed.)	人体解剖生理学
<i>Gene Cloning and DNA Analysis</i> (4th ed.)	基因克隆和 DNA 分析
<i>Principles of Gene Manipulation</i> (6th ed.)	基因操作原理
<i>An Introduction to Genetic Engineering</i> (2nd ed.)	遗传工程导论
<i>Essential Biology</i>	生物学导论



Preface

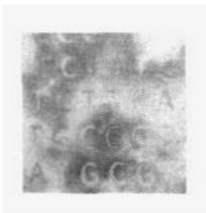
It is 22 years since the first edition of *Principles of Gene Manipulation* was published. In writing the first edition, our aim was to explain a new and rapidly growing technology. Our basic philosophy was to present the principles of gene manipulation, and its associated techniques, in sufficient detail to enable the non-specialist reader to understand them. In this new edition we have stuck to our original philosophy which we feel has been justified with the passage of time. Although originally intended as an advanced undergraduate textbook we know that many experienced researchers also used the previous editions as an informative reference source. In preparing this edition we have been conscious of the need to serve these two types of reader but herein lies a problem: the needs of experienced researchers are now vastly different from those of the undergraduate or early postgraduate student whereas this was not true a generation ago. Deliberately, we have not changed the level at which the book is written nor the general style. Rather, our solution has been to present the basic methodologies in the first half of the book, which is devoted to cloning in *Escherichia coli*, and to present the advanced techniques in the second half. The reader who has read and understood the material in the first half, or already knows it, should have no difficulty in understanding any of the material in the second half of the book. In the final chapter we present an overview of some important applications of recombinant DNA technology in a format that should appeal to all readers.

In writing the first half of the book much thought was given as to the amount of the early work on gene manipulation to include. Although older readers may feel that some of the material selected is 'dated' we have included it because an understanding of it is assumed in many publications today and is required if older papers are to be comprehended. In the second half of the book the emphasis is on cloning in eukaryotes, particularly in higher eukaryotes, and reflects current scientific focus. What is not included is methodology for analysing genomes since this is outwith the scope of 'gene manipulation'. However, this topic is covered in the third edition of our companion book *Principles of Genome Analysis and Genomics*.

We would like to thank our friends and colleagues for reading and commenting on various chapters prepared for this new edition. In particular we would like to thank Dylan Sweetman, Jane Pritchard, James Drummond, Roz Drummond, Phil Gardner, Gavin Craig, Ajay Kohli, Eva Stoger, Mark Leech, Victoria James, Paul Christou, Bruce Whitelaw, and Sue Goddard and her staff at CAMR for providing excellent library assistance.

Richard Twyman would like to dedicate this book to his parents, Peter and Irene, and his children, Emily and Lucy.

S.B. Primrose
R.M. Twyman
R.W. Old



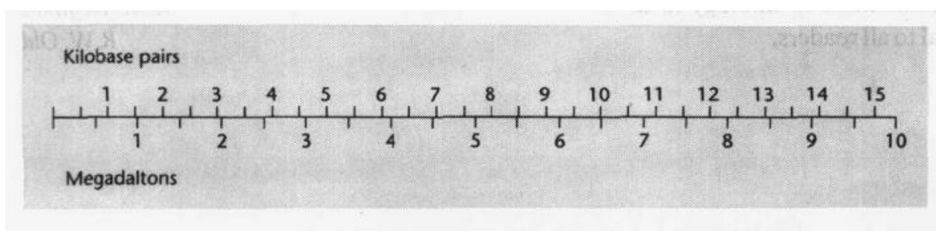
The genetic code and single-letter amino acid designations

5'-OH terminal base	Middle base				3'-OH terminal base
	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val*	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

*Codes for Met if in the initiator position.

Alanine	A	Leucine	L
Arginine	R	Lysine	K
Asparagine	N	Methionine	M
Aspartic acid	D	Phenylalanine	F
Cysteine	C	Proline	P
Glycine	G	Serine	S
Glutamic acid	E	Threonine	T
Glutamine	Q	Tryptophan	W
Histidine	H	Tyrosine	Y
Isoleucine	I	Valine	V

