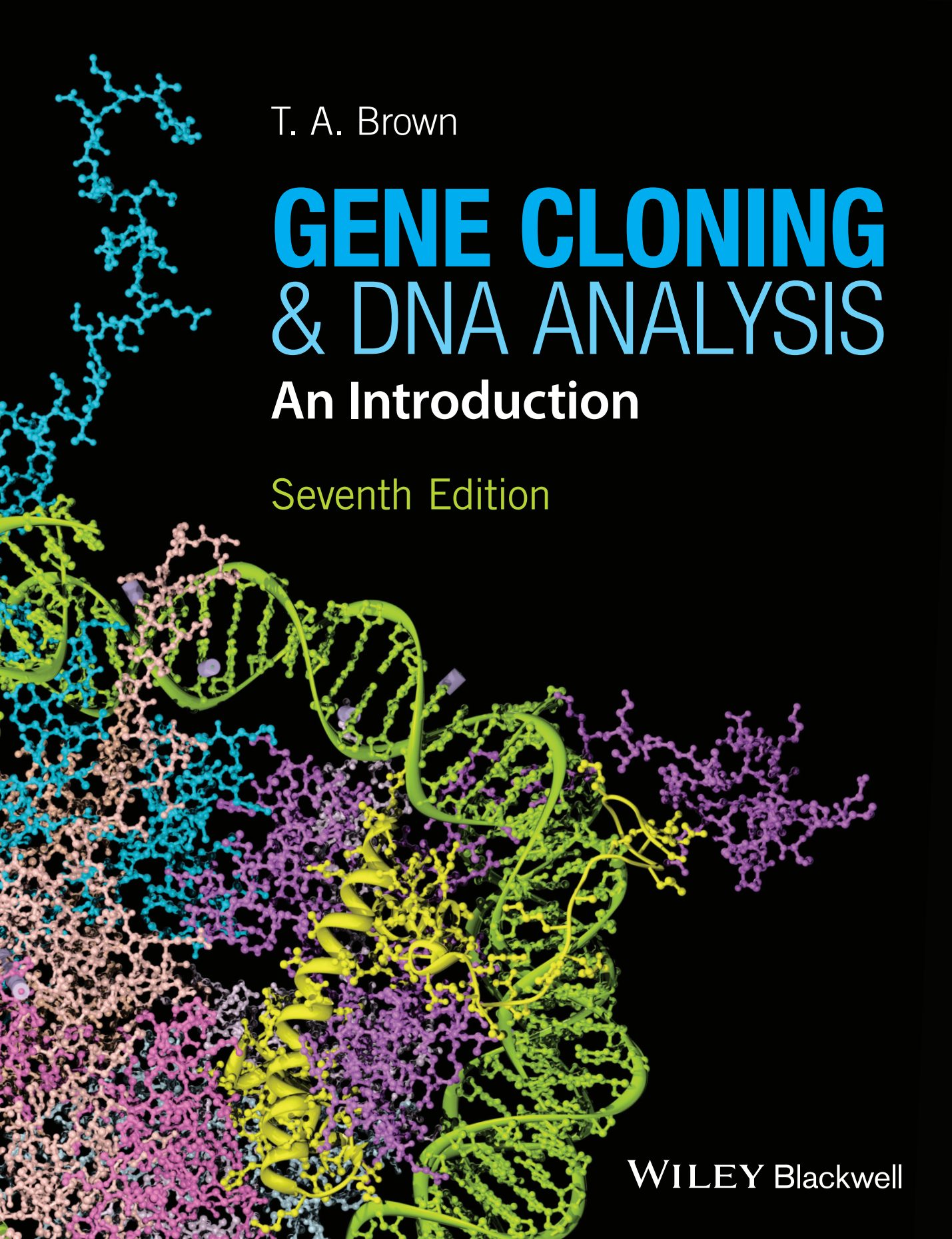

T. A. Brown
GENE CLONING
& DNA ANALYSIS
An Introduction
Seventh Edition
WILEY Blackwell

Contents in Brief

CONTENTS IN BRIEF

PART I

The Basic Principles of Gene Cloning and DNA Analysis

Chapter 1

Why Gene Cloning and DNA Analysis are Important

Chapter contents

CHAPTER CONTENTS

In the middle of the 19th century, Gregor Mendel formulated a set of rules to explain the inheritance of biological characteristics. The basic assumption of these rules is that each heritable property of an organism is controlled by a factor, called a **gene**, that is a physical particle present somewhere in the cell. The rediscovery of Mendel's laws in 1900 marks the birth of **genetics**, the science aimed at understanding what these genes are and exactly how they work.

1.1 The early development of genetics

For the first 30 years of its life this new science grew at an astonishing rate. The idea that genes reside on **chromosomes** was proposed by W. Sutton in 1903, and received experimental backing from T.H. Morgan in 1910. Morgan and his colleagues then developed the techniques for **gene mapping**, and by 1922 had produced a comprehensive analysis of the relative positions of over 2000 genes on the four chromosomes of the fruit fly, *Drosophila melanogaster*.

Despite the brilliance of these classical genetic studies, there was no real understanding of the molecular nature of the gene until the 1940s. Indeed, it was not until the

Gene Cloning and DNA Analysis: An Introduction, Seventh Edition. T.A. Brown.

Companion Website: www.wiley.com/go/brown/cloning

experiments of Avery, MacLeod, and McCarty in 1944, and of Hershey and Chase in 1952, that anyone believed that deoxyribonucleic acid (DNA) is the genetic material. Up until then it was widely thought that genes were made of protein. The discovery of the role of DNA was a tremendous stimulus to genetic research, and many famous biologists (Delbrück, Chargaff, Crick, and Monod were among the most influential) contributed to the second great age of genetics. During the 14 years between 1952 and 1966, the structure of DNA was elucidated, the genetic code cracked, and the processes of transcription and translation described.

1.2 The advent of gene cloning and the polymerase chain reaction

These years of activity and discovery were followed by a lull, a period of anticlimax when it seemed to some molecular biologists (as the new generation of geneticists styled themselves) that there was little of fundamental importance that was not understood. In truth, there was a frustration that the experimental techniques of the late 1960s were not sophisticated enough to allow the gene to be studied in any greater detail.

Then, in the years 1971–1973 genetic research was thrown back into gear by what at the time was described as a revolution in experimental biology. A whole new methodology was developed, enabling previously impossible experiments to be planned and carried out, if not with ease, then at least with success. These methods, referred to as **recombinant DNA technology** or **genetic engineering**, and having at their core the process of **gene cloning**, sparked another great age of genetics. They led to rapid and efficient **DNA sequencing** techniques that enabled the structures of individual genes to be determined, reaching a culmination at the turn of the century with the massive genome sequencing projects, including the human project which was completed in 2000. They led to procedures for studying the regulation of individual genes, which have allowed molecular biologists to understand how aberrations in gene activity can result in human diseases such as cancer. The techniques spawned modern **biotechnology**, which puts genes to work in production of proteins and other compounds needed in medicine and industrial processes.

During the 1980s, when the excitement engendered by the gene cloning revolution was at its height, it hardly seemed possible that another, equally novel and equally revolutionary process was just around the corner. According to DNA folklore, Kary Mullis invented the **polymerase chain reaction (PCR)** during a drive along the coast of California one evening in 1985. His brainwave was an exquisitely simple technique that acts as a perfect complement to gene cloning. PCR has made easier many of the techniques that were possible but difficult to carry out when gene cloning was used on its own. It has extended the range of DNA analysis and enabled molecular biology to find new applications in areas of endeavour outside of its traditional range of medicine, agriculture, and biotechnology. Archaeogenetics, molecular ecology, and DNA forensics are just three of the new disciplines that have become possible as a direct consequence of the invention of PCR, enabling molecular biologists to ask questions about human evolution and the impact of environmental change on the biosphere, and to bring their powerful tools to bear in the fight against crime. Forty years have passed since the dawning of the age of gene cloning, but we are still riding the rollercoaster and there is no end to the excitement in sight.

1.3 What is gene cloning?

What exactly is gene cloning? The easiest way to answer this question is to follow through the steps in a gene cloning experiment (Figure 1.1):

1. A fragment of DNA, containing the gene to be cloned, is inserted into a circular DNA molecule called a **vector**, to produce a **recombinant DNA molecule**.
2. The vector transports the gene into a host cell, which is usually a bacterium, although other types of living cell can be used.
3. Within the host cell the vector multiplies, producing numerous identical copies, not only of itself but also of the gene that it carries.
4. When the host cell divides, copies of the recombinant DNA molecule are passed to the progeny and further vector replication takes place.
5. After a large number of cell divisions, a colony, or **clone**, of identical host cells is produced. Each cell in the clone contains one or more copies of the recombinant DNA molecule. The gene carried by the recombinant molecule is now said to be cloned.

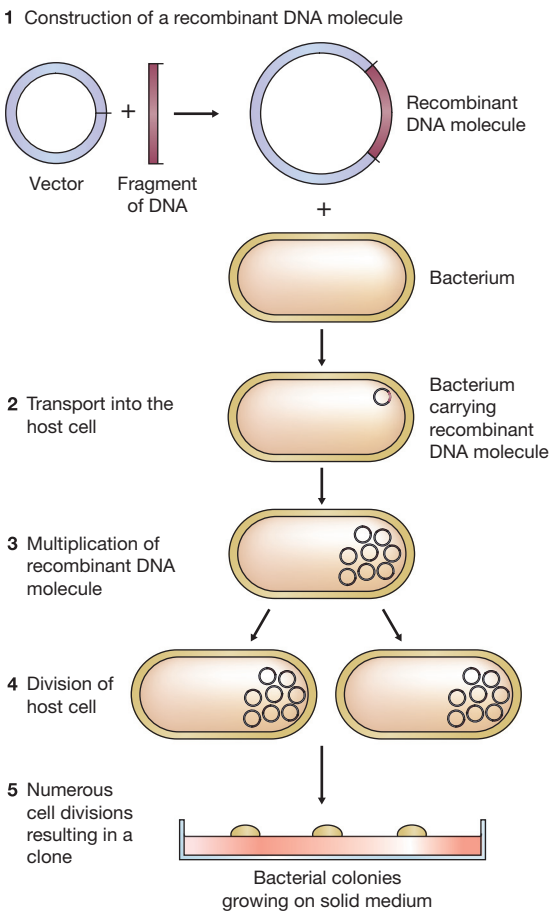


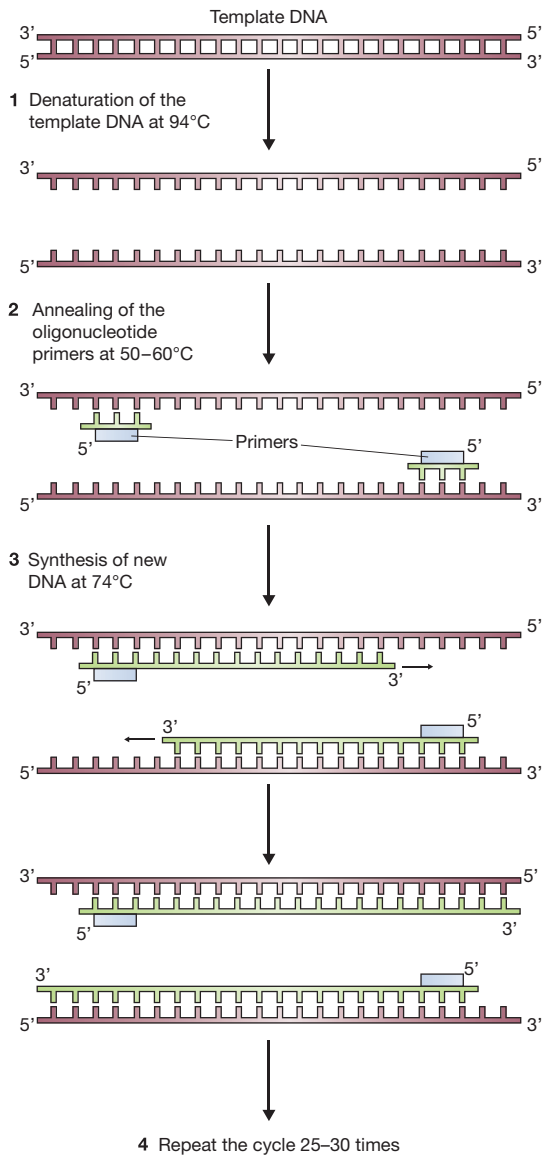
Figure 1.1
The basic steps in gene cloning.

1.4 What is PCR?

The polymerase chain reaction is very different from gene cloning. Rather than a series of manipulations involving living cells, PCR is carried out in a single test tube simply by mixing DNA with a set of reagents and placing the tube in a thermal cycler, a piece of equipment that enables the mixture to be incubated at a series of temperatures that are varied in a preprogrammed manner. The basic steps in a PCR experiment are as follows (Figure 1.2):

1 The mixture is heated to 94 °C, at which temperature the hydrogen bonds that hold together the two strands of the double-stranded DNA molecule are broken, causing the molecule to **denature**.

Figure 1.2
The basic steps in the polymerase chain reaction.



2 The mixture is cooled down to 50–60 °C. The two strands of each molecule could join back together at this temperature, but most do not because the mixture contains a large excess of short DNA molecules, called **oligonucleotides** or **primers**, which **anneal** to the DNA molecules at specific positions.

3 The temperature is raised to 74 °C. This is a good working temperature for the ***Taq* DNA polymerase** that is present in the mixture. We will learn more about **DNA polymerases** on p. 51. All we need to understand at this stage is that the *Taq* DNA polymerase attaches to one end of each primer and synthesizes new strands of DNA, complementary to the **template** DNA molecules, during this step of the PCR. Now we have four stands of DNA instead of the two that there were to start with.

4 The temperature is increased back to 94 °C. The double-stranded DNA molecules, each of which consists of one strand of the original molecule and one new strand of DNA, denature into single strands. This begins a second cycle of denaturation–annealing–synthesis, at the end of which there are eight DNA strands. By repeating the cycle 30 times the double-stranded molecule that we began with is converted into over 130 million new double-stranded molecules, each one a copy of the region of the starting molecule delineated by the annealing sites of the two primers.

1.5 Why gene cloning and PCR are so important

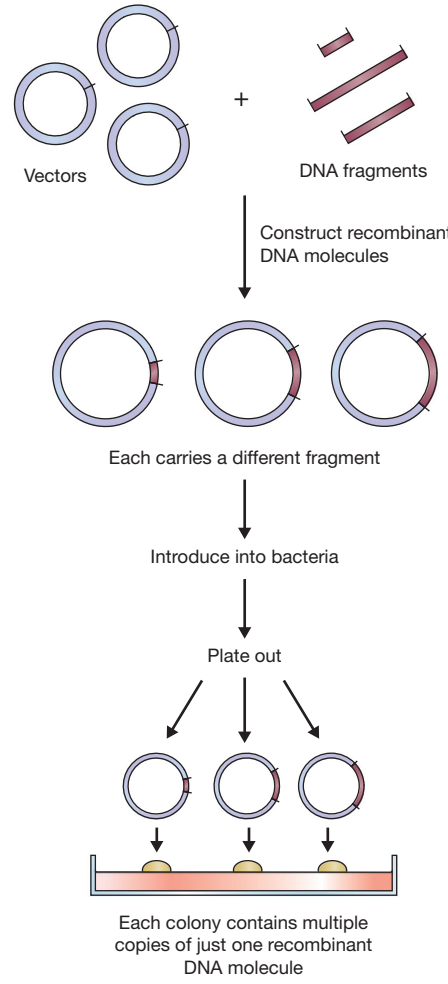
As can be seen from Figures 1.1 and 1.2, gene cloning and PCR are relatively straightforward procedures. Why, then, have they assumed such importance in biology? The answer is largely because both techniques can provide a pure sample of an individual gene, separated from all the other genes in the cell.

1.5.1 Obtaining a pure sample of a gene by cloning

To understand exactly how cloning can provide a pure sample of a gene, consider the basic experiment from Figure 1.1, but drawn in a slightly different way (Figure 1.3). In this example, the DNA fragment to be cloned is one member of a mixture of many different fragments, each carrying a different gene or part of a gene. This mixture could indeed be the entire genetic complement of an organism – a human, for instance. Each of these fragments becomes inserted into a different vector molecule to produce a family of recombinant DNA molecules, one of which carries the gene of interest. Usually, only one recombinant DNA molecule is transported into any single host cell, so that although the final set of clones may contain many different recombinant DNA molecules, each individual clone contains multiple copies of just one molecule. The gene is now separated away from all the other genes in the original mixture, and its specific features can be studied in detail.

In practice, the key to the success or failure of a gene cloning experiment is the ability to identify the particular clone of interest from the many different ones that are obtained. If we consider the **genome** of the bacterium *Escherichia coli*, which contains just over 4000 different genes, we might at first despair of being able to find just one gene among all the possible clones (Figure 1.4). The problem becomes even more overwhelming when we remember that bacteria are relatively simple organisms and that the human genome contains about five times as many genes. As explained in Chapter 8, a variety

Figure 1.3
Cloning allows individual fragments of DNA to be purified.



of different strategies can be used to ensure that the correct gene can be obtained at the end of the cloning experiment. Some of these strategies involve modifications to the basic cloning procedure, so that only cells containing the desired recombinant DNA molecule can divide and the clone of interest is automatically **selected**. Other methods involve techniques that enable the desired clone to be identified from a mixture of lots of different clones.

Once a gene has been cloned there is almost no limit to the information that can be obtained about its structure and expression. The availability of cloned material has stimulated the development of many different analytical methods for studying genes, with new techniques being introduced all the time. Methods for studying the structure and expression of a cloned gene are described in Chapters 10 and 11, respectively.

1.5.2 PCR can also be used to purify a gene

The polymerase chain reaction can also be used to obtain a pure sample of a gene. This is because the region of the starting DNA molecule that is copied during PCR

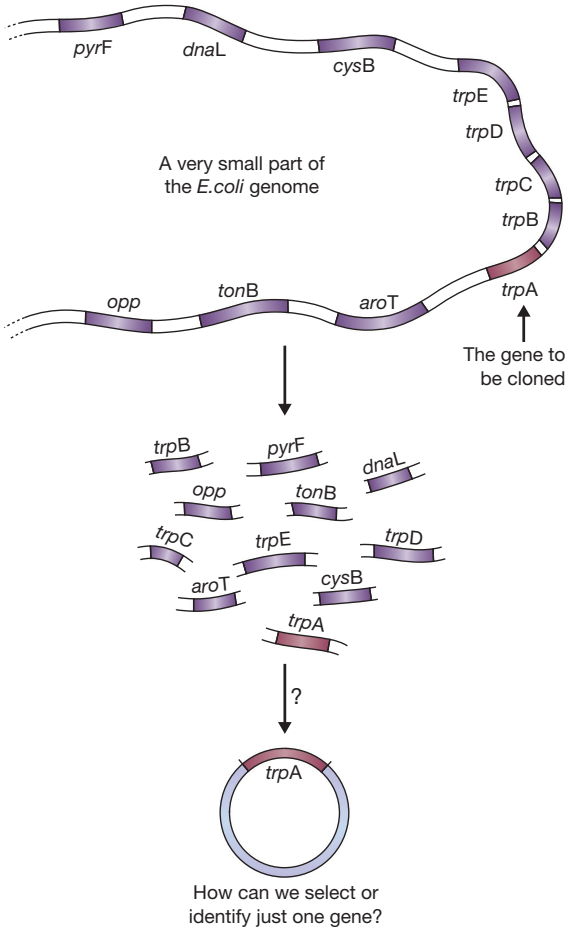


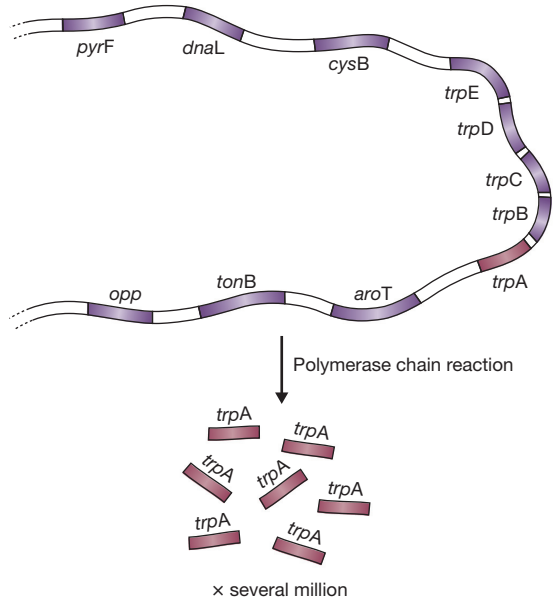
Figure 1.4
The problem of selection.

is the segment whose boundaries are marked by the annealing positions of the two oligonucleotide primers. If the primers anneal either side of the gene of interest, many copies of that gene will be synthesized (Figure 1.5). The outcome is the same as with a gene cloning experiment, although the problem of selection does not arise because the desired gene is automatically 'selected' as a result of the positions at which the primers anneal.

