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 (ϕX174, 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 ϕX174	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
Haemophilus influenzae	1995	1.8 Mb	First genome of cellular organism to be sequenced
Saccharomyces cerevisiae	1996	12 Mb	First eukaryotic genome to be sequenced
Ceanorhabditis elegans	1998	97 Mb	First genome of multicellular organism to be sequenced
Drosophila melanogaster	2000	165 Mb	
Homo sapiens	2000	3000 Mb	First mammalian genome to be sequenced
Arabidopsis thaliana	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

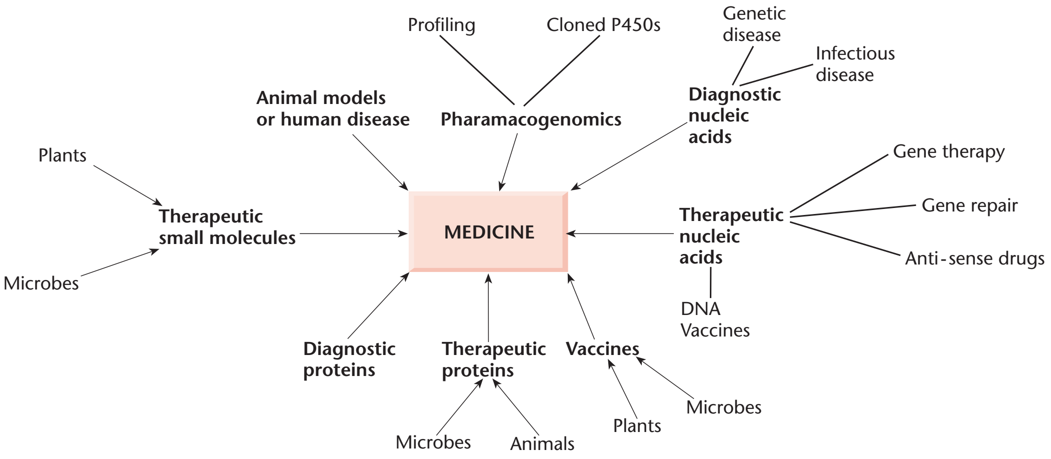

Fig. 1.1 The impact of gene manipulation on the practice of medicine.

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.

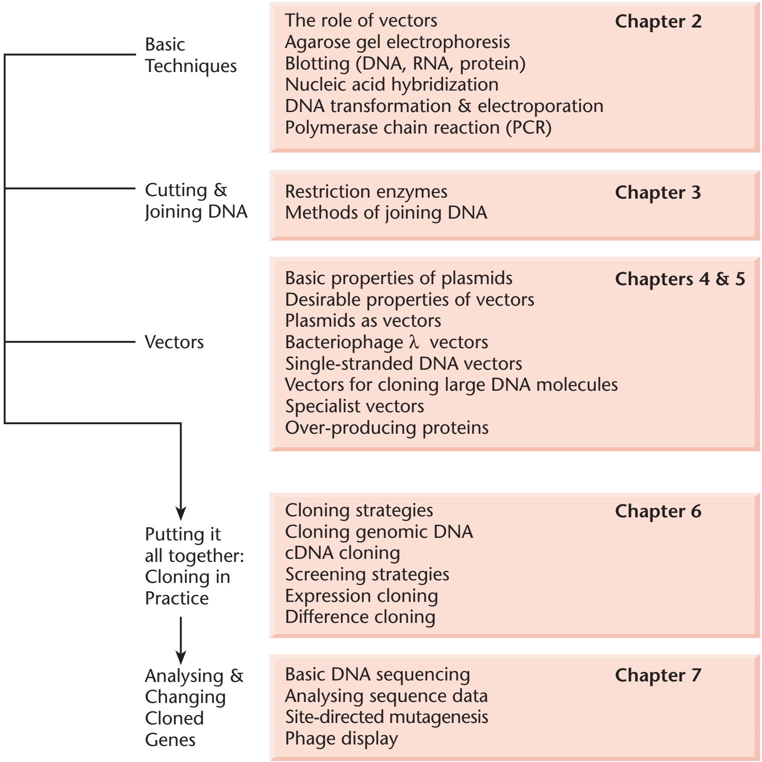

Fig. 1.2 'Roadmap' outlining the basic techniques in gene manipulation and their relationships.

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.

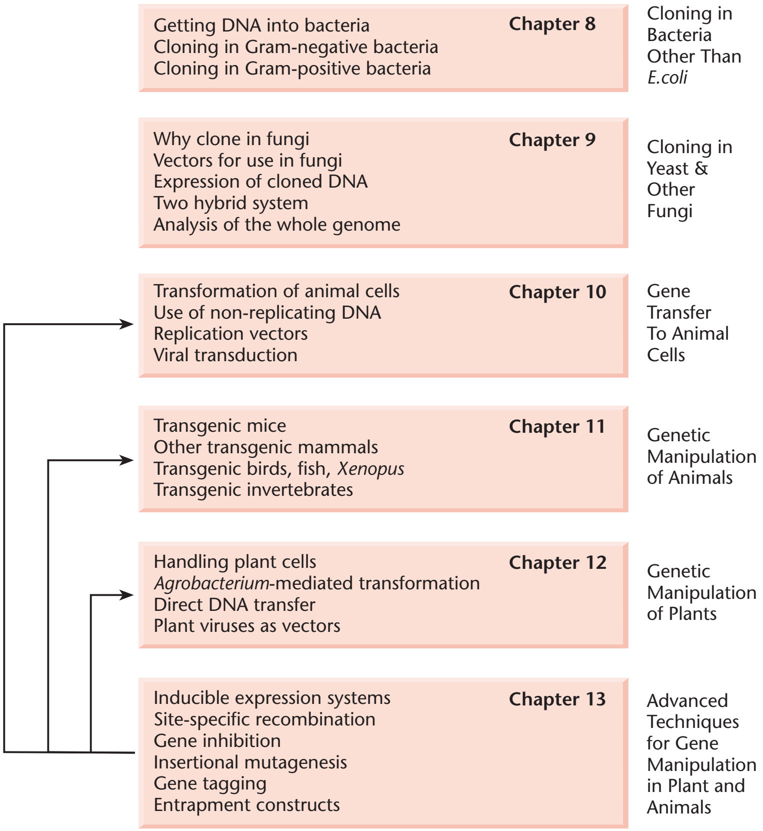
Getting DNA into bacteria
Cloning in Gram-negative bacteria
Cloning in Gram-positive bacteria
Chapter 8
Cloning in Bacteria Other Than *E.coli*
Why clone in fungi
Vectors for use in fungi
Expression of cloned DNA
Two hybrid system
Analysis of the whole genome
Chapter 9
Cloning in Yeast & Other Fungi
Transformation of animal cells
Use of non-replicating DNA
Replication vectors
Viral transduction
Chapter 10
Gene Transfer To Animal Cells
Transgenic mice
Other transgenic mammals
Transgenic birds, fish, *Xenopus*
Transgenic invertebrates
Chapter 11
Genetic Manipulation of Animals
Handling plant cells
Agrobacterium-mediated transformation
Direct DNA transfer
Plant viruses as vectors
Chapter 12
Genetic Manipulation of Plants
Inducible expression systems
Site-specific recombination
Gene inhibition
Insertional mutagenesis
Gene tagging
Entrapment constructs
Chapter 13
Advanced Techniques for Gene Manipulation in Plant and Animals

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.