Scale for conversion between kilobase pairs of duplex DNA and molecular weight





Contents

Preface vii

The genetic code and single-letter amino acid designations viii

Conversion scale viii

1 Gene manipulation: an all-embracing technique, 1

Introduction, 1

Sequence analysis, 1

In vivo biochemistry, 2

The new medicine, 3

Biotechnology: the new industry, 4

The central role of *E. coli*, 5

Outline of the rest of the book, 6

2 Basic techniques, 8

Introduction, 8

The basic problems, 8

The solutions: basic techniques, 8

Agarose gel electrophoresis, 9

Nucleic acid blotting, 11

Transformation of *E. coli*, 17

The polymerase chain reaction (PCR), 19

3 Cutting and joining DNA molecules, 26

Cutting DNA molecules, 26

Joining DNA molecules, 36

4 Basic biology of plasmid and phage vectors, 43

Plasmid biology and simple plasmid vectors, 43

The purification of plasmid DNA, 48

Desirable properties of plasmid cloning vehicles, 49

Bacteriophage λ , 53

DNA cloning with single-stranded DNA vectors, 60

5 Cosmids, phasmids and other advanced vectors, 64

Introduction, 64

Vectors for cloning large fragments of DNA, 64

Specialist-purpose vectors, 70

Putting it all together: vectors with combinations of features, 84

6 Cloning strategies, 85

Introduction, 85

Cloning genomic DNA, 86

cDNA cloning, 92

Screening strategies, 101

Difference cloning, 114

7 Sequencing and mutagenesis, 120

Introduction, 120

Basic DNA sequencing, 120

Whole-genome sequencing, 126

Analysing sequence data, 126

Changing genes: site-directed mutagenesis, 132

8 Cloning in bacteria other than *Escherichia coli*, 139

Introduction, 139

Introducing DNA into bacterial cells, 139

- Cloning in Gram-negative bacteria other than *E. coli*, 144
Cloning in Gram-positive bacteria, 148
Cloning in streptomycetes, 153
Homoeologous recombination, 155
- 9 Cloning in *Saccharomyces cerevisiae* and other fungi, 156**
Introduction, 156
Introducing DNA into fungi, 156
The fate of DNA introduced into fungi, 156
Plasmid vectors for use in fungi, 158
Retrovirus-like vectors, 159
Expression of cloned genes, 163
Overexpression of proteins in fungi, 165
Specialist vectors, 166
Yeast surface display, 168
Identifying genes encoding particular cellular activities, 171
Determining functions associated with particular genes, 171
- 10 Gene transfer to animal cells, 174**
Introduction, 174
DNA-mediated transformation, 174
Gene transfer by viral transduction, 187
Summary of expression systems for animal cells, 199
- 11 Genetic manipulation of animals, 202**
Introduction, 202
Genetic manipulation of mammals, 203
DNA transfer to other vertebrates, 215
DNA transfer to invertebrates, 218
- 12 Gene transfer to plants, 221**
Introduction, 221
Agrobacterium-mediated transformation, 224
Direct DNA transfer to plants, 237
In planta transformation, 239
Chloroplast transformation, 240
Plant viruses as vectors, 241
- 13 Advances in transgenic technology, 247**
Introduction, 247
Inducible expression systems, 247
Applications of site-specific recombination, 253
Further transgenic strategies for gene inhibition, 260
Transgenic technology for functional genomics, 266
- 14 Applications of recombinant DNA technology, 274**
Introduction, 274
Theme 1: Nucleic acid sequences as diagnostic tools, 274
Theme 2: New drugs and new therapies for genetic diseases, 283
Theme 3: Combating infectious disease, 293
Theme 4: Protein engineering, 299
Theme 5: Metabolic engineering, 303
Theme 6: Plant breeding in the twenty-first century, 311
Epilogue: from genes to genomes, 319
- References, 320
- Index, 377



CHAPTER 1

Gene manipulation: an all-embracing technique

Introduction

Occasionally technical developments in science occur that enable leaps forward in our knowledge and increase the potential for innovation. Molecular biology and biomedical research experienced such a revolutionary change in the mid-70s with the development of gene manipulation. Although the initial experiments generated much excitement, it is unlikely that any of the early workers in the field could have predicted the breadth of applications to which the technique has been put. Nor could they have envisaged that the methods they developed would spawn an entire industry comprising several hundred companies, of varying sizes, in the USA alone.

The term gene manipulation can be applied to a variety of sophisticated *in vivo* genetics as well as to *in vitro* techniques. In fact, in most Western countries there is a precise *legal* definition of gene manipulation as a result of government legislation to control it. In the UK, gene manipulation is defined as

the formation of new combinations of heritable material by the insertion of nucleic acid molecules, produced by whatever means outside the cell, into any virus, bacterial plasmid or other vector system so as to allow their incorporation into a host organism in which they do not naturally occur but in which they are capable of continued propagation.