A PCR experiment can be completed in a few hours, whereas it takes weeks – if not months – to obtain a gene by cloning. Why then is gene cloning still used? This is because PCR has two limitations:

- In order for the primers to anneal to the correct positions, on either side of the gene of interest, the sequences of these annealing sites must be known. It is easy to synthesize a primer with a predetermined sequence (see p. 149), but if the sequences of the annealing sites are unknown then the appropriate primers cannot be made. This means that PCR cannot be used to isolate genes that have not been studied before – that has to be done by cloning.

Figure 1.5
Gene isolation by PCR.



- There is a limit to the length of DNA sequence that can be copied by PCR. Five kilobases (kb) can be copied fairly easily, and segments up to 40 kb can be dealt with by using specialized techniques, but this is shorter than the lengths of many genes, especially those of humans and other vertebrates. Cloning must be used if an intact version of a long gene is required.

Gene cloning is therefore the only way of isolating long genes or those that have never been studied before. But PCR still has many important applications. For example, even if the sequence of a gene is not known, it may still be possible to determine the appropriate sequences for a pair of primers, based on what is known about the sequence of the equivalent gene in a different organism. A gene that has been isolated and sequenced from, say, mouse could therefore be used to design a pair of primers for isolation of the equivalent gene from humans.

In addition, there are many applications where it is necessary to isolate or detect genes whose sequences are already known. A PCR of human globin genes, for example, is used to test for the presence of mutations that might cause the blood disease called thalassaemia. The design of appropriate primers for this PCR is easy because the sequences of the human globin genes are known. After the PCR, the gene copies are sequenced or studied in some other way to determine if any of the thalassaemia mutations are present.

Another clinical application of PCR involves the use of primers specific for the DNA of a disease-causing virus. A positive result indicates that a sample contains the virus and that the person who provided the sample should undergo treatment to prevent onset of the disease. The PCR is tremendously sensitive, with a carefully set up reaction yielding detectable amounts of DNA even if there is just one DNA molecule in the starting mixture. This means that the technique can detect viruses at the earliest stages of an infection, increasing the chances of a treatment being successful. This great sensitivity

means that PCR can also be used with DNA from forensic material such as hairs and dried bloodstains, or even from the bones of long-dead humans (see Chapter 16).

1.6 How to find your way through this book

This book explains how gene cloning, PCR and other DNA analysis techniques are carried out, and describes the applications of these techniques in modern biology. The applications are covered in the second and third parts of the book. Part II describes how genes and genomes are studied, while Part III provides accounts of the broader applications of gene cloning and PCR in biotechnology, medicine, agriculture, and forensic science.

In Part I we deal with the basic principles. Most of the nine chapters in Part I are devoted to gene cloning because this technique is more complicated than PCR. When you have understood how cloning is carried out you will have understood many of the basic principles of how DNA is analysed. In Chapter 2 we look at the central component of a gene cloning experiment – the vector – which transports the gene into the host cell and is responsible for its replication. To act as a cloning vector a DNA molecule must be capable of entering a host cell and, once inside, replicating to produce multiple copies of itself. Two naturally occurring types of DNA molecule satisfy these requirements:

- **Plasmids**, which are small circles of DNA found in bacteria and some other organisms. Plasmids can replicate independently of the host cell chromosome.
- **Virus chromosomes**, in particular the chromosomes of **bacteriophages**, which are viruses that specifically infect bacteria. During infection the bacteriophage DNA molecule is injected into the host cell where it undergoes replication.

Chapter 3 describes how DNA is purified from living cells – both the DNA that will be cloned and the vector DNA – and Chapter 4 covers the various techniques for handling purified DNA molecules in the laboratory. There are many such techniques, but two are particularly important in gene cloning. These are the ability first to cut the vector at a specific point, and then to repair it in such a way that the gene is inserted (see Figure 1.1). These and other DNA manipulations were developed as an offshoot of basic research into DNA synthesis and modification in living cells, and most of the manipulations make use of purified enzymes. The properties of these enzymes, and the way they are used in DNA studies, are described in Chapter 4.

Once a recombinant DNA molecule has been constructed, it must be introduced into the host cell so that replication can take place. Transport into the host cell makes use of natural processes for the uptake of plasmid and viral DNA molecules. These processes and the ways they are utilized in gene cloning are described in Chapter 5, while the most important types of cloning vector are introduced, and their uses examined, in Chapters 6 and 7. To conclude the coverage of gene cloning, in Chapter 8 we investigate the problem of selection (see Figure 1.4), before returning in Chapter 9 to a more detailed description of PCR and its related techniques.

Chapter 2

Vectors for Gene Cloning: Plasmids and Bacteriophages

Chapter contents

CHAPTER CONTENTS

A DNA molecule needs to display several features to be able to act as a vector for gene cloning. Most importantly, it must be able to replicate within the host cell, so that numerous copies of the recombinant DNA molecule can be produced and passed to the daughter cells. A cloning vector also needs to be relatively small, ideally less than 10 kb in size, as large molecules tend to break down during purification, and are also more difficult to manipulate. Two kinds of DNA molecule that satisfy these criteria can be found in bacterial cells, namely plasmids and bacteriophage chromosomes.

2.1 Plasmids

Plasmids are circular molecules of DNA that lead an independent existence in the bacterial cell (Figure 2.1). Plasmids almost always carry one or more genes, and often these genes are responsible for a useful characteristic displayed by the host bacterium. For example, the ability to survive in normally toxic concentrations of antibiotics such as chloramphenicol or ampicillin is often due to the presence in the bacterium of a plasmid carrying antibiotic resistance genes. In the laboratory, antibiotic resistance is often used as a **selectable marker** to ensure that bacteria in a culture contain a particular plasmid (Figure 2.2).

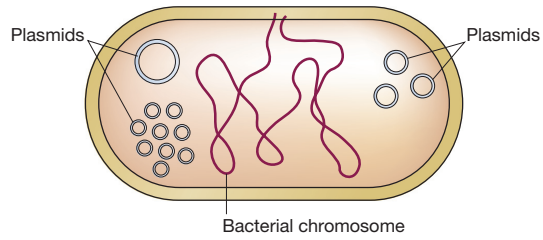
Most plasmids possess at least one DNA sequence that can act as an **origin of replication**, so they are able to multiply within the cell independently of the main bacterial

Gene Cloning and DNA Analysis: An Introduction, Seventh Edition. T.A. Brown.

Companion Website: www.wiley.com/go/brown/cloning

Figure 2.1

Plasmids: independent genetic elements found in bacterial cells.



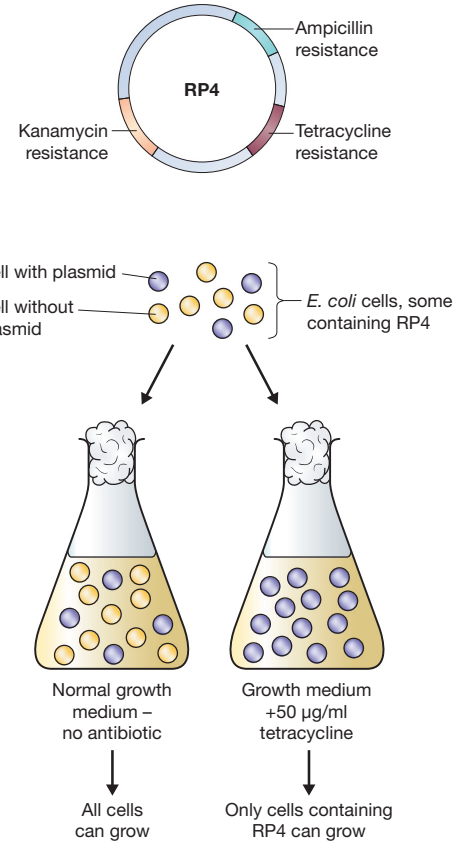
chromosome (Figure 2.3a). The smaller plasmids make use of the host cell's own DNA replicative enzymes in order to make copies of themselves, whereas some of the larger ones carry genes that code for special enzymes that are specific for plasmid replication. A few types of plasmid are also able to replicate by inserting themselves into the bacterial chromosome (Figure 2.3b). These integrative plasmids or **episomes** may be stably maintained in this form through numerous cell divisions, but they always at some stage exist as independent elements.

2.1.1 Size and copy number

The size and **copy number** of a plasmid are particularly important as far as cloning is concerned. We have already mentioned the relevance of plasmid size, and stated that less

Figure 2.2

The use of antibiotic resistance as a selectable marker for a plasmid. RP4 (top) carries genes for resistance to ampicillin, tetracycline, and kanamycin. Only those *E. coli* cells that contain RP4 (or a related plasmid) are able to survive and grow in a medium that contains toxic amounts of one or more of these antibiotics.



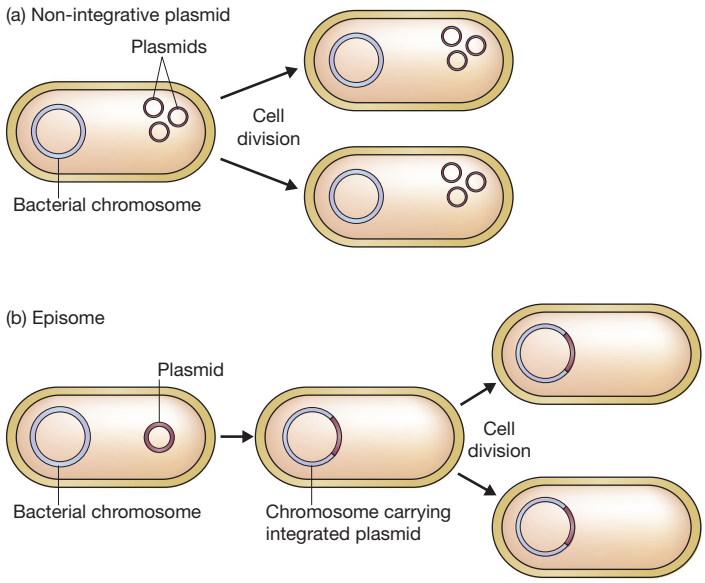


Figure 2.3

Replication strategies for (a) a non-integrative plasmid, and (b) an episome.

than 10 kb is desirable for a cloning vector. Plasmids range from about 1.0 kb for the smallest to over 250 kb for the largest plasmids (Table 2.1), so only a few are useful for cloning purposes. However, as we will see in Chapter 7, larger plasmids can be adapted for cloning under some circumstances.

The copy number refers to the number of molecules of an individual plasmid that are normally found in a single bacterial cell. The factors that control copy number are not well understood. Some plasmids, especially the larger ones, are **stringent** and have a low copy number of perhaps just one or two per cell; others, called **relaxed** plasmids, are present in multiple copies of 50 or more per cell. Generally speaking, a useful cloning vector needs to be present in the cell in multiple copies so that large quantities of the recombinant DNA molecule can be obtained.

Table 2.1

Sizes of representative plasmids.

PLASMID	SIZE		ORGANISM
	NUCLEOTIDE LENGTH (kb)	MOLECULAR MASS (MDa)	
pUC8	2.1	1.8	*E. coli*
ColE1	6.4	4.2	*E. coli*
RP4	54	36	*Pseudomonas* and others
F	95	63	*E. coli*
TOL	117	78	*Pseudomonas putida*
pTiAch5	213	142	*Agrobacterium tumefaciens*

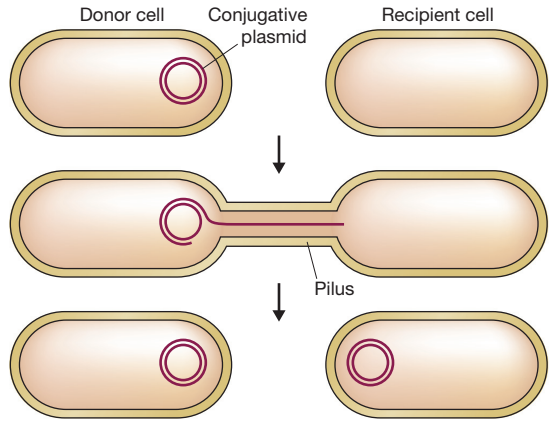


Figure 2.4
Plasmid transfer by conjugation between bacterial cells. The donor and recipient cells attach to each other by a pilus, a hollow appendage present on the surface of the donor cell. A copy of the plasmid is then passed to the recipient cell. Transfer is thought to occur through the pilus, but this has not been proven and transfer by some other means (e.g., directly across the bacterial cell walls) remains a possibility.

2.1.2 Conjugation and compatibility

Plasmids fall into two groups: conjugative and non-conjugative. Conjugative plasmids are characterized by the ability to promote sexual **conjugation** between bacterial cells (Figure 2.4), a process that can result in a conjugative plasmid spreading from one cell to all the other cells in a bacterial culture. Conjugation and plasmid transfer are controlled by a set of transfer or *tra* genes, which are present on conjugative plasmids but absent from the non-conjugative type. A non-conjugative plasmid may, under some circumstances, be cotransferred along with a conjugative plasmid when both are present in the same cell.

Several different kinds of plasmid may be found in a single cell, including more than one different conjugative plasmid at any one time. In fact, cells of *E. coli* have been known to contain up to seven different plasmids at once. To be able to coexist in the same cell, different plasmids must be **compatible**. If two plasmids are incompatible, then one or the other will be rapidly lost from the cell. Different types of plasmid can therefore be assigned to different **incompatibility groups** on the basis of whether or not they can coexist, and plasmids from a single incompatibility group are often related to each other in various ways. The basis of incompatibility is not well understood, but events during plasmid replication are thought to underlie the phenomenon.

2.1.3 Plasmid classification

The most useful classification of naturally occurring plasmids is based on the main characteristic coded by the plasmid genes. The five major types of plasmid according to this classification are as follows:

- **Fertility** or **F plasmids** carry only *tra* genes and have no characteristic beyond the ability to promote conjugal transfer of plasmids. A well-known example is the F plasmid of *E. coli*.
- **Resistance** or **R plasmids** carry genes conferring on the host bacterium resistance to one or more antibacterial agents, such as chloramphenicol, ampicillin, and mercury. R plasmids are very important in clinical microbiology as their spread through

natural populations can have profound consequences in the treatment of bacterial infections. An example is RP4, which is commonly found in *Pseudomonas*, but also occurs in many other bacteria.

- **Col plasmids** code for colicins, proteins that kill other bacteria. An example is ColE1 of *E. coli*.
- **Degradative plasmids** allow the host bacterium to metabolize unusual molecules such as toluene and salicylic acid, an example being TOL of *Pseudomonas putida*.
- **Virulence plasmids** confer pathogenicity on the host bacterium; these include the **Ti plasmids** of *Agrobacterium tumefaciens*, which induce crown gall disease on dicotyledonous plants.

2.1.4 Plasmids in organisms other than bacteria

Although plasmids are widespread in bacteria they are by no means as common in other organisms. The best-characterized eukaryotic plasmid is the **2 μm circle** that occurs in many strains of the yeast *Saccharomyces cerevisiae*. The discovery of the 2 μm plasmid was very fortuitous as it allowed the construction of cloning vectors for this very important industrial organism (p. 111). The search for plasmids in other eukaryotes (such as filamentous fungi, plants and animals) has proved disappointing, and it is suspected that many higher organisms simply do not harbour plasmids within their cells.

2.2 Bacteriophages

Bacteriophages, or phages as they are commonly known, are viruses that specifically infect bacteria. Like all viruses, phages are very simple in structure, consisting merely of a DNA (or occasionally ribonucleic acid (RNA)) molecule carrying a number of genes, including several for replication of the phage, surrounded by a protective coat or **capsid** made up of protein molecules (Figure 2.5).

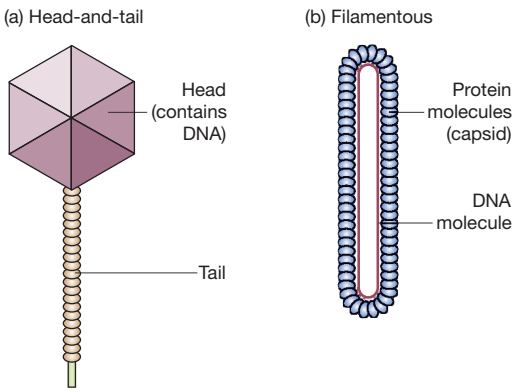


Figure 2.5
The two main types of phage structure. (a) Head-and-tail (e.g., λ). (b) Filamentous (e.g., M13).

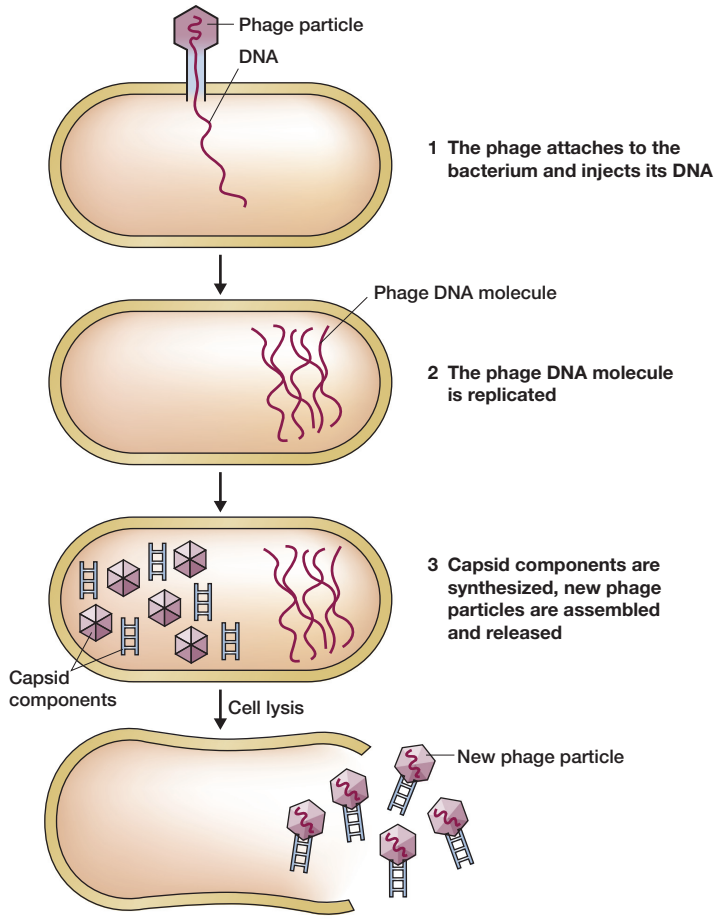


Figure 2.6
The general pattern of infection of a bacterial cell by a bacteriophage.

2.2.1 The phage infection cycle

The general pattern of infection, which is the same for all types of phage, is a three-step process (Figure 2.6):

1. The phage particle attaches to the outside of the bacterium and injects its DNA chromosome into the cell.
2. The phage DNA molecule is replicated, usually by specific phage enzymes coded by genes in the phage chromosome.
3. Other phage genes direct synthesis of the protein components of the capsid, and new phage particles are assembled and released from the bacterium.

With some phage types the entire infection cycle is completed very quickly, possibly in less than 20 minutes. This type of rapid infection is called a **lytic cycle**, as release of the new phage particles is associated with lysis of the bacterial cell. The characteristic feature of a lytic infection cycle is that phage DNA replication is immediately followed

by synthesis of capsid proteins, and the phage DNA molecule is never maintained in a stable condition in the host cell.

2.2.2 Lysogenic phages

In contrast to a lytic cycle, **lysogenic** infection is characterized by retention of the phage DNA molecule in the host bacterium, possibly for many thousands of cell divisions. With many lysogenic phages the phage DNA is inserted into the bacterial genome, in a manner similar to episomal insertion (see Figure 2.3b). The integrated form of the phage DNA (called the **prophage**) is quiescent, and a bacterium (referred to as a **lysogen**) that carries a prophage is usually physiologically indistinguishable from an uninfected cell. The prophage is eventually released from the host genome and the phage reverts to the lytic mode and lyses the cell. The infection cycle of **lambda (λ)**, a typical lysogenic phage of this type, is shown in Figure 2.7.