The definitions adopted by other countries are similar and all adequately describe the subject-matter of this book. Simply put, gene manipulation permits stretches of DNA to be isolated from their host organism and propagated in the same or a different host, a technique known as *cloning*. The ability to clone DNA has far-reaching consequences, as will be shown below.

Sequence analysis

Cloning permits the isolation of discrete pieces of a genome and their amplification. This in turn enables the DNA to be sequenced. Analysis of the sequences of some genetically well-characterized genes led to the identification of the sequences and structures which characterize the principal control elements of gene expression, e.g. promoters, ribosome binding sites, etc. As this information built up it became possible to scan new DNA sequences and identify potential new genes, or *open reading frames*, because they were bounded by characteristic motifs. Initially this sequence analysis was done manually but to the eye long runs of nucleotides have little meaning and patterns evade recognition. Fortunately such analyses have been facilitated by rapid increases in the power of computers and improvements in software which have taken place contemporaneously with advances in gene cloning. Now sequences can be scanned quickly for a whole series of structural features, e.g. restriction enzyme recognition sites, start and stop signals for transcription, inverted palindromes, sequence repeats, Z-DNA, etc., using programs available on the Internet.

From the nucleotide sequence of a gene it is easy to deduce the protein sequence which it encodes. Unfortunately, we are unable to formulate a set of general rules that allows us to predict a protein's three-dimensional structure from the amino acid sequence of its polypeptide chain. However, based on crystallographic data from over 300 proteins, certain structural motifs can be predicted. Nor does an amino acid sequence on its own give any clue to function. The solution is to compare the amino acid sequence with that of other better-characterized proteins: a high degree of homology suggests similarity in function. Again, computers are of great value since algorithms exist for comparing two sequences or for comparing one sequence with a group of other

sequences simultaneously. The Internet has made such comparisons easy because researchers can access all the protein sequence data that are stored in central databases, which are updated daily.

***In vivo* biochemistry**

Any living cell, regardless of its origin, carries out a plethora of biochemical reactions. To analyse these different reactions, biochemists break open cells, isolate the key components of interest and measure their levels. They purify these components and try to determine their performance characteristics. For example, in the case of an enzyme, they might determine its substrate specificity and kinetic parameters, such as K_m and V_{max} , and identify inhibitors and their mode of action. From these data they try to build up a picture of what happens inside the cell. However, the properties of a purified enzyme in a test-tube may bear little resemblance to its behaviour when it shares the cell cytoplasm or a cell compartment with thousands of other enzymes and chemical compounds. Understanding what happens inside cells has been facilitated by the use of mutants. These permit the determination of the consequences of altered regulation or loss of a particular component or activity. Mutants have also been useful in elucidating macromolecule structure and function. However, the use of mutants is limited by the fact that with classical technologies one usually has little control over the type of mutant isolated and/or location of the mutation.

Gene cloning provides elegant solutions to the above problems. Once isolated, entire genes or groups of genes can be introduced back into the cell type whence they came or into different cell types or completely new organisms, e.g. bacterial genes in plants or animals. The levels of gene expression can be measured directly or through the use of reporter molecules and can be modulated up or down at the whim of the experimenter. Also, specific mutations, ranging from a single base-pair to large deletions or additions, can be built into the gene at any position to permit all kinds of structural and functional analyses. Function in different cell types can also be analysed, e.g. do those structural features of a protein which result in its secretion from a yeast cell enable it to be exported from bacteria or higher eukaryotes? Experiments like these permit comparative studies of macromolecular processes and, in some cases, gene cloning and sequencing provides the only way to begin to understand such events as mitosis, cell division, telomere structure, intron splicing, etc. Again, the Internet has made such comparisons easy because researchers can access all the protein sequence data that are stored in central databases, which are updated daily.

The original goal of sequencing was to determine the precise order of nucleotides in a gene. Then the goal became the sequence of a small genome. First it was that of a small virus (ϕ X174, 5386 nucleotides). Then the goal was larger plasmid and viral genomes, then chromosomes and microbial genomes until ultimately the complete genomes of higher eukaryotes (humans, *Arabidopsis*) were sequenced (Table 1.1).

Table 1.1 Increases in sizes of genomes sequenced.