A limited number of lysogenic phages follow a rather different infection cycle. When **M13** or a related phage infects *E. coli*, new phage particles are continuously assembled and released from the cell. The M13 DNA is not integrated into the bacterial genome and does not become quiescent. With these phages, cell lysis never occurs, and the infected bacterium can continue to grow and divide, albeit at a slower rate than uninfected cells. Figure 2.8 shows the M13 infection cycle.

Although there are many different varieties of bacteriophage, only λ and M13 have found a major role as cloning vectors. We will now consider the properties of these two phages in more detail.

Gene organization in the λ DNA molecule

Lambda is a typical example of a head-and-tail phage (see Figure 2.5a). The DNA is contained in the polyhedral head structure and the tail serves to attach the phage to the bacterial surface and to inject the DNA into the cell (see Figure 2.7).

The λ DNA molecule is 49 kb in size and has been intensively studied by the techniques of gene mapping and DNA sequencing. As a result, the positions and identities of all of the genes in the λ DNA molecule are known (Figure 2.9). A feature of the λ genetic map is that genes related in terms of function are clustered together in the genome. For example, all of the genes coding for components of the capsid are grouped together in the left-hand third of the molecule, and genes controlling integration of the prophage into the host genome are clustered in the middle of the molecule. Clustering of related genes is profoundly important for controlling expression of the λ genome, as it allows genes to be switched on and off as a group rather than individually. Clustering is also important in the construction of λ-based cloning vectors, as we will discover when we return to this topic in Chapter 6.

The linear and circular forms of λ DNA

A second feature of λ that turns out to be of importance in the construction of cloning vectors is the conformation of the DNA molecule. The molecule shown in Figure 2.9 is linear, with two free ends, and represents the DNA present in the phage head structure. This linear molecule consists of two **complementary** strands of DNA, base-paired according to the **Watson–Crick rules** (that is, double-stranded DNA). At either end of the molecule is a short 12-nucleotide stretch in which the DNA is single-stranded (Figure 2.10a); these two single strands are complementary and so can

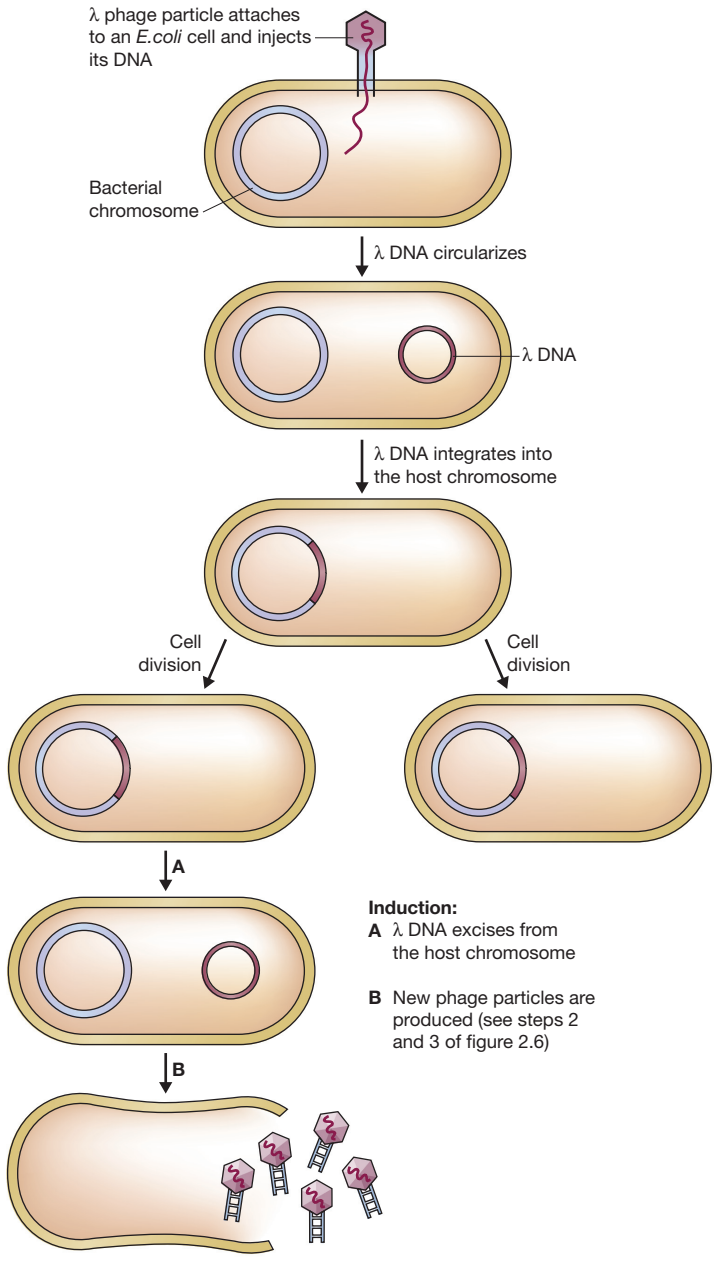


Figure 2.7

The lysogenic infection cycle of bacteriophage λ.

base pair with one another to form a circular, completely double-stranded molecule (Figure 2.10b).

Complementary single strands are often referred to as **'sticky' ends** or cohesive ends, because base pairing between them can 'stick' together the two ends of a DNA molecule (or the ends of two different DNA molecules). The λ cohesive ends are called the ***cos***

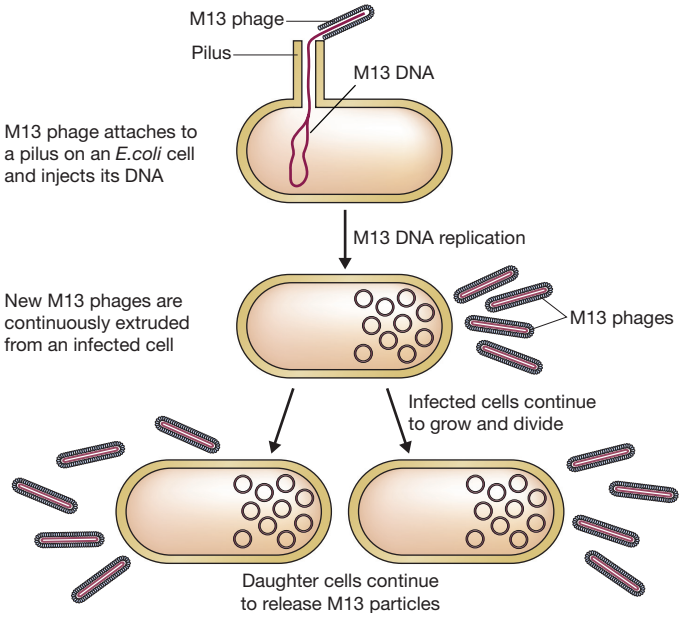


Figure 2.8

The infection cycle of bacteriophage M13.

sites and they play two distinct roles during the λ infection cycle. First, they allow the linear DNA molecule that is injected into the cell to be circularized, which is a necessary prerequisite for insertion into the bacterial genome (see Figure 2.7).

The second role of the *cos* sites is rather different, and comes into play after the prophage has excised from the host genome. At this stage a large number of new λ DNA molecules are produced by the rolling circle mechanism of replication (Figure 2.10c), in which a continuous DNA strand is 'rolled off' the template molecule. The result is a catenane consisting of a series of linear λ genomes joined together at the *cos* sites. The role of the *cos* sites is now to act as recognition sequences for an **endonuclease** that cleaves the catenane at the *cos* sites, producing individual λ genomes. This endonuclease, which is the product of gene *A* on the DNA molecule, creates the single-stranded sticky ends, and also acts in conjunction with other proteins to package each λ genome into a phage head structure. The cleavage and packaging processes recognize just the

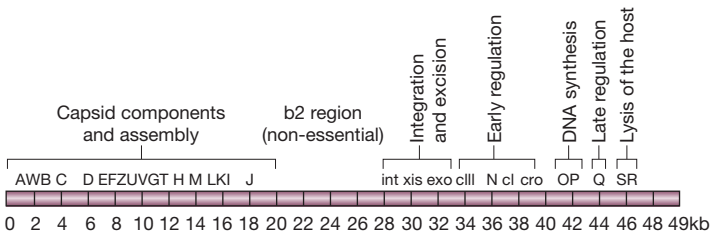


Figure 2.9

The λ genetic map, showing the positions of the important genes and the functions of the gene clusters.

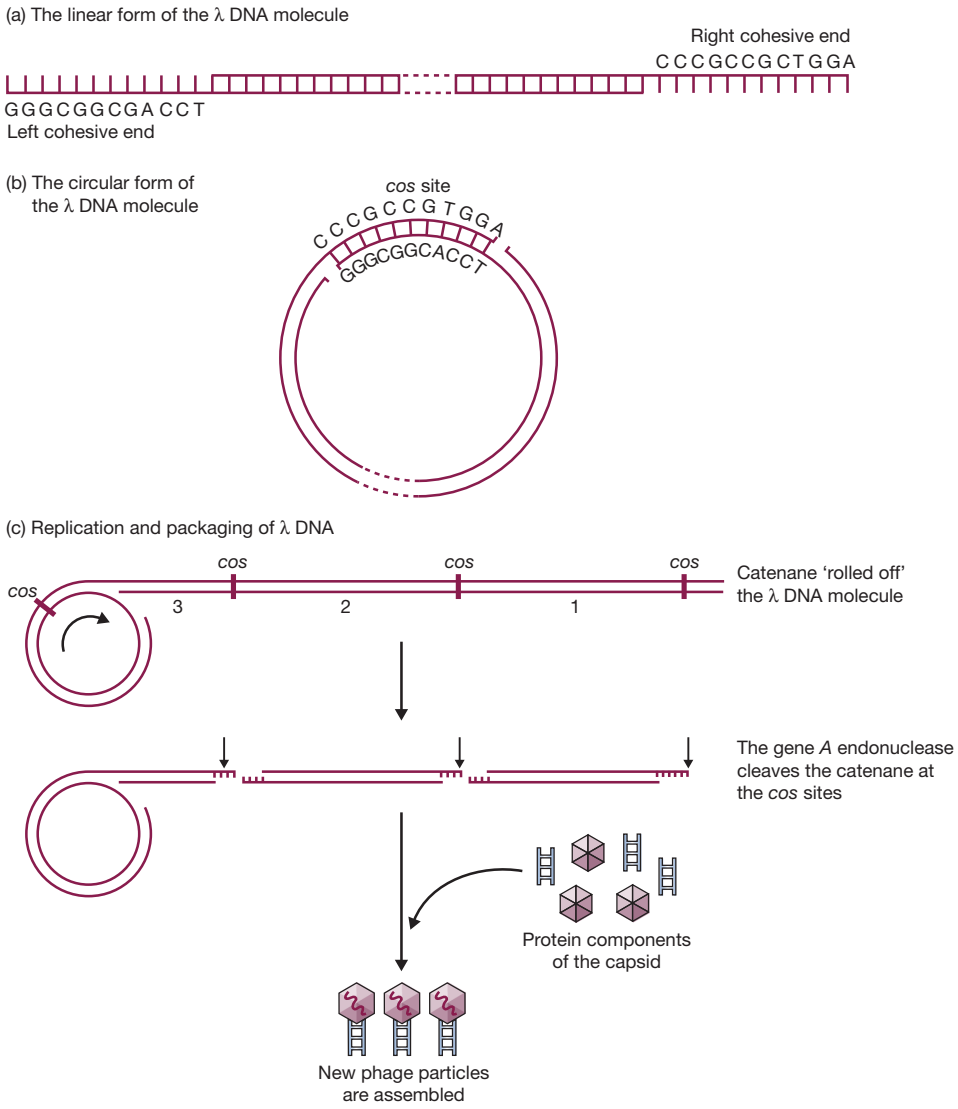


Figure 2.10

The linear and circular forms of λ DNA. (a) The linear form, showing the left and right cohesive ends. (b) Base pairing between the cohesive ends results in the circular form of the molecule. (c) Rolling circle replication produces a catenane of new linear λ DNA molecules, which are individually packaged into phage heads as new λ particles are assembled.

cos sites and the DNA sequences to either side of them, so changing the structure of the internal regions of the λ genome, for example by inserting new genes, has no effect on these events so long as the overall length of the λ genome is not altered too greatly.

M13 – a filamentous phage

M13 is an example of a filamentous phage (see Figure 2.5b) and is completely different in structure from λ. Furthermore, the M13 DNA molecule is much smaller than the λ

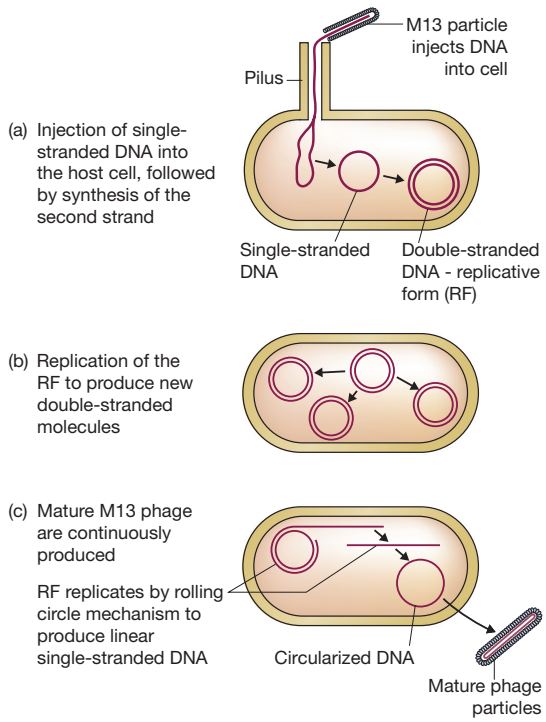


Figure 2.11

The M13 infection cycle, showing the different types of DNA replication that occur. (a) After infection, the single-stranded M13 DNA molecule is converted into the double-stranded replicative form (RF). (b) The RF replicates to produce multiple copies of itself. (c) Single-stranded molecules are synthesized by rolling circle replication and used in the assembly of new M13 particles.

genome, being only 6407 nucleotides in length. It is circular and is unusual in that it consists entirely of single-stranded DNA.

The smaller size of the M13 DNA molecule means that it has room for fewer genes than the λ genome. This is possible because the M13 capsid is constructed from multiple copies of just three proteins (requiring only three genes), whereas synthesis of the λ head-and-tail structure involves over 15 different proteins. In addition, M13 follows a simpler infection cycle than λ, and does not need genes for insertion into the host genome.

Injection of an M13 DNA molecule into an *E. coli* cell occurs via the **pilus**, the structure that connects two cells during sexual conjugation (see Figure 2.4). Once inside the cell the single-stranded molecule acts as the template for synthesis of a complementary strand, resulting in normal double-stranded DNA (Figure 2.11a). This molecule is not inserted into the bacterial genome, but instead replicates until over 100 copies are present in the cell (Figure 2.11b). When the bacterium divides, each daughter cell receives copies of the phage genome, which continues to replicate, thereby maintaining its overall numbers per cell. As shown in Figure 2.11c, new phage particles are continuously assembled and released, with about 1000 new phages being produced during each generation of an infected cell.

Several features of M13 make this phage attractive as a cloning vector. The genome is less than 10 kb in size, well within the range desirable for a potential vector. In addition, the double-stranded **replicative form (RF)** of the M13 genome behaves very much like a plasmid, and can be treated as such for experimental purposes. It is easily prepared from a culture of infected *E. coli* cells (p. 43) and can be reintroduced by **transfection** (p. 83). Most importantly, genes cloned with an M13-based vector can be obtained in

the form of single-stranded DNA. Single-stranded versions of cloned genes are useful for several techniques, such as *in vitro* mutagenesis (p. 216). Cloning in an M13 vector is an easy and reliable way of obtaining single-stranded DNA for this type of work. M13 vectors are also used in **phage display**, a technique for identifying pairs of genes whose protein products interact with one another (p. 241).

2.2.3 Viruses as cloning vectors for other organisms

Most living organisms are infected by viruses, and it is not surprising that there has been great interest in the possibility that viruses might be used as cloning vectors for higher organisms. This is especially important when it is remembered that plasmids are not commonly found in organisms other than bacteria and yeast. Several eukaryotic viruses have been employed as cloning vectors for specialized applications. For example, human **adenoviruses** are used in **gene therapy** (p. 286), **baculoviruses** are used to synthesize important pharmaceutical proteins in insect cells (p. 262), and **caulimoviruses** and **geminiviruses** have been used for cloning in plants (p. 127). These vectors are discussed more fully in Chapter 7.

Further reading

FURTHER READING

Dale, J.W. and Park, S.T. (2010) *Molecular Genetics of Bacteria*, 5th edn, Wiley-Blackwell, Chichester. [Provides a detailed description of plasmids and bacteriophages.]

Willey, J., Sherwood, L., and Woolverton, C. (2013) *Prescott's Microbiology*, 9th edn, McGraw-Hill Higher Education, Maidenhead. [A good introduction to microbiology, including plasmids and phages.]

Chapter 5

Introduction of DNA into Living Cells

Chapter contents

CHAPTER CONTENTS

The manipulations described in Chapter 4 allow the molecular biologist to create novel recombinant DNA molecules. The next step in a gene cloning experiment is to introduce these molecules into living cells, usually bacteria, which then grow and divide to produce clones (see Figure 1.1). Strictly speaking, the word 'cloning' refers only to the later stages of the procedure, and not to the construction of the recombinant DNA molecule itself.

Cloning serves two main purposes. First, it allows a large number of recombinant DNA molecules to be produced from a limited amount of starting material. At the outset only a few nanograms of recombinant DNA may be available, but each bacterium that takes up a plasmid subsequently divides numerous times to produce a colony, each cell of which contains multiple copies of the molecule. Several micrograms of recombinant DNA can usually be prepared from a single bacterial colony, representing a thousandfold increase over the starting amount (Figure 5.1). If the colony is used not as a source of DNA but as an inoculum for a liquid culture, the resulting cells may provide milligrams of DNA, a millionfold increase in yield. In this way, cloning can supply the large amounts of DNA needed for molecular biological studies of gene structure and expression (see Chapters 10 and 11).

The second important function of cloning can be described as purification. The manipulations that result in a recombinant DNA molecule can only rarely be controlled to the extent that no other DNA molecules are present at the end of the procedure. The

Gene Cloning and DNA Analysis: An Introduction, Seventh Edition. T.A. Brown.

Companion Website: www.wiley.com/go/brown/cloning

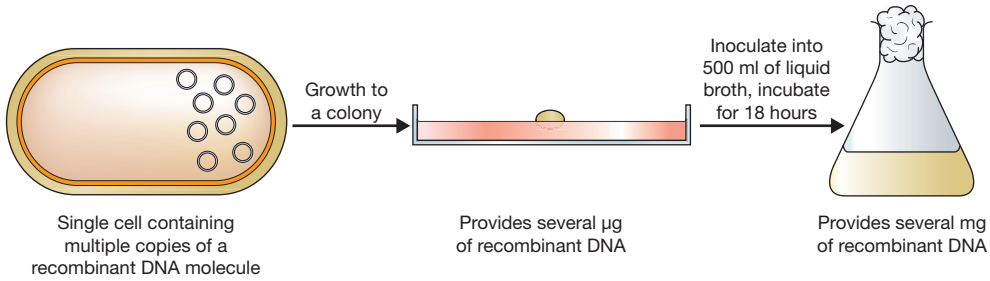


Figure 5.1

Cloning can supply large amounts of recombinant DNA.

ligation mixture may contain, in addition to the desired recombinant molecule, any number of the following (Figure 5.2a):

- Unligated vector molecules.
- Unligated DNA fragments.
- Vector molecules that have recircularized without new DNA being inserted ('self-ligated' vector).
- Recombinant DNA molecules that carry the wrong inserted DNA fragment.

Unligated molecules rarely cause a problem because, even though they may be taken up by bacterial cells, only under exceptional circumstances will they be replicated. It is much more likely that enzymes within the host bacteria will degrade these pieces of DNA. Self-ligated vector molecules and incorrect recombinant plasmids are more important because they are replicated just as efficiently as the desired molecule (Figure 5.2b). Purification of the desired molecule can still be achieved through cloning because it is extremely unusual for any one cell to take up more than one DNA molecule. As each cell gives rise to a single colony, each of the resulting clones will be made up of cells that all contain the same molecule. Of course, different colonies contain different molecules: some contain the desired recombinant DNA molecule, some have different recombinant molecules, and some contain self-ligated vector. The problem therefore becomes a question of identifying those colonies that contain the correct recombinant plasmids.

This chapter is concerned with the way in which plasmid and phage vectors, and recombinant molecules derived from them, are introduced into bacterial cells. During the course of the chapter it will become apparent that selection for colonies containing recombinant molecules, as opposed to colonies containing self-ligated vector, is relatively easy. The more difficult proposition of how to distinguish clones containing the correct recombinant DNA molecule from all the other recombinant clones will be tackled in Chapter 8.

5.1 Transformation: The uptake of DNA by bacterial cells

Most species of bacteria are able to take up DNA molecules from the medium in which they grow. Often, a DNA molecule taken up in this way will be degraded, but

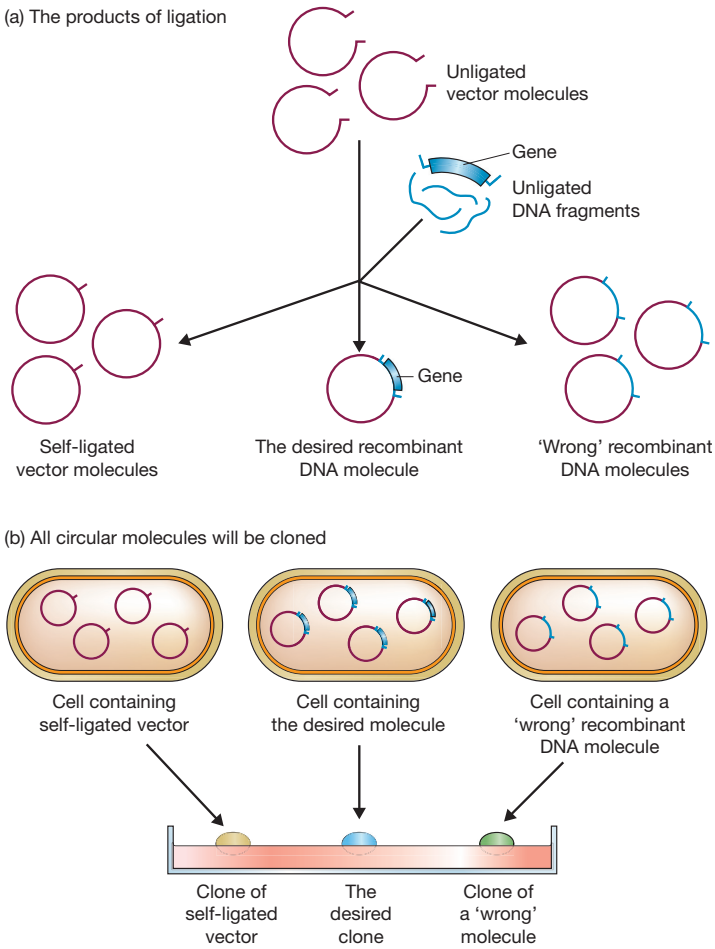


Figure 5.2

Cloning is analogous to purification. From a mixture of different molecules, clones containing copies of just one molecule can be obtained.

occasionally it is able to survive and replicate in the host cell. In particular, this will happen if the DNA molecule is a plasmid with an origin of replication that is recognized by the host.

5.1.1 Not all species of bacteria are equally efficient at DNA uptake

In nature, transformation is probably not a major process by which bacteria obtain genetic information. This is reflected by the fact that, in the laboratory, only a few species (notably members of the genera *Bacillus* and *Streptococcus*) can be transformed with ease. Close studies of these organisms have revealed that they possess sophisticated mechanisms for DNA binding and uptake.

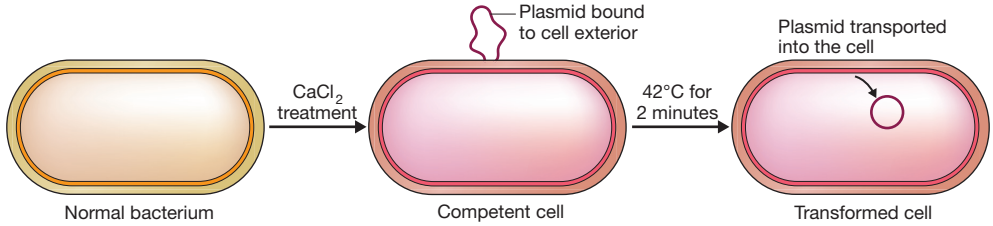


Figure 5.3

The binding and uptake of DNA by a competent bacterial cell.

Most species of bacteria, including *E. coli*, take up only limited amounts of DNA under normal circumstances. In order to transform these species efficiently, the bacteria have to undergo some form of physical and/or chemical treatment that enhances their ability to take up DNA. Cells that have undergone this treatment are said to be **competent**.

5.1.2 Preparation of competent E. coli cells

As with many breakthroughs in recombinant DNA technology, the key development as far as transformation is concerned occurred during the early 1970s, when it was observed that *E. coli* cells that had been soaked in an ice-cold salt solution were more efficient at DNA uptake than unsoaked cells. A solution of 50 mM calcium chloride ($CaCl_2$) is traditionally used for this purpose, although other salts (notably rubidium chloride) are also effective.

Exactly why this treatment works is not understood. At one time it was thought that the $CaCl_2$ causes the DNA to precipitate onto the outside of the cells, or is responsible for some type of physical change in the cell wall that improves DNA binding. More recent studies have suggested that the salt treatment induces an over-production of certain outer membrane proteins, including one or more that bind DNA. In any case, soaking in $CaCl_2$ affects only DNA binding, and not the actual uptake into the cell. When DNA is added to treated cells it remains attached to the cell exterior, and is not at this stage transported into the cytoplasm (Figure 5.3). The actual movement of DNA into competent cells is stimulated by briefly raising the temperature to 42 °C. It is possible that this heat shock changes the permeability of the membrane to DNA, or, as with $CaCl_2$ treatment, the heat shock might induce the activity of a membrane protein that transports DNA into the cell.

5.1.3 Selection for transformed cells

The transformation of competent cells is an inefficient procedure, however carefully the cells have been prepared. Although 1 ng of the plasmid vector called pUC8 (p. 97) can yield between 1000 and 10 000 transformants, this represents the uptake of only 0.01% of all the available molecules. Furthermore, 10 000 transformants is only a very small proportion of the total number of cells that are present in a competent culture. This latter fact means that some way must be found to distinguish a cell that has taken up a plasmid from the many thousands that have not been transformed.

The uptake and stable retention of a plasmid is usually detected by looking for expression of the genes carried by the plasmid. For example, *E. coli* cells are normally sensitive

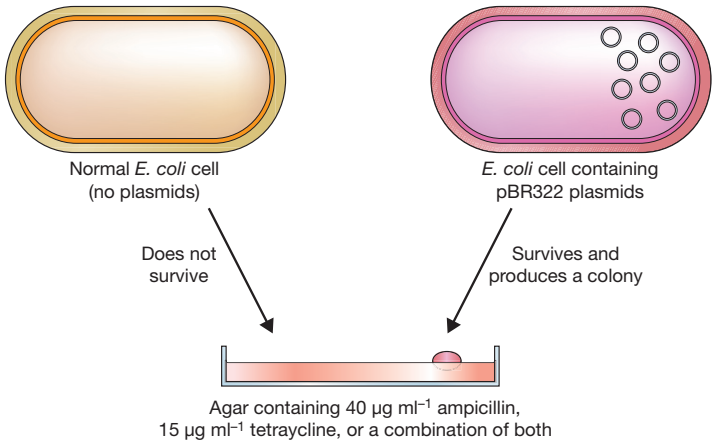


Figure 5.4

Selecting cells that contain pBR322 plasmids by plating onto agar medium containing ampicillin and/or tetracycline.

to the growth-inhibitory effects of the antibiotics ampicillin and tetracycline. Cells that contain the plasmid pBR322 (p. 94), which was one of the first cloning vectors to be developed back in the 1970s, are resistant to these antibiotics. This is because pBR322 carries two sets of genes: one gene set that codes for a β-lactamase enzyme which modifies ampicillin into a form that is non-toxic to the bacterium, and a second set of genes that code for enzymes that detoxify tetracycline. After a transformation experiment with pBR322, only those *E. coli* cells that have taken up a plasmid are $amp^R tet^R$ and able to form colonies on an agar medium that contains ampicillin or tetracycline (Figure 5.4). Non-transformants, which are still $amp^S tet^S$, do not produce colonies on the selective medium. Transformants and non-transformants are therefore easily distinguished from one another.

Most plasmid cloning vectors carry at least one gene that confers antibiotic resistance on the host cells, with the selection of transformants being achieved by plating onto an agar medium that contains the relevant antibiotic. Bear in mind, however, that resistance to the antibiotic is not due merely to the presence of the plasmid in the transformed cells. The resistance gene on the plasmid must also be expressed, so that the enzyme which detoxifies the antibiotic is synthesized. Although expression of the resistance gene begins immediately after transformation, it takes a few minutes before the cell contains enough of the enzyme to be able to withstand the toxic effects of the antibiotic. For this reason, the transformed bacteria should not be plated onto the selective medium immediately after a heat-shock treatment, but should first be placed in a small volume of liquid medium, in the absence of the antibiotic, and incubated for a short time. Plasmid replication and expression can then start so that, when the cells are plated out and encounter the antibiotic, they will already have synthesized sufficient resistance enzymes to be able to survive (Figure 5.5).

5.2 Identification of recombinants

Plating onto a selective medium enables transformants to be distinguished from non-transformants. The next problem is to determine which of the transformed colonies comprise cells that contain recombinant DNA molecules, and which contain self-ligated

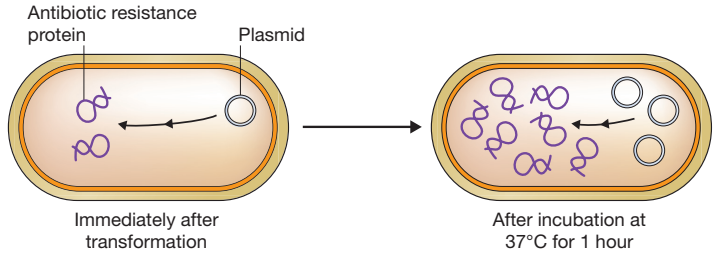


Figure 5.5

Phenotypic expression. Incubation at 37 °C for 1 h before plating out improves the survival of transformants on a selective medium, because the bacteria have had time to begin synthesis of the antibiotic-resistance enzymes.

vector molecules (see Figure 5.2). With most cloning vectors, the insertion of a DNA fragment into the plasmid destroys the integrity of one of the genes present on the molecule. **Recombinants** can therefore be identified because the characteristic coded by the inactivated gene is no longer displayed by the host cells (Figure 5.6). We will explore the general principles of **insertional inactivation** by looking at the different methods used with the two cloning vectors mentioned in Section 5.1, namely pBR322 and pUC8.

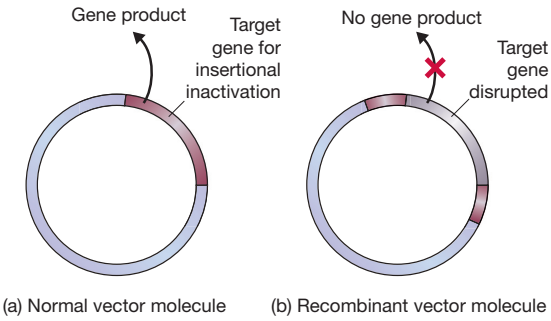
5.2.1 Recombinant selection with pBR322: Insertional inactivation of an antibiotic resistance gene

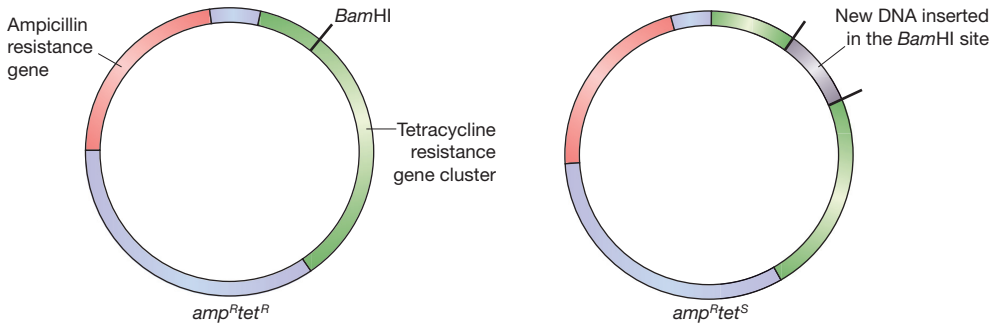
pBR322 has several unique restriction sites that can be used to open up the vector before the insertion of a new DNA fragment (Figure 5.7a). *Bam*HI, for example, cuts pBR322 at just one position, within the cluster of genes that code for resistance to tetracycline. A recombinant pBR322 molecule, one that carries an extra piece of DNA in the *Bam*HI site (Figure 5.7b), is no longer able to confer tetracycline resistance on its host, as one of the necessary genes is now disrupted by the inserted DNA. Cells containing this recombinant pBR322 molecule are still resistant to ampicillin, but are sensitive to tetracycline ($amp^R tet^S$).

Screening for pBR322 recombinants is performed in the following way. After transformation, the cells are plated onto an ampicillin medium and incubated until colonies appear (Figure 5.8a). All of these colonies are transformants (remember, untransformed cells are amp^S and so do not produce colonies on the selective medium) but only a few contain recombinant pBR322 molecules; most will contain the normal, self-ligated plasmid. To identify the recombinants the colonies are **replica plated** onto agar

Figure 5.6

Insertional inactivation. (a) The normal, non-recombinant vector molecule carries a gene whose product confers a selectable or identifiable characteristic on the host cell. (b) This gene is disrupted when new DNA is inserted into the vector; as a result the recombinant host does not display the relevant characteristic.





(a) The normal vector molecule (b) A recombinant pBR322 molecule

Figure 5.7

The cloning vector pBR322. (a) The normal vector molecule. (b) A recombinant molecule containing an extra piece of DNA inserted into the *Bam*HI site. For a more detailed map of pBR322, see Figure 6.1.

medium that contains tetracycline (Figure 5.8b). After incubation, some of the original colonies regrow, but others do not (Figure 5.8c). Those that do grow consist of cells that carry the normal pBR322 with no inserted DNA and therefore a functional tetracycline resistance gene cluster ($amp^R tet^R$). The colonies that do not grow on tetracycline agar are recombinants ($amp^R tet^S$). Reference back to the original ampicillin agar plate reveals the positions of these colonies, enabling samples to be recovered for further study.

5.2.2 Insertional inactivation does not always involve antibiotic resistance

Although the insertional inactivation of an antibiotic resistance gene provides an effective means of recombinant identification, the method is made inconvenient by the need to carry out two screenings: one with the antibiotic that selects for transformants; and a second screen, after replica plating, with the antibiotic that distinguishes recombinants. Most modern plasmid vectors therefore make use of a different system. An example is pUC8 (Figure 5.9a), which carries the ampicillin resistance gene and a gene called *lacZ′*, which codes for part of the enzyme β-galactosidase. Cloning with pUC8 involves insertional inactivation of the *lacZ′* gene, with recombinants identified because of their inability to synthesize β-galactosidase (Figure 5.9b).

β-Galactosidase is one of a series of enzymes involved in the breakdown of lactose to glucose plus galactose. It is normally coded by the gene *lacZ*, which resides on the *E. coli* chromosome. Some strains of *E. coli* have a modified *lacZ* gene, one that lacks the segment referred to as *lacZ′* and coding for the α-peptide portion of β-galactosidase (Figure 5.10a). These mutants can synthesize the enzyme only when they harbour a plasmid, such as pUC8, that carries the missing *lacZ′* segment of the gene.

A cloning experiment with pUC8 involves the selection of transformants on ampicillin agar, followed by screening for β-galactosidase activity to identify recombinants. Cells that harbour a normal pUC8 plasmid are amp^R and able to synthesize β-galactosidase. Recombinants are also amp^R but unable to make β-galactosidase.

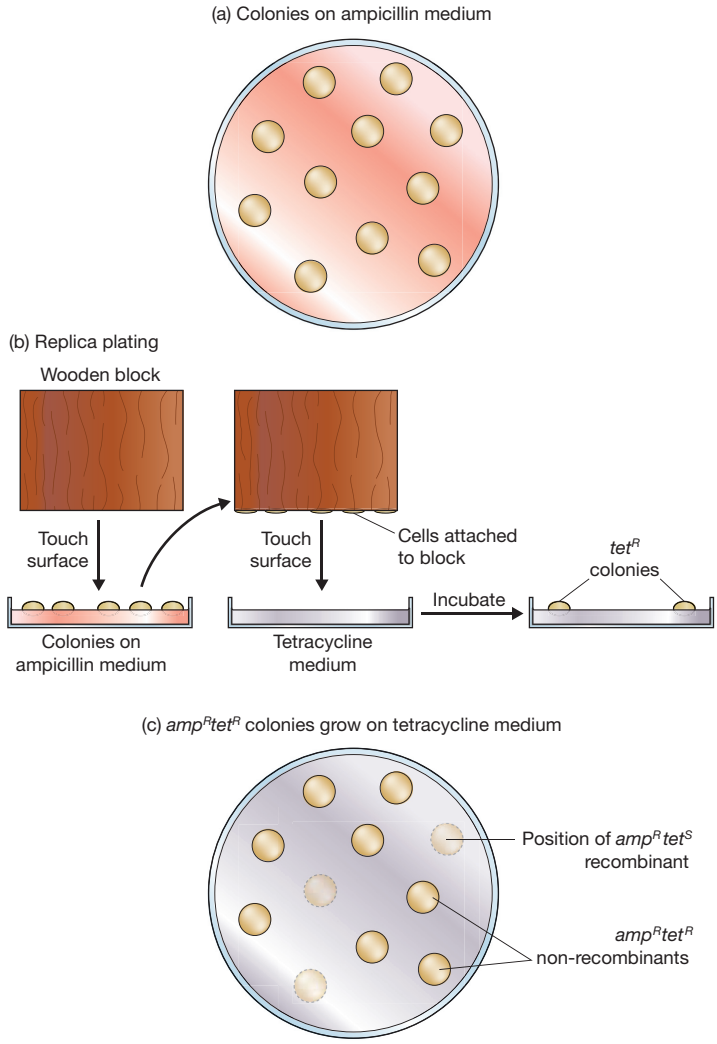


Figure 5.8

Screening for pBR322 recombinants by insertional inactivation of the tetracycline resistance gene. (a) Cells are plated onto ampicillin agar: all the transformants produce colonies. (b) The colonies are replica plated onto tetracycline medium. (c) The colonies that grow on tetracycline medium are $amp^R tet^R$ and therefore non-recombinants. Recombinants ($amp^R tet^S$) do not grow, but their position on the ampicillin plate is now known.

Screening for the presence or absence of β-galactosidase is in fact quite easy. Rather than assay for lactose being split to glucose and galactose, the test is for a slightly different reaction that is also catalysed by β-galactosidase. This involves a lactose analogue termed X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) which is broken down by β-galactosidase to a product that is coloured deep blue. If X-gal (plus an inducer of the enzyme such as isopropylthiogalactoside, IPTG) is added to the agar, along with ampicillin, then non-recombinant colonies, the cells of which synthesize β-galactosidase, will be coloured blue, whereas recombinants with a disrupted *lacZ′* gene and unable to make β-galactosidase, will be white. This system, which is known as **Lac selection**, is summarized in Figure 5.10b. Note that both ampicillin resistance and

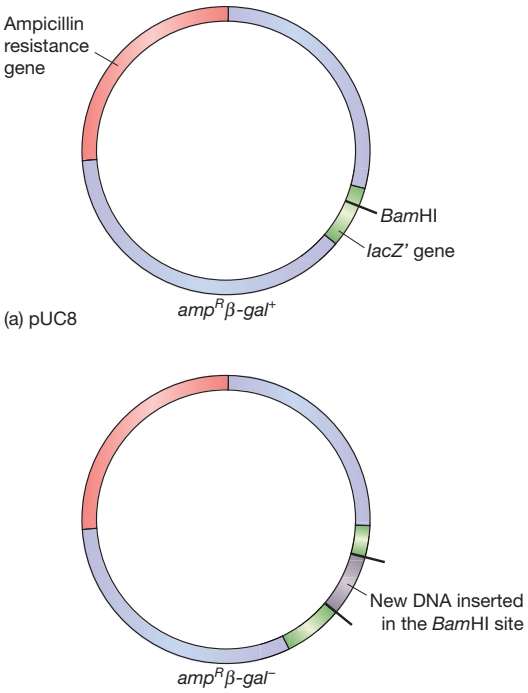


Figure 5.9

The cloning vector pUC8. (a) The normal vector molecule. (b) A recombinant molecule containing an extra piece of DNA inserted into the *Bam*HI site. For a more detailed map of pUC8, see Figure 6.3.

the presence or absence of β-galactosidase are tested for on a single agar plate. The two screenings are therefore carried out together and there is no need for the time-consuming replica plating step that is necessary with plasmids such as pBR322.