Genome sequenced	Year	Genome size	Comment
Bacteriophage ϕ X174	1977	5.38 kb	First genome sequenced
Plasmid pBR322	1979	4.3 kb	First plasmid sequenced
Bacteriophage λ	1982	48.5 kb	
Epstein-Barr virus	1984	172 kb	
Yeast chromosome III	1992	315 kb	First chromosome sequenced
<i>Haemophilus influenzae</i>	1995	1.8 Mb	First genome of cellular organism to be sequenced
<i>Saccharomyces cerevisiae</i>	1996	12 Mb	First eukaryotic genome to be sequenced
<i>Ceanorhabditis elegans</i>	1998	97 Mb	First genome of multicellular organism to be sequenced
<i>Drosophila melanogaster</i>	2000	165 Mb	
<i>Homo sapiens</i>	2000	3000 Mb	First mammalian genome to be sequenced
<i>Arabidopsis thaliana</i>	2000	125 Mb	First plant genome to be sequenced

Now the sequencing of large genomes has become routine, albeit in specialist laboratories. Having the complete genome sequence of an organism provides us with fascinating insights into certain aspects of its biology. For example, we can determine the metabolic capabilities of a new microbe without knowing anything about its physiology. However, there are many aspects of cellular biology that cannot be ascertained from sequence data alone. For example, what RNA species are made when in the cell or organism life cycle and how fast do they turn over? What proteins are made when and how do the different proteins in a cell interact? How does environment affect gene expression? The answers to these questions are being provided by the new disciplines of genomics, proteomics and environomics which rely heavily on the *techniques* of gene manipulation, which are discussed in later chapters. A detailed presentation of whole-genome sequencing, genomics and proteomics can be found in Primrose and Twyman (2002).

The new medicine

The developments in gene manipulation that have taken place in the last 25 years have revolutionized the study of biology. There is no subject area within biology where recombinant DNA is not being used and as a result the old divisions between subject areas such as botany, genetics, zoology, biochemistry, etc. are fast breaking down. Nowhere has the impact of

recombinant DNA technology been greater than on the practice of medicine.

The first medical benefit to arise from recombinant DNA technology was the availability of significant quantities of therapeutic proteins, such as human growth hormone (HGH). This protein is used to treat adolescents suffering from pituitary dwarfism to enable them to achieve a normal height. Originally HGH was purified from pituitary glands removed from cadavers. However, a very large number of pituitary glands are required to produce sufficient HGH to treat just one child. Furthermore, some children treated with pituitary-derived HGH have developed Creutzfeld–Jakob syndrome. Following the cloning and expression of the HGH gene in *Escherichia coli*, it is possible to produce enough HGH in a 10 litre fermenter to treat hundreds of children. Since then, many different therapeutic proteins have become available for the first time. Many of these proteins are also manufactured in *E. coli* but others are made in yeast or animal cells and some in plants or the milk of animals. The only common factor is that the relevant gene has been cloned and overexpressed using the techniques of gene manipulation.

Medicine has benefited from recombinant DNA technology in other ways (Fig. 1.1). New routes to vaccines have been developed. The current hepatitis B vaccine is based on the expression of a viral antigen on the surface of yeast cells and a recombinant vaccine has been used to eliminate rabies from foxes in a large part of Europe. Gene manipulation can