5.3 Introduction of phage DNA into bacterial cells

There are two different methods by which a recombinant DNA molecule constructed with a phage vector can be introduced into a bacterial cell, namely transfection and ***in vitro* packaging**.

5.3.1 Transfection

Transfection is equivalent to transformation, the only difference being that phage DNA rather than a plasmid is involved. Just as with a plasmid, the purified phage DNA or recombinant phage molecule is mixed with competent *E. coli* cells and DNA uptake is induced by heat shock. Transfection is the standard method for introducing the double-stranded RF form of an M13 cloning vector into *E. coli*.

5.3.2 In vitro *packaging of λ cloning vectors*

Transfection with λ DNA molecules is not a very efficient process when compared with the infection of a culture of cells with mature λ phage particles. It would therefore be

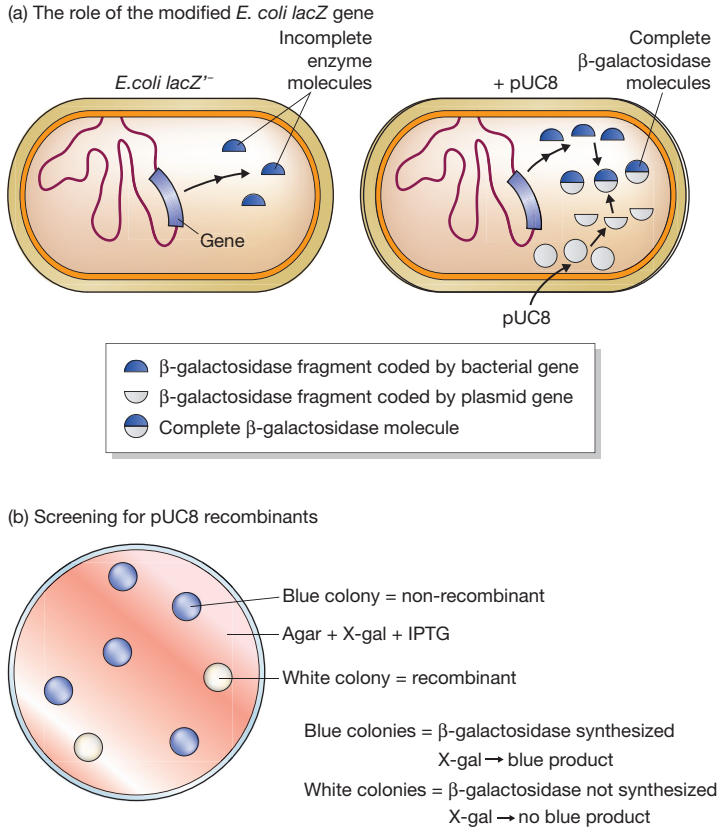


Figure 5.10

The rationale behind insertional inactivation of the *lacZ'* gene carried by pUC8. (a) The bacterial and plasmid genes complement each other to produce a functional β-galactosidase molecule. (b) Recombinants are screened by plating onto agar containing X-gal and IPTG.

useful if recombinant λ molecules could be packaged into their λ head-and-tail structures in the test tube.

This may sound difficult but it is actually relatively easy to achieve. Packaging requires a number of different proteins coded by the λ genome, but these can be prepared at a high concentration from cells infected with defective λ phage strains. Currently, two different systems are in use.

With the single strain system, the defective λ phage carries a mutation in the *cos* sites, so that these are not recognized by the endonuclease that normally cleaves the λ catenanes during phage replication (p. 21). This means that the defective phage cannot replicate, though it does direct the synthesis of all the proteins needed for packaging. These proteins accumulate in the bacterium and can be purified from cultures of *E. coli* infected with the mutated λ. The protein preparation is then used for *in vitro* packaging of recombinant λ molecules (Figure 5.11a).

With the second system, two defective λ strains are needed, both of which carry a mutation in a gene for one of the components of the phage protein coat. With one strain the mutation is in gene *D*, and with the second strain it is in gene *E* (see Figure 2.9).

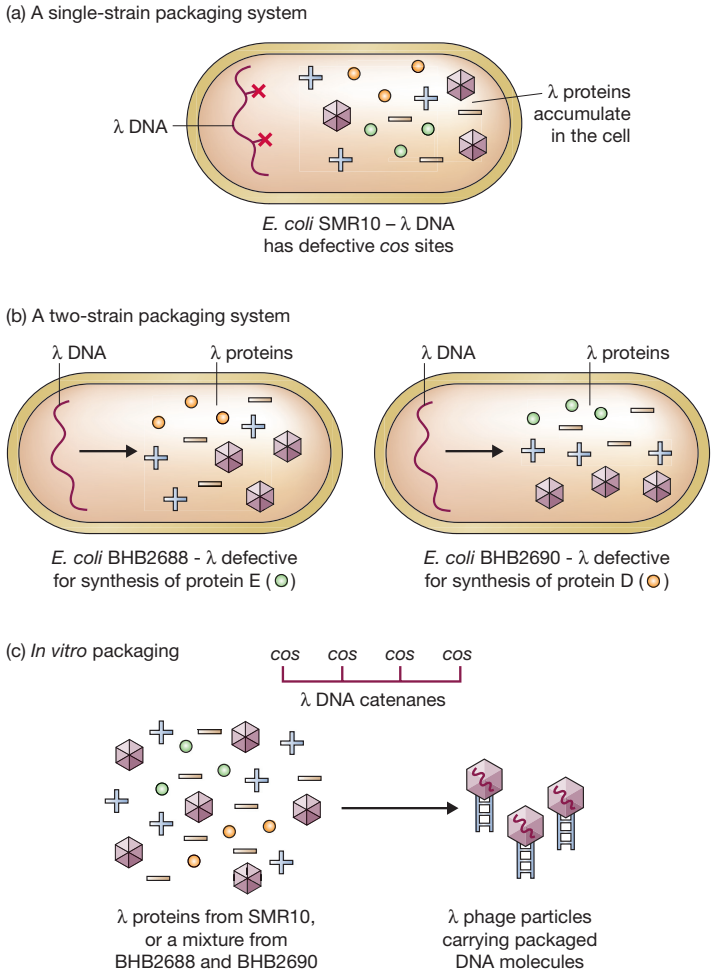


Figure 5.11

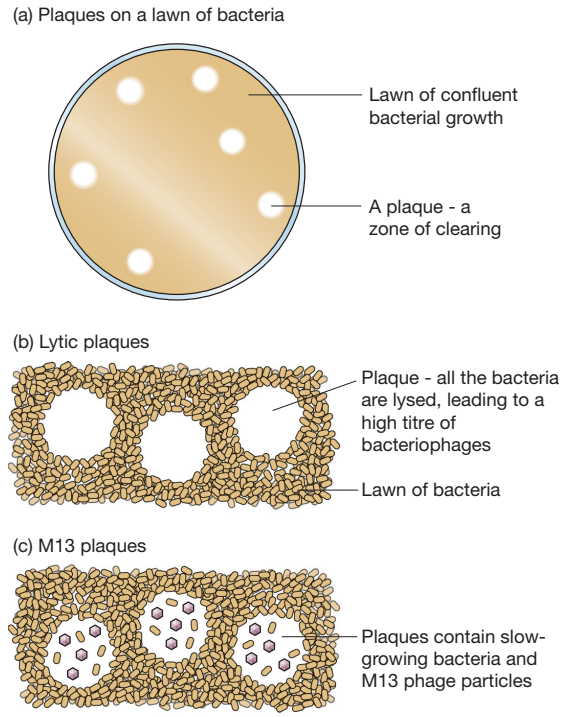
In vitro packaging. (a) Synthesis of λ capsid proteins by *E. coli* strain SMR10, which carries a λ phage that has defective *cos* sites. (b) Synthesis of incomplete sets of λ capsid proteins by *E. coli* strains BHB2688 and BHB2690. (c) The cell lysates provide the complete set of capsid proteins and can package λ DNA molecules in the test tube.

Neither strain is able to complete an infection cycle in *E. coli* because, in the absence of the product of the mutated gene, the complete capsid structure cannot be made. Instead, the products of all the other coat protein genes accumulate (Figure 5.11b). An *in vitro* packaging mix can therefore be prepared by combining lysates of two cultures of cells, one infected with the λ D^- strain, the other infected with the E^- strain. The mixture now contains all the necessary components for *in vitro* packaging.

With both systems, the formation of phage particles is achieved simply by mixing the packaging proteins with λ DNA, because assembly of the particles occurs automatically in the test tube (Figure 5.11c). The packaged λ DNA is then introduced into *E. coli* cells simply by adding the assembled phages to the bacterial culture and allowing the normal λ infective process to take place.

Figure 5.12

Bacteriophage plaques. (a) The appearance of plaques on a lawn of bacteria. (b) Plaques produced by a phage that lyses the host cell (e.g., λ in the lytic infection cycle); the plaques contain lysed cells plus many phage particles. (c) Plaques produced by M13; these plaques contain slow-growing bacteria plus many M13 phage particles.



5.3.3 Phage infection is visualized as plaques on an agar medium

The final stage of the phage infection cycle is cell lysis (p. 18). If infected cells are spread onto a solid agar medium immediately after addition of the phage particles, or immediately after transfection with phage DNA, cell lysis can be visualized as **plaques** on a lawn of bacteria (Figure 5.12a). Each plaque is a zone of clearing produced as the phages lyse the cells and move on to infect and eventually lyse the neighbouring bacteria (Figure 5.12b).

Both, λ and M13 form plaques. In the case of λ these are true plaques, produced by cell lysis, but M13 plaques are slightly different as M13 does not lyse the host cells (p. 19). Instead, M13 causes a decrease in the growth rate of infected cells, sufficient to produce a zone of relative clearing on a bacterial lawn. Although not true plaques, these zones of clearing are visually identical to normal phage plaques (Figure 5.12c).

The end result of a gene cloning experiment using a λ or M13 vector is therefore an agar plate covered in phage plaques, each of which is derived from a single transfected or infected cell and therefore contains identical phage particles. These may contain self-ligated vector molecules, or they may be recombinants.

5.4 Identification of recombinant phages

A variety of methods for distinguishing recombinant plaques have been devised, the following being the most important.

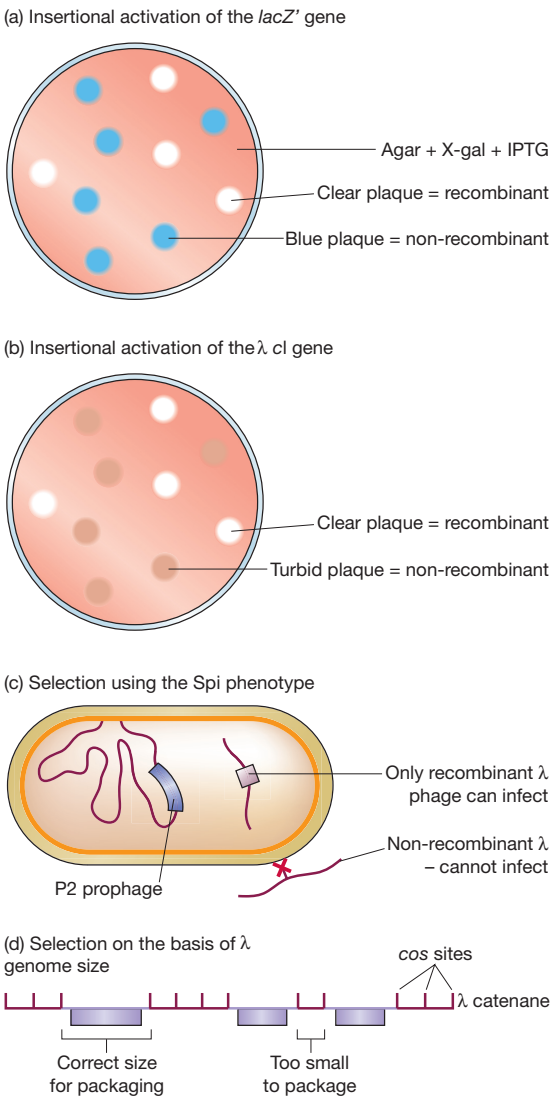


Figure 5.13
Strategies for the selection of recombinant phage.

5.4.1 Insertional inactivation of a lacZ′ gene carried by the phage vector

All M13 cloning vectors, as well as several λ vectors, carry a copy of the *lacZ′* gene. The insertion of new DNA into this gene inactivates β-galactosidase synthesis, just as with the plasmid vector pUC8. Recombinants are distinguished by plating cells onto X-gal agar. Plaques comprising normal phages are blue; recombinant plaques are clear (Figure 5.13a).

5.4.2 Insertional inactivation of the λ cI gene

Several types of λ cloning vector have unique restriction sites in the *c*I gene, the insertional inactivation of which causes a change in plaque morphology. Normal plaques

appear 'turbid', whereas recombinants with a disrupted *c*I gene are 'clear' (Figure 5.13b). The difference is readily apparent to the experienced eye.

5.4.3 Selection using the Spi phenotype

λ phages cannot normally infect *E. coli* cells that already possess an integrated form of a related phage called P2; λ is therefore said to be Spi^+ (sensitive to P2 prophage inhibition). Some λ cloning vectors are designed so that the insertion of new DNA causes a change from Spi^+ to Spi^-, enabling the recombinants to infect cells that carry P2 prophages. Such cells are used as the host for cloning experiments with these vectors. As only the recombinants are Spi^-, only recombinants will form plaques (Figure 5.13c).

5.4.4 Selection on the basis of λ genome size

The λ packaging system, which assembles the mature phage particles, can only insert DNA molecules of between 37 and 52 kb into the head structure; anything less than 37 kb is not packaged. Many λ vectors have been constructed by deleting large segments of the λ DNA molecule and so are less than 37 kb in length. These can only be packaged into mature phage particles after extra DNA has been inserted, bringing the total genome size up to 37 kb or more (Figure 5.13d). Therefore, with these vectors only recombinant phages are able to replicate.

5.5 Introduction of DNA into non-bacterial cells

Various ways of introducing DNA into yeast, fungi, animals and plants are also needed if these organisms are to be used as the hosts for gene cloning. Strictly speaking, these processes are not 'transformation', as that term has a specific meaning that applies only to the uptake of DNA by bacteria, but molecular biologists have forgotten this over the years and 'transformation' is now used to describe uptake of DNA by any organism.

In general terms, soaking cells in salt is effective only with a few species of bacteria, although treatment with lithium chloride or lithium acetate does enhance DNA uptake by yeast cells, and is frequently used in the transformation of *Saccharomyces cerevisiae*. However, for most higher organisms, more sophisticated methods are needed.

5.5.1 Transformation of individual cells

With most organisms the main barrier to DNA uptake is the cell wall. Cultured animal cells, which usually lack cell walls, are easily transformed, especially if the DNA is precipitated onto the cell surface with calcium phosphate (Figure 5.14a) or enclosed in **liposomes** that fuse with the cell membrane (Figure 5.14b). For other types of cell, the answer is often to remove the cell wall, and enzymes that degrade yeast, fungal and plant cell walls are currently available. Under the correct conditions, intact **protoplasts** can be obtained (Figure 5.14c) that generally take up DNA quite readily. Transformation can

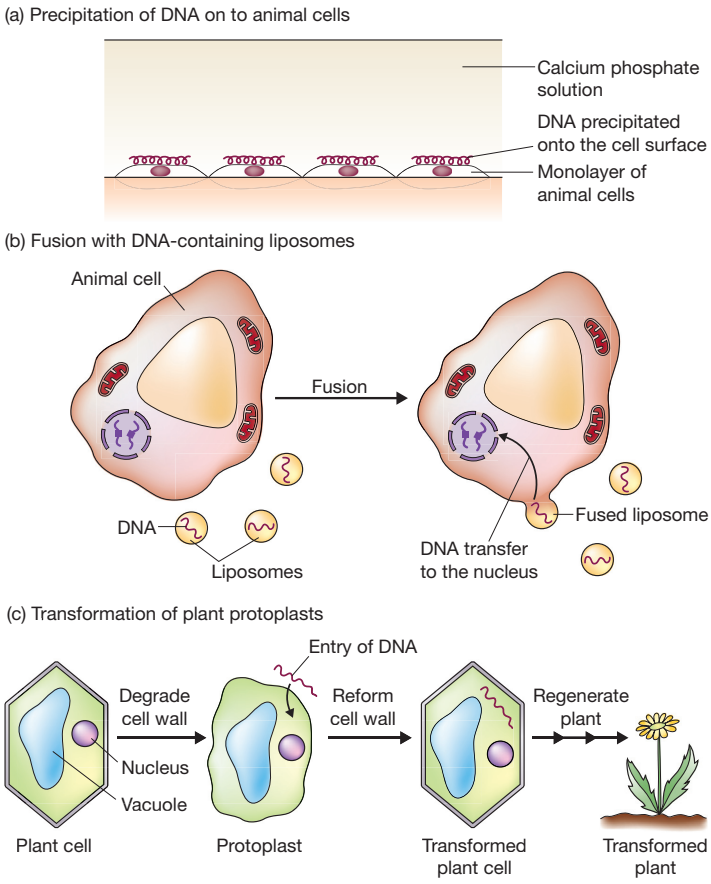


Figure 5.14

Strategies for introducing new DNA into animal and plant cells. (a) Precipitation of DNA onto animal cells. (b) Introduction of DNA into animal cells by liposome fusion. (c) Transformation of plant protoplasts.

be stimulated by applying specialized techniques such as **electroporation**, whereby the cells are subjected to a short electrical pulse that is thought to induce the transient formation of pores in the cell membrane, through which DNA molecules are able to enter the cell. After transformation, the protoplasts are washed to remove the degradative enzymes, after which their cell walls are re-formed spontaneously.

In contrast to the transformation systems described above, two physical methods are available for introducing DNA into cells. The first of these methods is **microinjection**, in which a very fine pipette is used to inject DNA molecules directly into the nucleus of the cells to be transformed (Figure 5.15a). This technique was initially applied to animal cells but has subsequently been used successfully with plant cells. The second method involves bombarding the cells with high-velocity microprojectiles, usually particles of gold or tungsten that have been coated with DNA. The microprojectiles are fired at the cells using a particle gun (Figure 5.15b). This unusual technique is called biolistics and has been used with several different types of cell.

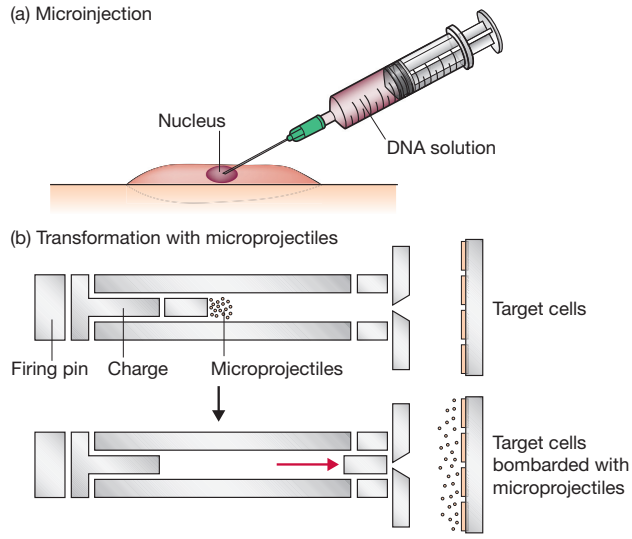


Figure 5.15
Two physical methods for introducing DNA into cells.

5.5.2 Transformation of whole organisms

With animals and plants, the desired end product might not be transformed cells but a transformed organism. Plants are relatively easy to regenerate from cultured cells, although problems have been experienced in developing regeneration procedures for monocotyledonous species such as cereals and grasses. A single transformed plant cell can therefore give rise to a transformed plant, which carries the cloned DNA in every cell, and passes the cloned DNA on to its progeny following flowering and seed formation. Animals, of course, cannot be regenerated from cultured cells, so the creation of transformed animals requires a rather more subtle approach. One possible technique that could be used in mammals such as mice would be to remove fertilized eggs from the oviduct, to microinject DNA, and then to reimplant the transformed cells into the mother's reproductive tract. We will look more closely at these methods for obtaining transformed animals in Chapter 13.

Further reading

FURTHER READING

Aich, P., Patra, M., Chatterjee, A.K. *et al.* (2012) Calcium chloride made *E. coli* competent for uptake of extraneous DNA through overproduction of OmpC protein. *Protein Journal*, **31**, 366–373. [Describes the effect of calcium chloride on *E. coli* cells and how this treatment might stimulate DNA uptake.]

Calvin, N.M. and Hanawalt, P.C. (1988) High efficiency transformation of bacterial cells by electroporation. *Journal of Bacteriology*, **170**, 2796–2801.

Capecchi, M.R. (1980) High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. *Cell*, **22**, 479–488.

Hammer, R.E., Pursel, V.G., Rexroad, C.E. *et al.* (1985) Production of transgenic rabbits, sheep and pigs by microinjection. *Nature*, **315**, 680–683.

Hohn, B. (1979) *In vitro* packaging of lambda and cosmid DNA. *Methods in Enzymology*, **68**, 299–309.

Klein, T.M., Wolf, E.D., Wu, R., and Sanford, J.C. (1987) High velocity microprojectiles for delivering nucleic acids into living cells. *Nature*, **327**, 70–73. [Biolistics.]

Lederberg, J. and Lederberg, E.M. (1952) Replica plating and indirect selection of bacterial mutants. *Journal of Bacteriology*, **63**, 399–406.

Mandel, M. and Higa, A. (1970) Calcium-dependent bacteriophage DNA infection. *Journal of Molecular Biology*, **53**, 159–162. [The first description of the use of calcium chloride to prepare competent *E. coli* cells.]

Chapter 6

Cloning Vectors for *Escherichia coli*

Chapter contents

CHAPTER CONTENTS

The basic experimental techniques involved in gene cloning have now been described. In Chapters 3, 4 and 5, we have seen how DNA is purified from cell extracts, how recombinant DNA molecules are constructed in the test tube, how DNA molecules are reintroduced into living cells, and how recombinant clones are distinguished. Now, we must look more closely at the cloning vector itself, in order to consider the range of vectors available to the molecular biologist, and to understand the properties and uses of each individual type.

The greatest variety of cloning vectors exists for use with *E. coli* as the host organism. This is not surprising in view of the central role that this bacterium has played in basic research over the past 50 years. The tremendous wealth of information that exists concerning the microbiology, biochemistry and genetics of *E. coli* has meant that virtually all fundamental studies of gene structure and function were initially carried out with this bacterium as the experimental organism. Even when a eukaryote is being studied, *E. coli* is still used for the construction of recombinant genes that will subsequently be placed back in the eukaryotic host in order to study their function and expression. During recent years, gene cloning and molecular biological research have become mutually synergistic: breakthroughs in gene cloning have acted as a stimulus to research, and the needs of research have spurred on the development of new, more sophisticated cloning vectors.

Gene Cloning and DNA Analysis: An Introduction, Seventh Edition. T.A. Brown.

Companion Website: www.wiley.com/go/brown/cloning

In this chapter the most important types of *E. coli* cloning vector will be described, and their specific uses outlined. Cloning vectors for yeast, fungi, plants and animals will be considered in Chapter 7.

6.1 Cloning vectors based on *E. coli* plasmids

The simplest cloning vectors, and those most widely used in gene cloning, are based on small bacterial plasmids. A large number of different plasmid vectors are available for use with *E. coli*, many of which are obtainable from commercial suppliers. These combine ease of purification with desirable properties such as high transformation efficiency, convenient selectable markers for transformants and recombinants, and the ability to clone reasonably large (up to about 8 kb) pieces of DNA. Most routine gene cloning experiments make use of one or other of these plasmid vectors.

One of the first vectors to be developed was pBR322, which was introduced in Chapter 5 to illustrate the general principles of transformant selection and recombinant identification (p. 79). Although pBR322 lacks the more sophisticated features of the newest cloning vectors, and consequently is no longer used extensively in research, it still illustrates very well the important and fundamental properties of any plasmid cloning vector. We will therefore begin our study of *E. coli* vectors by looking more closely at pBR322.

6.1.1 The nomenclature of plasmid cloning vectors

The name 'pBR322' conforms with the standard rules for vector nomenclature:

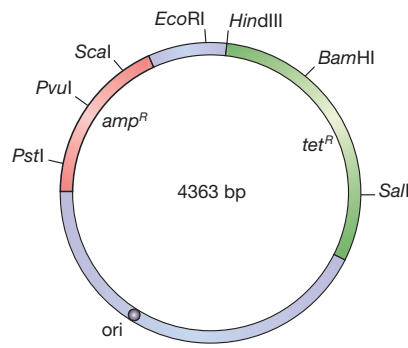
- 'p' indicates that this is indeed a plasmid.
- 'BR' identifies the laboratory in which the vector was originally constructed (BR stands for Bolivar and Rodriguez, the two investigators who developed pBR322).
- '322' distinguishes this plasmid from others developed in the same laboratory (there are also plasmids called pBR325, pBR327, pBR328, etc.).

6.1.2 The useful properties of pBR322

The genetic and physical map of pBR322 (Figure 6.1) gives an indication of why this plasmid was such a popular cloning vector.

Figure 6.1

A map of pBR322 showing the positions of the ampicillin resistance (amp^R) and tetracycline resistance (tet^R) genes, the origin of replication (ori) and some of the most important restriction sites.



The first useful feature of pBR322 is its size. In Chapter 2, it was stated that a cloning vector should be less than 10 kb in size in order to avoid problems such as DNA breakdown during purification. pBR322 is 4363 bp, which means that not only can the vector itself be purified with ease, but so too can any recombinant DNA molecules constructed with it. Even with 6 kb of additional DNA, a recombinant pBR322 molecule is still of a manageable size.

The second feature of pBR322 is that, as described in Chapter 5, it carries two sets of antibiotic resistance genes. Either ampicillin or tetracycline resistance can be used as a selectable marker for cells containing the plasmid, and each marker gene contains unique restriction sites that can be used in cloning experiments. The insertion of new DNA into pBR322 that has been restricted with *Pst*I, *Pvu*I or *Sca*I inactivates the amp^R gene, and insertion using any one of eight restriction endonucleases (notably *Bam*HI and *Hin*dIII) inactivates tetracycline resistance. This great variety of restriction sites that can be used for insertional inactivation means that pBR322 can be used to clone DNA fragments with any of several types of sticky end.

A third advantage of pBR322 is that it has a reasonably high copy number. Generally, there are about 15 molecules present in a transformed *E. coli* cell, but this number can be increased up to between 1000 and 3000 by plasmid amplification in the presence of a protein synthesis inhibitor such as chloramphenicol (p. 39). An *E. coli* culture therefore provides a good yield of recombinant pBR322 molecules.

6.1.3 The pedigree of pBR322

The remarkable convenience of pBR322 as a cloning vector did not arise by chance. The plasmid was in fact designed in such a way that the final construct would possess these desirable properties. An outline of the scheme used to construct pBR322 is shown in Figure 6.2a. It is clear that the production of pBR322 was a tortuous business that required the full and skilful use of the DNA manipulative techniques described in Chapter 4. A summary of the results of these manipulations is provided in Figure 6.2b, from which it can be seen that pBR322 comprises DNA derived from three different naturally occurring plasmids:

- The amp^R gene originally resided on the plasmid R1, a typical antibiotic resistance plasmid that occurs in natural populations of *E. coli* (p. 16).
- The tet^R gene is derived from R6-5, a second antibiotic resistance plasmid.
- The replication origin of pBR322, which directs multiplication of the vector in host cells, is originally from pMB1, which is closely related to the colicin-producing plasmid ColE1 (p. 17).

6.1.4 More sophisticated E. coli plasmid cloning vectors

pBR322 was developed during the late 1970s, with the first report describing its use appearing in 1977. Since then, many other plasmid cloning vectors have been constructed, the majority of these derived from pBR322 by manipulations similar to those summarized in Figure 6.2a. One of the first of these was pBR327, which was produced by removing a 1089 bp segment from pBR322. This deletion left the amp^R and tet^R genes intact, but changed the replicative and conjugative abilities of the resulting plasmid. This was an important step in the development of more sophisticated vectors, for two reasons:

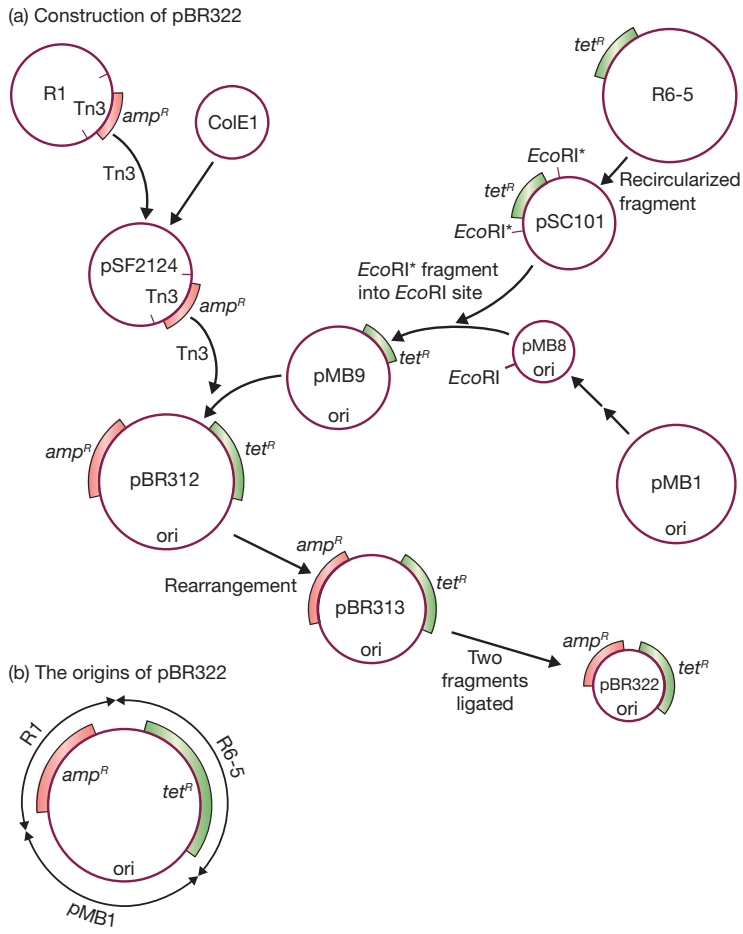


Figure 6.2

The pedigree of pBR322. (a) The manipulations involved in construction of pBR322. The *amp*^{R} gene was obtained from Tn3, a type of **transposable element** carried by the R1 plasmid. The *tet*^{R} gene was excised from pSC101 by treatment with *Eco*R1 in a low-salt solution, which decreases the specificity of the enzyme so that, as well as cutting at its standard GAATTC recognition sequence, it also cuts at related sequences such TAATTC. This is called **star activity**, and the related sequences are referred to as *Eco*RI* sites. (b) A summary of the origins of pBR322.

- The change in the replicative function means that pBR327 has a higher copy number than pBR322, being present at about 30–45 molecules per *E. coli* cell. This is not of great relevance as far as plasmid yield is concerned, as both plasmids can be amplified to copy numbers greater than 1000. The higher copy number of pBR327 in normal cells makes this vector more suitable if the aim of the experiment is to study the function of the cloned gene. In these cases, gene dosage becomes important because the more copies there are of a cloned gene the more likely it is that the effect of the cloned gene on the host cell will be detectable.
- The deletion also destroys the conjugative ability of pBR322, making pBR327 a non-conjugative plasmid that cannot direct its own transfer to other *E. coli* cells.

This is important for **biological containment**, averting the possibility of a recombinant pBR327 molecule escaping from the test tube and colonizing bacteria in the gut of a careless molecular biologist. In contrast, pBR322 could in theory be passed to natural populations of *E. coli* by conjugation, although in fact pBR322 also has safeguards (though less comprehensive ones) to minimize the chances of this happening.

Although pBR327, like pBR322, is no longer widely used, its properties have been inherited by most of today's modern plasmid vectors. There are a great number of these, and it would be pointless to attempt to describe them all. Two additional examples will suffice to illustrate the most important features.

pUC8: A Lac selection plasmid

This vector was mentioned in Chapter 5 when the identification of recombinants by insertional inactivation of the β-galactosidase gene was described (p. 81). pUC8 (Figure 6.3a) is descended from pBR322, although only the replication origin and the amp^R gene remain. The nucleotide sequence of the amp^R gene has been changed so that it no longer contains the unique restriction sites. All of these cloning sites are now clustered into a short segment of the *lacZ′* gene carried by pUC8.

pUC8 has two important advantages that have led to it becoming one of the most popular *E. coli* cloning vectors. The first of these was fortuitous. The manipulations

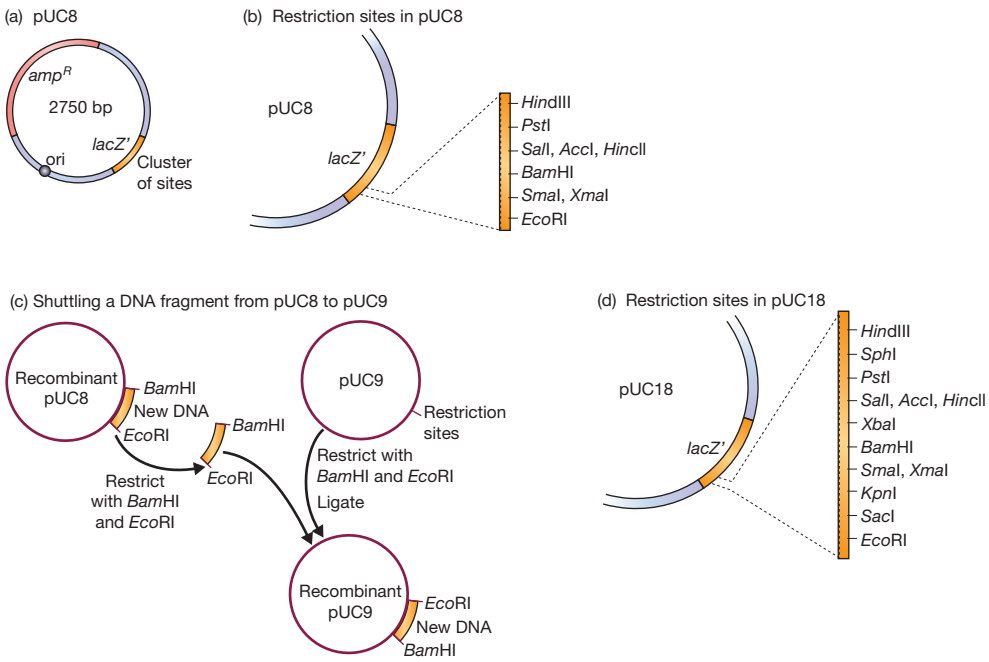


Figure 6.3

The pUC plasmids. (a) The structure of pUC8. (b) The restriction site cluster in the *lacZ′* gene of pUC8. (c) Shuttling a DNA fragment between pUC8 and pUC9, which enables the fragment to be cloned in two different directions. (d) The restriction site cluster in puC18.

involved in constructing pUC8 were accompanied by a chance mutation, within the origin of replication, which results in the plasmid having a copy number of 500–700, even before amplification. This has a significant effect on the yield of cloned DNA obtainable from *E. coli* cells transformed with recombinant pUC8 plasmids.

The second advantage is that identification of recombinant cells can be achieved by a single-step process, by plating onto agar medium containing ampicillin plus X-gal (p. 82). With both pBR322 and pBR327, selection of recombinants is a two-step procedure, requiring replica plating from one antibiotic medium to another (p. 80). A cloning experiment with pUC8 can therefore be carried out in half the time needed with pBR322 or pBR327.

The cluster of cloning sites are contained in a short artificial oligonucleotide, called a **polylinker**, which was inserted into the *lacZ′* gene when the first pUC8 plasmid was created. The polylinker is designed so that it does not totally disrupt the *lacZ′* gene, the reading frame being maintained throughout the polylinker, so that a functional (though altered) β-galactosidase enzyme is still produced. The polylinker in pUC8 contains nine unique restriction sites, ones that are not found elsewhere in the vector (Figure 6.3b). Because these sites are clustered, a DNA fragment with two different sticky ends (say *Eco*RI at one end and *Bam*HI at the other) can be cloned without resorting to additional manipulations such as linker attachment.

pUC8 is one of a family of vectors, differing only in the identity of the cloning sites. A second member of the family, pUC9, has the same polylinker as pUC8, but inserted into the *lacZ′* gene in the opposite orientation. This pair of vectors can therefore be used to clone a DNA fragment in both forward and reverse directions (Figure 6.3c), which enables **antisense RNA** to be prepared (p. 288). Other pUC vectors carry different combinations of restriction sites and provide even greater flexibility in the types of DNA fragment that can be cloned (Figure 6.3d).

pGEM3Z: In vitro *transcription of cloned DNA*

pGEM3Z (Figure 6.4a) is very similar to a pUC vector in that it carries the amp^R and *lacZ′* genes, the latter containing a cluster of restriction sites, and it is almost exactly the same size. The distinction is that pGEM3Z has two additional short pieces of DNA, each of which acts as the recognition site for attachment of an RNA polymerase enzyme. These two **promoter** sequences lie on either side of the cluster of restriction sites used for the introduction of new DNA into the pGEM3Z molecule. This means that if a recombinant pGEM3Z molecule is mixed with purified RNA polymerase in the test tube, transcription occurs and RNA copies of the cloned fragment are synthesized (Figure 6.4b). The RNA that is produced could be used as a hybridization probe (p. 143), or it might be required for experiments aimed at studying RNA processing (e.g., the removal of introns) or protein synthesis.

The promoters carried by pGEM3Z and other vectors of this type are not the standard sequences recognized by the *E. coli* RNA polymerase. Instead, one of the promoters is specific for the RNA polymerase coded by T7 bacteriophage and the other for the RNA polymerase of SP6 phage. These RNA polymerases are synthesized during the infection of *E. coli* with one or other of the phages, and are responsible for transcribing the phage genes. They are chosen for *in vitro* transcription as they are very active enzymes (remember that the entire lytic infection cycle takes only 20 min – see p. 18 – so the phage genes must be transcribed very quickly). These polymerases are

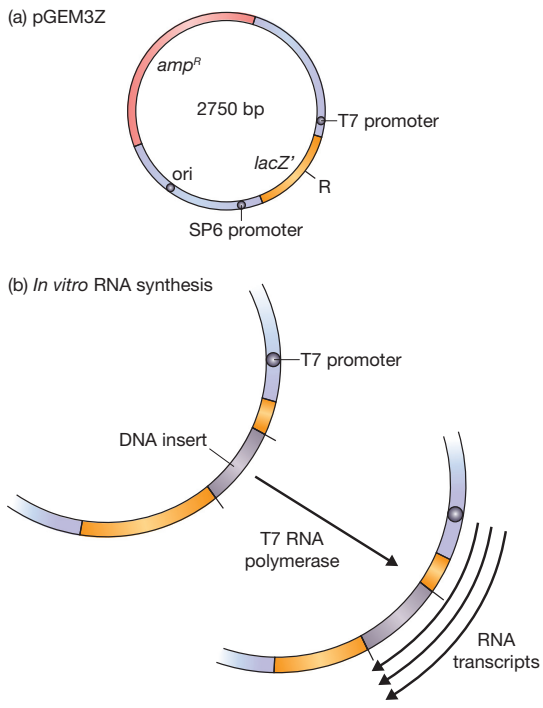


Figure 6.4
pGEM3Z. (a) Map of the vector. 'R' indicates a polylinker containing restriction sites for *Eco*RI, *Sac*I, *Kpn*I, *Ava*I, *Sma*I, *Bam*HI, *Xba*I, *Sal*I, *Acc*I, *Hin*cII, *Pst*I, *Sph*I, and *Hin*dIII. (b) *In vitro* RNA synthesis from the T7 promoter.

able to synthesize 1–2 mg of RNA per minute, which is substantially more than can be produced by the standard *E. coli* enzyme.

6.2 Cloning vectors based on λ bacteriophage

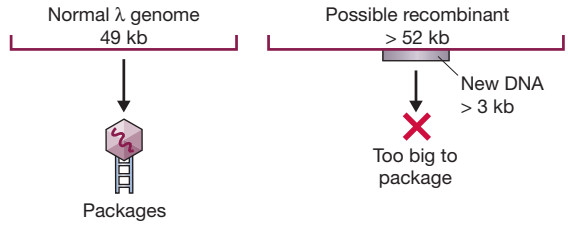
As well as plasmid cloning vectors, a variety of vectors based on λ bacteriophage have been developed for cloning DNA in *E. coli*. The primary use of these vectors is to clone large pieces of DNA, from 5 to 25 kb, most of which are too large to be handled by plasmid vectors.