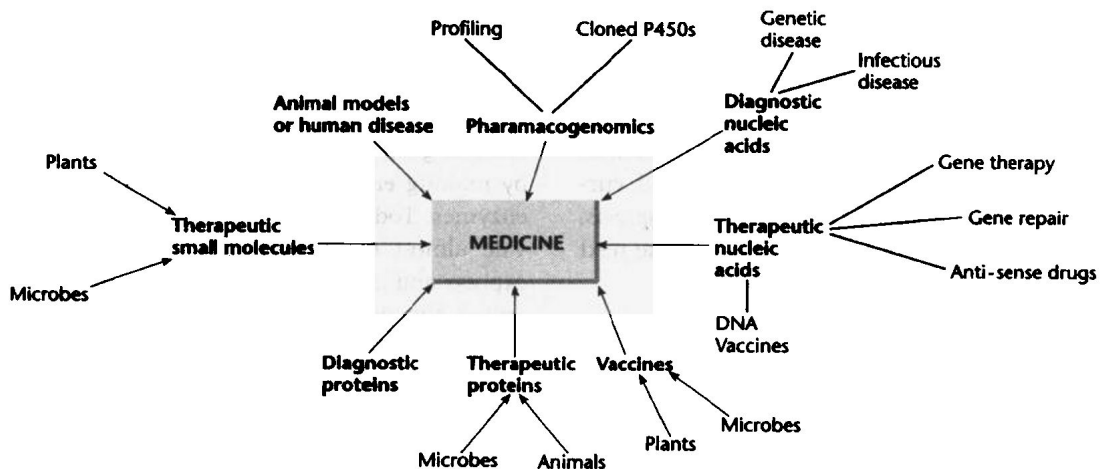


Fig. 1.1 The impact of gene manipulation on the practice of medicine.

also be used to increase the levels of small molecules within microbial cells. This can be done by cloning all the genes for a particular biosynthetic pathway and overexpressing them. Alternatively, it is possible to shut down particular metabolic pathways and thus redirect particular intermediates towards the desired end-product. This approach has been used to facilitate production of chiral intermediates and antibiotics. Novel antibiotics can also be created by mixing and matching genes from organisms producing different but related molecules in a technique known as combinatorial biosynthesis.

Gene cloning enables nucleic acid probes to be produced readily and such probes have many uses in medicine. For example, they can be used to determine or confirm the identity of a microbial pathogen or to diagnose pre- or perinatally an inherited genetic disease. Increasingly, probes are being used to determine the likelihood of adverse reactions to drugs or to select the best class of drug to treat a particular illness (pharmacogenomics). A variant of this technique is to use cloned cytochrome P450s to determine how a new drug will be metabolized and if any potentially toxic by-products will result.

Nucleic acids are also being used as therapeutic entities in their own right. For example, antisense nucleic acids are being used to down-regulate gene expression in certain diseases. In other cases, nucleic acids are being administered to correct or repair inherited gene defects (gene therapy/gene repair) or as vaccines. In the reverse of gene repair, animals are being generated that have mutations identical to those found in human disease. Note that the use of antisense nucleic acids and gene therapy/repair depends on the availability of information on the exact *cause* of a disease. For most medical conditions such information is lacking and currently available drugs are used to treat *symptoms*. This situation will change significantly in the next decade.

Biotechnology: the new industry

The early successes in overproducing mammalian proteins in *E. coli* suggested to a few entrepreneurial individuals that a new company should be formed to

exploit the potential of recombinant DNA technology. Thus was Genentech born (Box 1.1). Since then thousands of biotechnology companies have been formed worldwide. As soon as major new developments in the science of gene manipulation are reported, a rash of new companies are formed to commercialize the new technology. For example, many recently formed companies are hoping the data from the Human Genome Sequencing Project will result in the identification of a large number of new proteins with potential for human therapy. Others are using gene manipulation to understand the regulation of transcription of particular genes, arguing that it would make better therapeutic sense to modulate the process with low-molecular-weight, orally active drugs.

Although there are thousands of biotechnology companies, fewer than 100 have sales of their products and even fewer are profitable. Already many biotechnology companies have failed, but the technology advances at such a rate that there is no shortage of new company start-ups to take their place. One group of biotechnology companies that has prospered is those supplying specialist reagents to laboratory workers engaged in gene manipulation. In the very beginning, researchers had to make their own restriction enzymes and this restricted the technology to those with protein chemistry skills. Soon a number of companies were formed which catered to the needs of researchers by supplying high-quality enzymes for DNA manipulation. Despite the availability of these enzymes, many people had great difficulty in cloning DNA. The reason for this was the need for careful quality control of all the components used in the preparation of reagents, something researchers are not good at! The supply companies responded by making easy-to-use cloning kits in addition to enzymes. Today, these supply companies can provide almost everything that is needed to clone, express and analyse DNA and have thereby accelerated the use of recombinant DNA technology in all biological disciplines. In the early days of recombinant DNA technology, the development of methodology was an end in itself for many academic researchers. This is no longer true. The researchers have gone back to using the tools to further our

Box 1.1 The birth of an industry

Biotechnology is not new. Cheese, bread and yoghurt are products of biotechnology and have been known for centuries. However, the stock-market excitement about biotechnology stems from the potential of gene manipulation, which is the subject of this book. The birth of this modern version of biotechnology can be traced to the founding of the company Genentech.