6.2.1 Segments of the λ genome can be deleted without impairing viability

The first problem that had to be solved when λ cloning vectors were first developed was the capacity of the phage capsid. The λ DNA molecule can be increased in size by only about 5%, representing the addition of only 3 kb of new DNA. If the total size of the molecule is more than 52 kb, then it cannot be packaged into the λ head structure and infective phage particles are not formed. This severely limits the size of a DNA fragment that can be inserted into an unmodified λ vector (Figure 6.5).

The way forward for the development of λ cloning vectors was provided by the discovery that a large segment in the central region of the λ DNA molecule can be removed without affecting the ability of the phage to infect *E. coli* cells and produce new phage particles. A means of replicating within the host cell is, of course, the most essential

Figure 6.5
The size limitation placed on the λ genome by the need to package it into the phage head.



requirement for any cloning vector. For plasmid vectors this requirement is easy to satisfy, as relatively short DNA sequences are able to act as plasmid origins of replication, and most – if not all – of the enzymes needed for replication are provided by the host cell. Elaborate manipulations, such as those that resulted in pBR322 (see Figure 6.2a), are therefore possible as long as the final construction has an intact, functional replication origin. With a bacteriophage, the situation regarding replication is more complex. Phage DNA molecules generally carry several genes that are essential for replication, including genes coding for components of the phage protein coat and phage-specific DNA replicative enzymes. The alteration or deletion of any of these genes will impair or destroy the replicative ability of the resulting molecule.

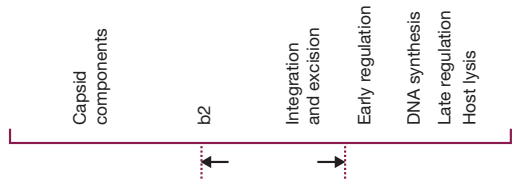
The segment that can be removed from the λ DNA molecule is called the 'non-essential' region. This contains most of the genes involved in the integration and excision of the λ prophage from the *E. coli* chromosome. Removal of this segment therefore destroys the ability of the phage to follow its lysogenic infection cycle, but the lytic cycle is unaffected. This in itself is desirable for a cloning vector as it means that induction is not needed before plaques are formed (p. 41).

The non-essential region lies between positions 20 and 35 on the map shown in Figure 2.9. Its removal decreases the size of the resulting λ molecule by up to 15 kb, which means that as much as 18 kb of new DNA can now be added before the cut-off point for packaging is reached (Figure 6.6).

6.2.2 Natural selection was used to isolate modified λ that lack certain restriction sites

A second essential requirement for a cloning vector is the presence of restriction sites into which new DNA can be cloned. Unfortunately, the λ genome is so large that it has more than one recognition sequence for virtually every restriction endonuclease. Consequently, restriction cannot be used to cleave the normal λ molecule in a way that will allow the insertion of new DNA, because the molecule would be cut into several small fragments that would be very unlikely to re-form a viable λ genome on religation (Figure 6.7). Even a deleted λ genome, with the non-essential region removed, has multiple recognition sites for most restriction endonucleases. This is a problem that is often

Figure 6.6
The λ genetic map, showing the position of the main non-essential region that can be deleted without affecting the ability of the phage to follow the lytic infection cycle. There are other, much shorter non-essential regions in other parts of the genome.



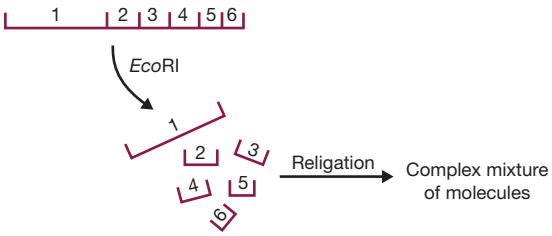


Figure 6.7
λ DNA has multiple recognition sites for almost all restriction endonucleases.

encountered when a new vector is being developed. If just one or two sites need to be removed, then *in vitro* mutagenesis (p. 216) can be used to change their sequences. For example, an *Eco*RI site, GAATTC, could be changed to GGATTC, which is not recognized by the enzyme. However, *in vitro* mutagenesis was in its infancy when the first λ vectors were under development, and even today it would not be an efficient means of changing more than a few sites in a single molecule.

Instead, natural selection was used to provide strains of λ that lack the unwanted restriction sites. Natural selection can be brought into play by using as a host an *E. coli* strain that produces *Eco*RI. Most λ DNA molecules that invade the cell are destroyed by this restriction endonuclease, but a few survive and produce plaques. These are mutant phages, from which one or more *Eco*RI sites have been lost spontaneously (Figure 6.8).

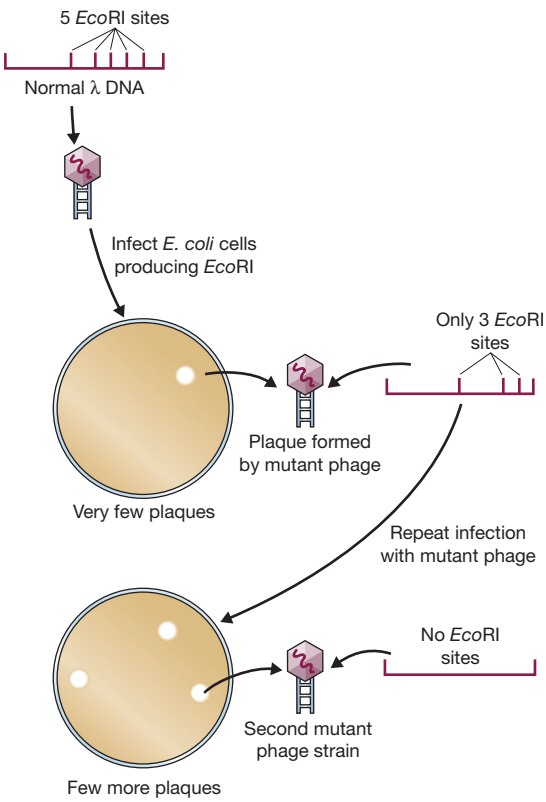
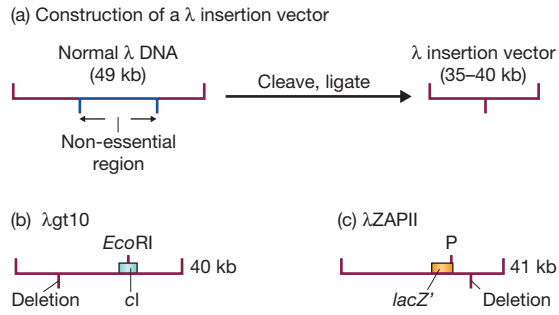


Figure 6.8
Using natural selection to isolate λ phage lacking *Eco*RI restriction sites.

Figure 6.9

λ insertion vectors. P = polylinker in the *lacZ′* gene of λZAPII, containing unique restriction sites for *Sac*I, *Not*I, *Xba*I, *Spe*I, *Eco*RI, and *Xho*I.



Several cycles of infection will eventually result in λ molecules that lack all or most of the *Eco*RI sites.

6.2.3 Insertion and replacement vectors

Once the problems posed by packaging constraints and by the multiple restriction sites had been solved, the way was open for the development of different types of λ-based cloning vectors. The first two classes of vector to be produced were λ **insertion** and λ **replacement** (or substitution) vectors.

Insertion vectors

With an insertion vector (Figure 6.9a), a large segment of the non-essential region has been deleted, and the two arms ligated together. An insertion vector possesses at least one unique restriction site into which new DNA can be inserted. The size of the DNA fragment that an individual vector can carry depends, of course, on the extent to which the non-essential region has been deleted. Two popular insertion vectors are:

- **λgt10** (Figure 6.9b), which can carry up to 8 kb of new DNA, inserted into a unique *Eco*RI site located in the *c*I gene. Insertional inactivation of this gene means that recombinants are distinguished as clear rather than turbid plaques (p. 87).
- **λZAPII** (Figure 6.9c), with which insertion of up to 10 kb DNA into any of six restriction sites within a polylinker inactivates the *lacZ′* gene carried by the vector. Recombinants give clear rather than blue plaques on X-gal agar.

Replacement vectors

A λ replacement vector has two recognition sites for the restriction endonuclease used for cloning. These sites flank a segment of DNA that is replaced by the DNA to be cloned (Figure 6.10a). Often, the replaceable fragment (called the **stuffer fragment**) carries additional restriction sites that can be used to cut it up into small pieces, so that its own re-insertion during a cloning experiment is very unlikely. Replacement vectors are generally designed to carry larger pieces of DNA than insertion vectors can handle. Recombinant selection is often on the basis of size, with non-recombinant vectors being too small to be packaged into λ phage heads (p. 88).

An example of a replacement vector is:

- **λDASHII** (Figure 6.10b), which can carry inserted DNA of between 9 and 23 kb by replacing a segment flanked by various restriction sites, any of which can be used to remove the stuffer fragment, so that DNA fragments with a variety of sticky ends

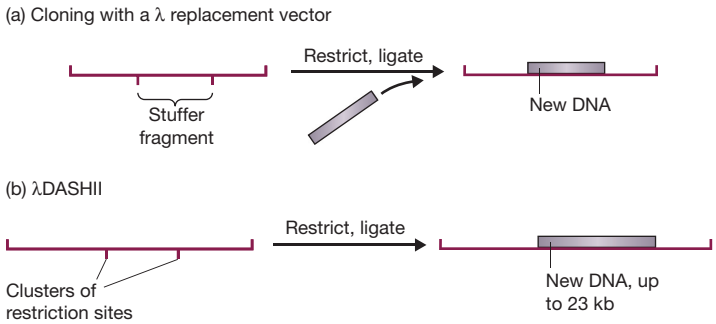


Figure 6.10

λ replacement vectors. (a) Cloning with a λ replacement vector. (b) Cloning with λDASHII.

can be cloned. Recombinant selection with λDASHII can be on the basis of size, or can utilize the Spi phenotype (p. 88). As the inserted DNA is flanked by the promoter sequences of T3 and T7 phage, RNA copies can be obtained in the same way as for pGEM3Z.

6.2.4 Cloning experiments with λ insertion or replacement vectors

A cloning experiment with a λ vector can proceed along the same lines as with a plasmid vector: the λ molecules are restricted, new DNA is added, the mixture is ligated, and the resulting molecules are used to transfect a competent *E. coli* host (Figure 6.11a). This type of experiment requires that the vector be in its circular form, with the *cos* sites hydrogen-bonded to each other.

Although satisfactory for many purposes, a procedure based on transfection is not particularly efficient, and a greater number of recombinants will be obtained if one or two refinements are introduced. The first improvement would be to use the linear form of the vector. When the linear form of the vector is digested with the relevant restriction endonuclease, the left and right arms are released as separate fragments. A recombinant molecule can be constructed by mixing together the DNA to be cloned with the vector arms. Ligation results in several molecular arrangements, including catenanes comprising left arm–DNA–right arm repeated many times (Figure 6.11b). If the inserted DNA is the correct size, the *cos* sites that separate these structures will be the correct distance apart for *in vitro* packaging (p. 83). In this way, recombinant phage are produced in the test tube and can be used to infect an *E. coli* culture. This strategy – and in particular the use of *in vitro* packaging – results in a large number of recombinant plaques.

6.2.5 Long DNA fragments can be cloned using a cosmid

The final and most sophisticated type of λ-based vector is the **cosmid**. Cosmids are hybrids between a phage DNA molecule and a bacterial plasmid, and their design centres on the fact that the enzymes which package the λ DNA molecule into the phage protein coat need only the *cos* sites in order to function (p. 20). The *in vitro* packaging reaction works not only with λ genomes, but also with any molecule that carries *cos* sites separated by 37–52 kb of DNA.

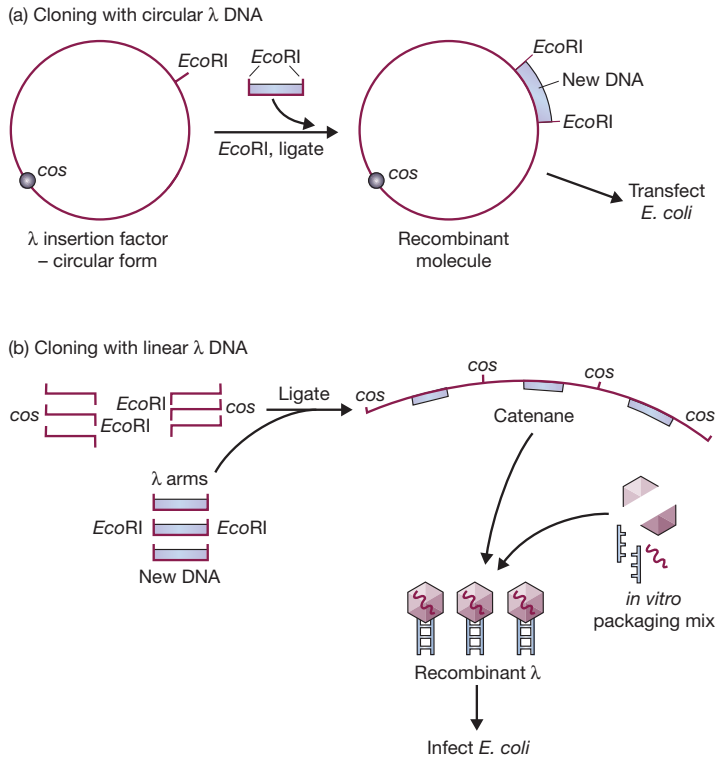


Figure 6.11

Different strategies for cloning with a λ vector. (a) Using the circular form of λ as a plasmid. (b) Using the left and right arms of the λ vector, plus *in vitro* packaging, to achieve a greater number of recombinant plaques.

A cosmid is basically a plasmid that carries a *cos* site (Figure 6.12a). It also needs a selectable marker (such as the ampicillin resistance gene) and a plasmid origin of replication, since cosmids lack all the λ genes and so do not produce plaques. Instead, colonies are formed on selective media, just as with a plasmid vector.

A cloning experiment with a cosmid is carried out as follows (Figure 6.12b). The cosmid is opened at its unique restriction site and new DNA fragments are inserted. These fragments are usually produced by partial digestion with a restriction endonuclease, as total digestion will almost invariably result in fragments that are too small to be cloned with a cosmid. Ligation is carried out so that catenanes are formed. Provided that the inserted DNA is the correct size, *in vitro* packaging cleaves the *cos* sites and places the recombinant cosmids into mature phage particles. These λ phage are then used to infect an *E. coli* culture, although of course plaques are not formed. Instead, infected cells are plated onto a selective medium and antibiotic-resistant colonies are grown. All colonies are recombinants, as non-recombinant linear cosmids are too small to be packaged into λ heads.

6.2.6 λ and other high-capacity vectors enable genomic libraries to be constructed

The main use of λ-based vectors is to clone DNA fragments that are too long to be handled by plasmid vectors. A replacement vector, such as λDASHII, can carry up to 23 kb

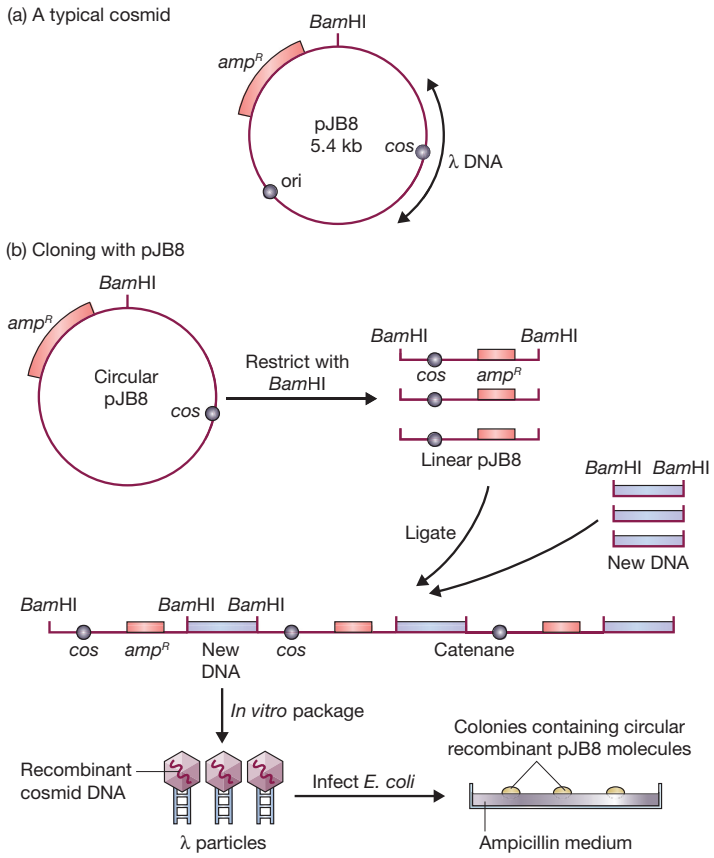


Figure 6.12

A typical cosmid and the way it is used to clone long fragments of DNA.

of new DNA, and some cosmids can manage fragments up to 40 kb. This compares with a maximum insert size of about 8 kb for most plasmids.

The ability to clone such long DNA fragments means that genomic libraries can be generated. A genomic library is a set of recombinant clones that contains all of the DNA present in an individual organism. An *E. coli* genomic library, for example, contains all the *E. coli* genes, so any desired gene can be withdrawn from the library and studied. Genomic libraries can be retained for many years, and propagated so that copies can be sent from one research group to another.

The big question is how many clones are needed for a genomic library? The answer can be calculated with the formula:

$$N = \frac{\ln(1-p)}{\ln\left(1-\frac{a}{b}\right)}$$

where N is the number of clones that are required, p is probability that any given gene will be present, a is the average size of the DNA fragments inserted into the vector, and b is the total size of the genome.

Table 6.1
Number of clones needed for genomic libraries of a variety of organisms.

SPECIES	GENOME SIZE (bp)	NUMBER OF CLONES[a]	
		17 kb FRAGMENTS[b]	35 kb FRAGMENTS[c]
E. coli	4.6×10^6	820	410
Saccharomyces cerevisiae	1.8×10^7	3225	1500
Drosophila melanogaster	1.2×10^8	21 500	10 000
Rice	5.7×10^8	100 000	49 000
Human	3.2×10^9	564 000	274 000
Frog	2.3×10^{10}	4 053 000	1 969 000

[a] Calculated for a probability (p) of 95% that any particular gene will be present in the library.
[b] Fragments suitable for a replacement vector such as λDASHII.
[c] Fragments suitable for a cosmid.

Table 6.1 shows the number of clones needed for genomic libraries of a variety of organisms, constructed using a λ replacement vector or a cosmid. For humans and other mammals, several hundred thousand clones are required. It is by no means impossible to obtain several hundred thousand clones, and the methods used to identify a clone carrying a desired gene (see Chapter 8) can be adapted to handle such large numbers, so genomic libraries of these sizes are by no means unreasonable. Nonetheless, ways of reducing the numbers of clones needed for mammalian genomic libraries are continually being sought.

One solution is to develop new cloning vectors capable of handling longer DNA inserts. The most popular of these vectors are **bacterial artificial chromosomes (BACs)**, which are based on the F plasmid (p. 16). The F plasmid is relatively large, and vectors derived from it have a higher capacity than normal plasmid vectors. BACs can handle DNA inserts up to 300 kb in size, reducing the size of the human genomic library to just 30 000 clones. Other high-capacity vectors have been constructed from bacteriophage P1, which has the advantage over λ of being able to squeeze 110 kb of DNA into its capsid structure. Cosmid-type vectors based on P1 have been designed and used to clone DNA fragments ranging in size from 75 to 100 kb. Vectors that combine the features of P1 vectors and BACs, called **P1-derived artificial chromosomes (PACs)**, also have a capacity of up to 300 kb.

6.3 Cloning vectors for the synthesis of single-stranded DNA

Cloning vectors capable of providing single-stranded versions of cloned DNA were extremely important for many years, as the DNA sequencing methods used during most of the 1980s and 1990s required single-stranded DNA as the starting material. Although those sequencing methods have now been superseded by techniques that use double-stranded DNA, the special vectors designed during the 1980s and 1990s for the synthesis of single-stranded DNA are still important because this type of DNA is also needed for specialist applications such as *in vitro* mutagenesis (p. 216) and phage display (p. 241).

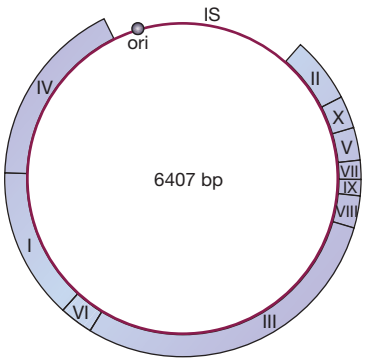


Figure 6.13
The M13 genome, showing the positions of genes I to X.