In 1976, a 27-year-old venture capitalist called Robert Swanson had a discussion over a few beers with a University of California professor, Herb Boyer. The discussion centred on the commercial potential

of gene manipulation. Swanson's enthusiasm for the technology and his faith in it was contagious. By the close of the meeting the decision was taken to found Genentech (Genetic Engineering Technology). Though Swanson and Boyer faced scepticism from both the academic and business communities they forged ahead with their idea. Successes came thick and fast (see Table B1.1) and within a few years they had proved their detractors wrong. Over 1000 biotechnology companies have been set up in the USA alone since the founding of Genentech but very, very few have been as successful.

Table B1.1 Key events at Genentech.

1976	Genentech founded
1977	Genentech produced first human protein (somatostatin) in a microorganism
1978	Human insulin cloned by Genentech scientists
1979	Human growth hormone cloned by Genentech scientists
1980	Genentech went public, raising \$35 million
1982	First recombinant DNA drug (human insulin) marketed (Genentech product licensed to Eli Lilly & Co.)
1984	First laboratory production of factor VIII for therapy of haemophilia. Licence granted to Cutter Biological
1985	Genentech launched its first product, Protropin (human growth hormone), for growth hormone deficiency in children
1987	Genentech launched Activase (tissue plasminogen activator) for dissolving blood clots in heart-attack patients
1990	Genentech launched Actimmune (interferon- $\gamma_{1\beta}$) for treatment of chronic granulomatous disease
1990	Genentech and the Swiss pharmaceutical company Roche complete a \$2.1 billion merger

knowledge of biology, and the development of new methodologies has largely fallen to the supply companies.

The central role of *E. coli*

E. coli has always been a popular model system for molecular geneticists. Prior to the development of recombinant DNA technology, there existed a large number of well-characterized mutants, gene regulation was understood and there was a ready availability of a wide selection of plasmids. Compared with other microbial systems it was matchless. It is not surprising, therefore, that the first

cloning experiments were undertaken in *E. coli*. Subsequently, cloning techniques were extended to a range of other microorganisms, such as *Bacillus subtilis*, *Pseudomonas* sp., yeasts and filamentous fungi, and then to higher eukaryotes. Curiously, cloning in *E. coli* is technically easier than in any other organism. As a result, it is rare for researchers to clone DNA directly in other organisms. Rather, DNA from the organism of choice is first manipulated in *E. coli* and subsequently transferred back to the original host. Without the ability to clone and manipulate DNA in *E. coli*, the application of recombinant DNA technology to other organisms would be greatly hindered.

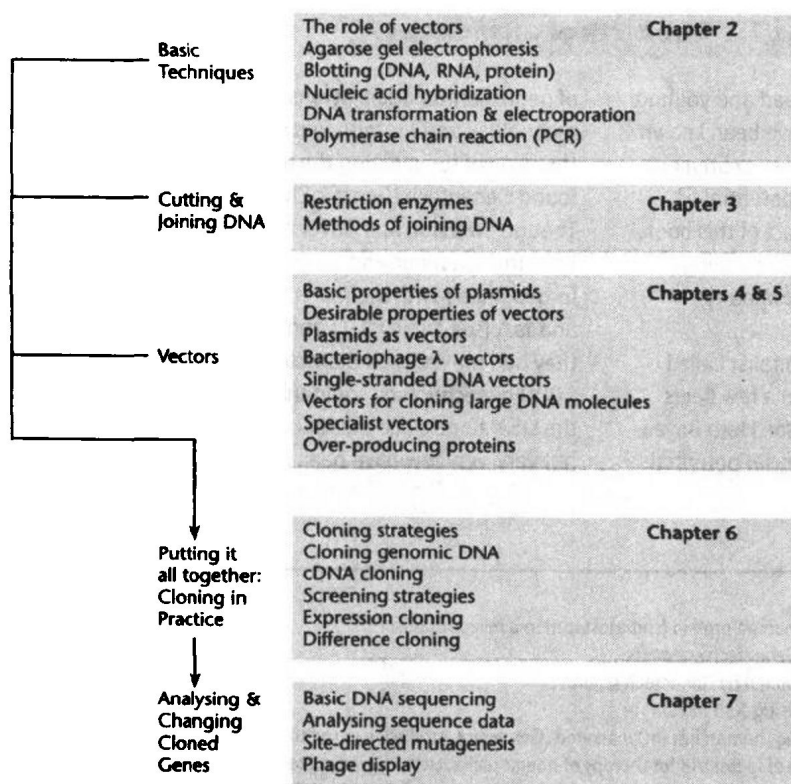


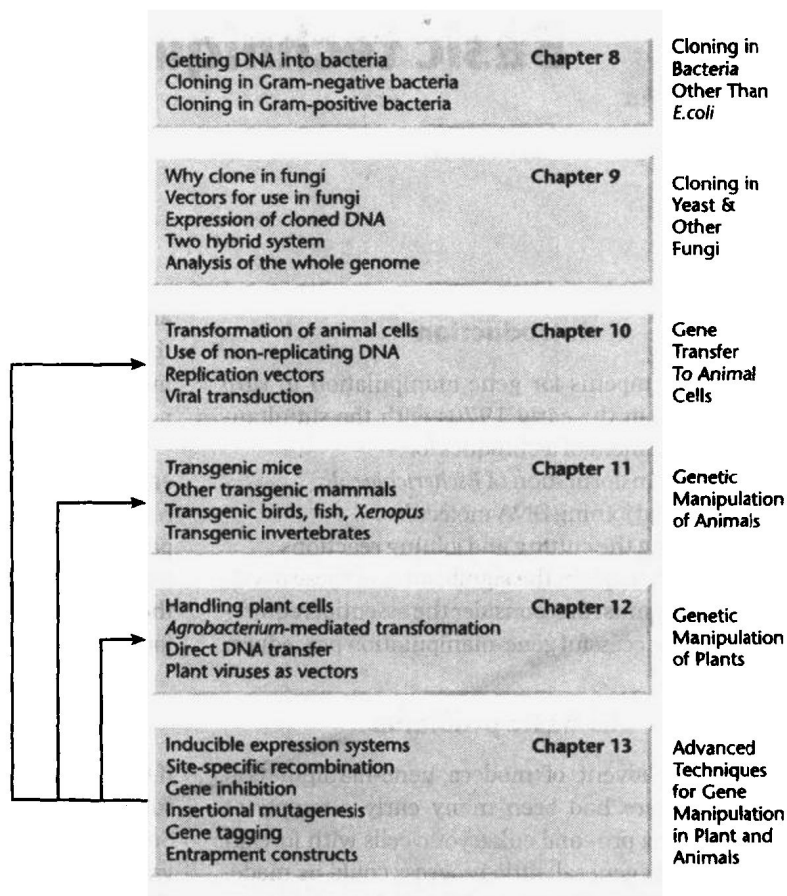
Fig. 1.2 'Roadmap' outlining the basic techniques in gene manipulation and their relationships.

Outline of the rest of the book

As noted above, *E. coli* has an essential role in recombinant DNA technology. Therefore, the first half of the book is devoted to the methodology for manipulating genes in this organism (Fig. 1.2). Chapter 2 covers many of the techniques that are common to all cloning experiments and are fundamental to the success of the technology. Chapter 3 is devoted to methods for selectively cutting DNA molecules into fragments that can be readily joined together again. Without the ability to do this, there would be no recombinant DNA technology. If fragments of DNA are inserted into cells, they fail to replicate except in those rare cases where they integrate into the chromosome. To enable such fragments to be propagated, they are inserted into DNA molecules (vectors) that are capable of extrachromosomal replication. These vectors are derived from plasmids and bacteriophages and their basic properties are described in Chapter 4. Originally, the purpose of

vectors was the propagation of cloned DNA but today vectors fulfil many other roles, such as facilitating DNA sequencing, promoting expression of cloned genes, facilitating purification of cloned gene products, etc. The specialist vectors for these tasks are described in Chapter 5. With this background in place it is possible to describe in detail how to clone the particular DNA sequences that one wants. There are two basic strategies. Either one clones all the DNA from an organism and then selects the very small number of clones of interest or one amplifies the DNA sequences of interest and then clones these. Both these strategies are described in Chapter 6. Once the DNA of interest has been cloned, it can be sequenced and this will yield information on the proteins that are encoded and any regulatory signals that are present. There might also be a wish to modify the DNA and/or protein sequence and determine the biological effects of such changes. The techniques for sequencing and changing cloned genes are described in Chapter 7.

Fig. 1.3 'Roadmap' of the advanced techniques in gene manipulation and their application to organisms other than *E. coli*.



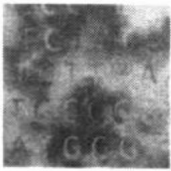
In the second half of the book the specialist techniques for cloning in organisms other than *E. coli* are described (Fig. 1.3). Each of these chapters can be read in isolation from the other chapters in this section, provided that there is a thorough understanding of the material from the first half of the book. Chapter 8 details the methods for cloning in other bacteria. Originally it was thought that some of these bacteria, e.g. *B. subtilis*, would usurp the position of *E. coli*. This has not happened and gene manipulation techniques are used simply to better understand the biology of these bacteria. Chapter 9 focuses on cloning in fungi, although the emphasis is on the yeast *Saccharomyces cerevisiae*. Fungi are eukaryotes and are useful model systems for investigating topics such as meiosis and mitosis, control of cell division, etc. Animal cells can be cultured like microorganisms and the techniques for cloning in them are described in Chapter 10. Chapters 11 and 12 are devoted to the intricacies of cloning in animal

and plant representatives of higher eukaryotes and Chapter 13 covers some cutting-edge techniques for these same systems.

The concluding chapter is a survey of the different applications of recombinant DNA technology that are being exploited by the biotechnology industry. Rather than going through application after application, we have opted to show the interplay of different technologies by focusing on six themes:

- Nucleic acid sequences as diagnostic tools.
- New drugs and new therapies for genetic diseases.
- Combating infectious disease.
- Protein engineering.
- Metabolic engineering.
- Plant breeding in the twenty-first century.

By treating the topic in this way we have been able to show the interplay between some of the basic techniques and the sophisticated analysis now possible with genome sequence information.



CHAPTER 2

Basic techniques

Introduction

The initial impetus for gene manipulation *in vitro* came about in the early 1970s with the simultaneous development of techniques for:

- genetic transformation of *Escherichia coli*;
- cutting and joining DNA molecules;
- monitoring the cutting and joining reactions.

In order to explain the significance of these developments we must first consider the essential requirements of a successful gene-manipulation procedure.

The basic problems

Before the advent of modern gene-manipulation methods there had been many early attempts at transforming pro- and eukaryotic cells with foreign DNA. But, in general, little progress could be made. The reasons for this are as follows. Let us assume that the exogenous DNA is taken up by the recipient cells. There are then two basic difficulties. First, where detection of uptake is dependent on gene expression, failure could be due to lack of accurate transcription or translation. Secondly, and more importantly, the exogenous DNA may not be maintained in the transformed cells. If the exogenous DNA is integrated into the host genome, there is no problem. The exact mechanism whereby this integration occurs is not clear and it is usually a rare event. However this occurs, the result is that the foreign DNA sequence becomes incorporated into the host cell's genetic material and will subsequently be propagated as part of that genome. If, however, the exogenous DNA fails to be integrated, it will probably be lost during subsequent multiplication of the host cells. The reason for this is simple. In order to be replicated, DNA molecules must contain an *origin of replication*, and in bacteria and viruses there is usually only one per genome. Such molecules are called *replicons*. Fragments of DNA are not replicons

and in the absence of replication will be diluted out of their host cells. It should be noted that, even if a DNA molecule contains an origin of replication, this may not function in a foreign host cell.

There is an additional, subsequent problem. If the early experiments were to proceed, a method was required for assessing the fate of the donor DNA. In particular, in circumstances where the foreign DNA was maintained because it had become integrated in the host DNA, a method was required for mapping the foreign DNA and the surrounding host sequences.

The solutions: basic techniques

If fragments of DNA are not replicated, the obvious solution is to attach them to a suitable replicon. Such replicons are known as *vectors* or *cloning vehicles*. Small plasmids and bacteriophages are the most suitable vectors for they are replicons in their own right, their maintenance does not necessarily require integration into the host genome and their DNA can be readily isolated in an intact form. The different plasmids and phages which are used as vectors are described in detail in Chapters 4 and 5. Suffice it to say at this point that initially plasmids and phages suitable as vectors were only found in *E. coli*. An important consequence follows from the use of a vector to carry the foreign DNA: simple methods become available for purifying the vector molecule, complete with its foreign DNA insert, from transformed host cells. Thus not only does the vector provide the replicon function, but it also permits the easy bulk preparation of the foreign DNA sequence, free from host-cell DNA.

Composite molecules in which foreign DNA has been inserted into a vector molecule are sometimes called DNA *chimeras* because of their analogy with the Chimaera of mythology – a creature with the head of a lion, body of a goat and tail of a serpent. The construction of such composite or *artificial*