6.3.1 Vectors based on M13 bacteriophage

The first single-stranded DNA vectors were based on M13 bacteriophage. The normal M13 genome is 6.4 kb in length, with most of the DNA taken up by ten closely packed genes (Figure 6.13), each essential for the replication of the phage. There is only a single 507-nucleotide intergenic sequence into which new DNA could be inserted without disrupting one of these genes, and this region includes the replication origin which must itself remain intact. This means that there is only limited scope for modifying the M13 genome.

The first step in the construction of an M13 cloning vector was to introduce the *lacZ′* gene into the intergenic sequence. This gave rise to M13mp1, which forms blue plaques on X-gal agar (Figure 6.14a). Polylinkers were then inserted into the *lacZ′* gene to create a series of M13 vectors with different sets of cloning sites (Figure 6.14b). The polylinkers are the same as those used in the pUC plasmids, which means that the cloned DNA can be shuttled between the equivalent M13 and pUC vectors, so that single- and double-stranded versions can be obtained.

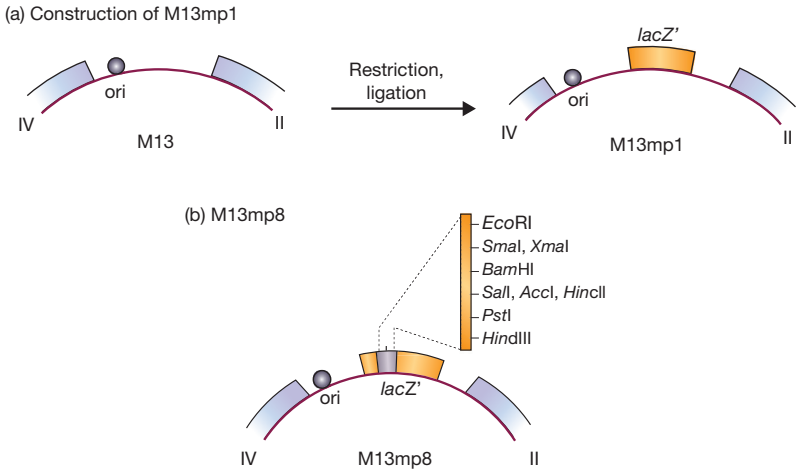


Figure 6.14
M13 vectors. (a) Construction of M13mp1. (b) M13mp8, which contains the same set of restriction sites as pUC8.

A cloning experiment with an M13 vector simply involves inserting the new DNA into one of the restriction sites in the polylinker, and then transfecting competent *E. coli* cells. After plating out on X-gal agar, recombinant plaques are identified by the blue-white colour test.

6.3.2 Hybrid plasmid–M13 vectors

Although M13 vectors are very useful for the production of single-stranded versions of cloned genes, they do suffer one major disadvantage. There is a limit to the size of DNA fragment which can be cloned with an M13 vector. Typically, 1500 bp is considered to be the maximum capacity, although fragments of up to 3 kb have occasionally been cloned. To overcome this problem, a number of hybrid vectors (**phagemids**) have been developed by combining a part of the M13 genome with plasmid DNA.

An example is provided by pEMBL8 (Figure 6.15a), which was made by transferring into pUC8 a 1300 bp fragment of the M13 genome. This piece of M13 DNA

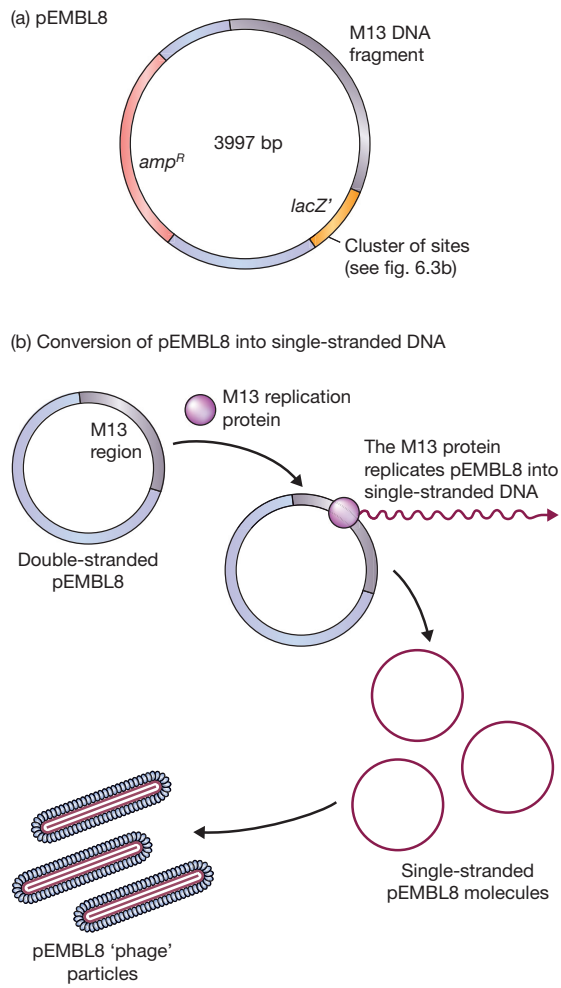


Figure 6.15
pEMBL8: a hybrid plasmid–M13 vector that can be converted into single-stranded DNA.

contains the signal sequence recognized by the enzymes that convert the normal double-stranded M13 molecule into single-stranded DNA before the secretion of new phage particles. This signal sequence is still functional even though detached from the rest of the M13 genome, so pEMBL8 molecules are also converted into single-stranded DNA and secreted as defective phage particles (Figure 6.15b). All that is necessary is that the *E. coli* cells used as hosts for a pEMBL8 cloning experiment are subsequently infected with normal M13 to act as a **helper phage**, providing the necessary replicative enzymes and phage coat proteins. pEMBL8, being derived from pUC8, has the polylinker cloning sites within the *lacZ′* gene, so recombinant plaques can be identified in the standard way on agar containing X-gal. With pEMBL8, single-stranded versions of cloned DNA fragments up to 10 kb in length can be obtained, greatly extending the range of the M13 cloning system.

6.4 Vectors for other bacteria

Cloning vectors have also been developed for several other species of bacteria, including *Streptomyces*, *Bacillus*, and *Pseudomonas*. These vectors are used for basic studies of gene function and activity in these species, and also for the production of **recombinant protein**. A recombinant protein is a protein synthesized from a cloned gene, with cloning being used in this way to obtain large quantities of important proteins (e.g., insulin) that are used as pharmaceuticals to treat human disorders (see Chapter 13). As some recombinant proteins cannot be synthesized efficiently in *E. coli*, other species – both bacterial and eukaryotic – are also used for recombinant protein synthesis (p. 259).

Some cloning vectors for these other types of bacteria are based on plasmids specific to the host species, and some on **broad host range plasmids** that are able to replicate in a variety of bacteria. The origin of replication of the RK2 plasmid, for example, functions in most Gram-negative bacteria, and is used in cloning vectors that replicate in *Pseudomonas* species as well as *E. coli*. A cloning vector that carries the RK2 origin is therefore a type of **shuttle vector**, one that can be used in two different species. Shuttle vectors are useful because the initial stages of a gene cloning experiment, up to the point when the correct recombinant DNA molecule has been identified, are usually easier to carry out with *E. coli* as the host. Once the desired clone has been obtained, the broad host range properties of the shuttle vector can then be exploited to transfer the molecule to the second species.

Bacteriophages have also been used as cloning vectors for bacteria other than *E. coli*. An example is the SV1 phage of *Streptomyces venezuelae*, which has been developed into a set of cloning vectors for *Streptomyces* species. SV1 is a lysogenic phage, but rather than deleting the genes responsible for the lysogenic infection cycle, vectors based on SV1 retain this function so that the cloned gene is inserted into the *Streptomyces* genome. This results in a very stable transformant, with little possibility of the cloned gene being lost from the host even after many cycles of bacterial replication. Stability is an important consideration if the aim of the gene cloning project is to obtain a recombinant bacterial strain that creates a valuable pharmaceutical protein (p. 269).

Further reading

FURTHER READING

Bolivar, F., Rodriguez, R.L., Green, P.J., *et al*. (1977) Construction and characterization of new cloning vectors. II. A multipurpose cloning system. *Gene*, **2**, 95–113. [pBR322.]

Choi, K.-H., Trunck, L.A., Kumar, A., *et al*. (2008) Genetic tools for *Pseudomonas*, in *Pseudomonas Genomics and Molecular Biology* (ed. P. Cornells), Caister Academic Press, Norfolk, pp. 65–86. [Describes cloning vectors for *Pseudomonas*.]

Fayed, B., Younger, E., Taylor, G., and Smith, M.C.M. (2014) A novel *Streptomyces* spp. integration vector, derived from the *S. venezuelae* phage, SV1. *BMC Biotechnology*, **14**, 51.

Iouannou, P.A., Amemiya, C.T., Garnes, J., *et al*. (1994) P1-derived vector for the propagation of large human DNA fragments. *Nature Genetics*, **6**, 84–89.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Research*, **12**, 7035–7056. [RNA synthesis from DNA cloned in a plasmid such as pGEM3Z.]

Sanger, F., Coulson, A.R., Barrell, B.G., *et al*. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology*, **143**, 161–178. [M13 vectors.]

Shiyuza, H., Birren, B., Kim, U.J., *et al*. (1992) Cloning and stable maintenance of 300 kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 8794–8797. [The first description of a BAC.]

Sternberg, N. (1992) Cloning high molecular weight DNA fragments by the bacteriophage P1 system. *Trends in Genetics*, **8**, 11–16.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103–119.

Chapter 7

Cloning Vectors for Eukaryotes

Chapter contents

CHAPTER CONTENTS

Most cloning experiments are carried out with *E. coli* as the host, and the widest variety of cloning vectors are available for this organism. *E. coli* is particularly popular when the aim of the cloning experiment is to study the basic features of molecular biology such as gene structure and function. However, under some circumstances it may be desirable to use a eukaryotic host for a gene cloning experiment. This is especially true in biotechnology (Chapter 13), where the aim may not be to study a gene, but to use cloning to obtain recombinant protein, or to change the properties of an organism (e.g., to introduce herbicide resistance into a crop plant). We must therefore consider cloning vectors for eukaryotes.

7.1 Vectors for yeast and other fungi

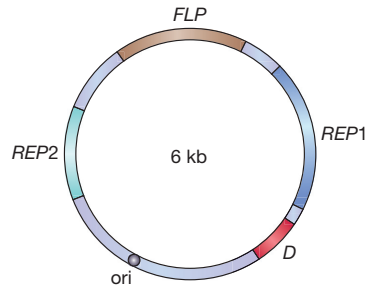
The yeast *Saccharomyces cerevisiae* is one of the most important organisms in biotechnology. As well as its role in brewing and breadmaking, yeast is used as a host organism for the production of important pharmaceuticals from cloned genes (p. 260). The development of cloning vectors for yeast was initially stimulated by the discovery of a plasmid that is present in most strains of *S. cerevisiae* (Figure 7.1). The **2 μm plasmid**, as it is called, is one of only a very limited number of plasmids found in eukaryotic cells.

Gene Cloning and DNA Analysis: An Introduction, Seventh Edition. T.A. Brown.

Companion Website: www.wiley.com/go/brown/cloning

Figure 7.1

The yeast 2 μm plasmid. *REP1* and *REP2* are involved in replication of the plasmid, and *FLP* codes for a protein that can convert the A form of the plasmid (shown here) to the B form, in which the gene order has been rearranged by intramolecular recombination. The function of gene *D* is not exactly known.



7.1.1 Selectable markers for the 2 μm plasmid

The 2 μm plasmid is an excellent basis for a cloning vector. It is 6 kb in size, which is ideal for a vector, and exists in the yeast cell at a copy number of between 70 and 200. Replication makes use of a plasmid origin, several enzymes provided by the host cell, and the proteins coded by the *REP1* and *REP2* genes carried by the plasmid.

However, all is not perfectly straightforward in using the 2 μm plasmid as a cloning vector. First, there is the question of a selectable marker. Some yeast cloning vectors carry genes conferring resistance to inhibitors such as methotrexate and copper, but most of the popular yeast vectors make use of a radically different type of selection system. In practice, a normal yeast gene is used, generally one that codes for an enzyme involved in amino acid biosynthesis. An example is the gene *LEU2*, which codes for β-isopropylmalate dehydrogenase, one of the enzymes involved in the conversion of pyruvic acid to leucine.

In order to use *LEU2* as a selectable marker, a special kind of host organism is needed. The host must be an **auxotrophic** mutant that has a non-functional *LEU2* gene. Such a *leu2*$^-$ yeast is unable to synthesize leucine and can survive only if this amino acid is supplied as a nutrient in the growth medium (Figure 7.2a). Selection is possible because transformants contain a plasmid-borne copy of the *LEU2* gene, and so are able to grow in the absence of the amino acid. In a cloning experiment, cells are plated out onto **minimal medium**, which contains no added amino acids. Only transformed cells are able to survive and form colonies (Figure 7.2b).

7.1.2 Vectors based on the 2 μm plasmid: Yeast episomal plasmids

Vectors derived from the 2 μm plasmid are called **yeast episomal plasmids (YEps)**. Some YEps contain the entire 2 μm plasmid while others include just the 2 μm origin of replication. An example of the latter type is YEp13 (Figure 7.3).

As well as the 2 μm origin of replication and the selectable *LEU2* gene, YEp13 also includes the entire pBR322 sequence. Yep13 is therefore a shuttle vector, and can replicate and be selected for in both yeast and *E. coli*. Most yeast cloning vectors are shuttle vectors because it is sometimes difficult to recover the recombinant DNA molecule from a transformed yeast colony. This is not a major problem with YEps as they are present in yeast cells primarily as plasmids, but with other yeast vectors that may integrate into one of the yeast chromosomes, purification might be impossible. This is a disadvantage because in many cloning experiments the purification of recombinant DNA

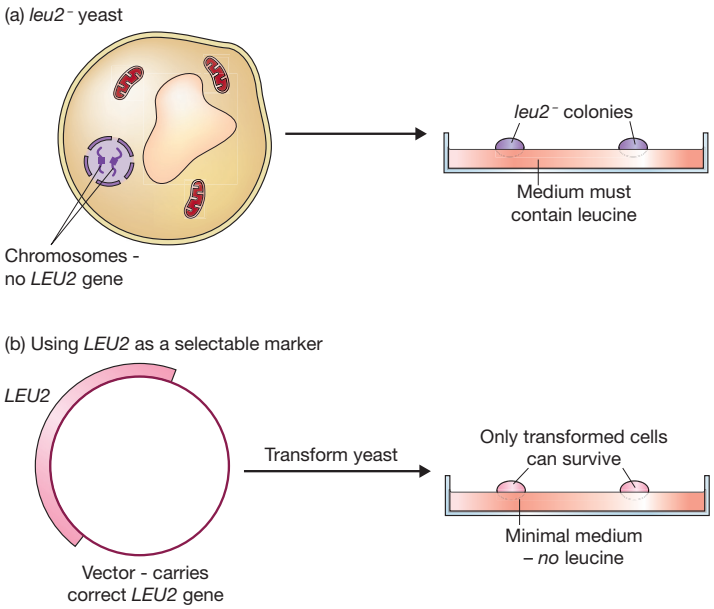


Figure 7.2

Using the *LEU2* gene as a selectable marker in a yeast cloning experiment.

is essential in order for the correct construct to be identified by, for example, DNA sequencing.

The standard procedure when cloning in yeast is therefore to perform the initial cloning experiment with *E. coli*, and to select recombinants in this organism. Recombinant plasmids can then be purified, characterized, and the correct molecule introduced into yeast (Figure 7.4).

7.1.3 A YEp may insert into yeast chromosomal DNA

The word 'episomal' indicates that a YEp can replicate as an independent plasmid, but also implies that the integration into one of the yeast chromosomes can occur (see the definition of 'episome' on p. 14). Integration occurs because the gene carried on the

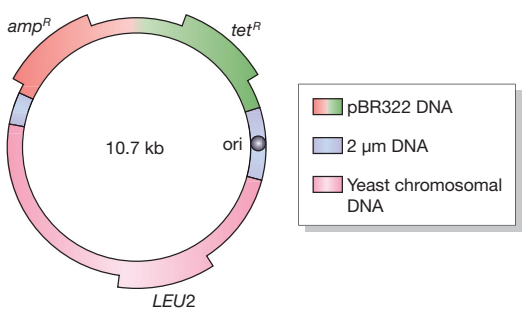
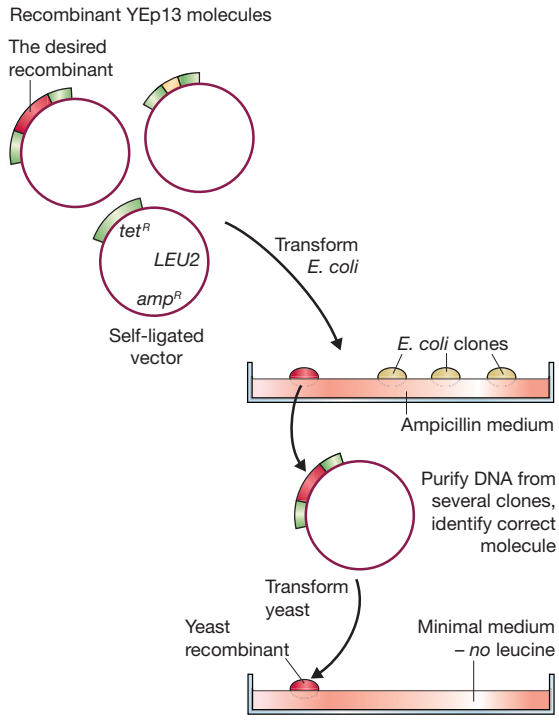


Figure 7.3

A yeast episomal plasmid, YEp13.

Figure 7.4

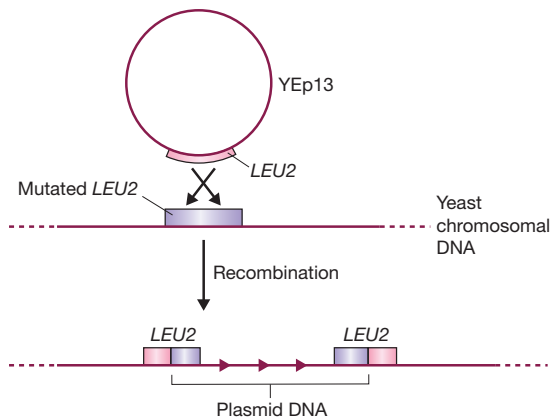
Cloning with an *E. coli*–yeast shuttle vector such as YEp13.



vector as a selectable marker is very similar to the mutant version of the gene present in the yeast chromosomal DNA. With YEp13, for example, **homologous recombination** can occur between the plasmid *LEU2* gene and the yeast mutant *LEU2* gene, resulting in insertion of the entire plasmid into one of the yeast chromosomes (Figure 7.5). The plasmid may remain integrated, or a later recombination event may result in it being excised again.

Figure 7.5

Recombination between plasmid and chromosomal *LEU2* genes can integrate YEp13 into yeast chromosomal DNA. After integration there are two copies of the *LEU2* gene. Usually one of these copies is functional, and the other mutated.



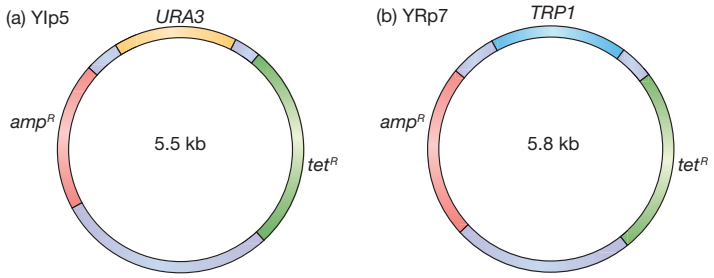


Figure 7.6
A YIp and a YRp.

7.1.4 Other types of yeast cloning vector

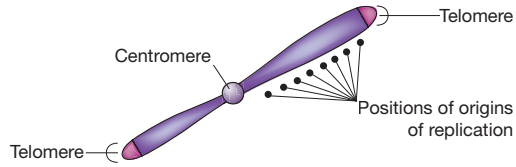
In addition to YEps, there are several other types of cloning vector for use with *S. cerevisiae*. Two important ones are as follows:

- **Yeast integrative plasmids (YIps)** are basically bacterial plasmids carrying a yeast gene. An example is YIp5, which is pBR322 with an inserted *URA3* gene (Figure 7.6a). This gene codes for orotidine-5′-phosphate decarboxylase (an enzyme that catalyses one of the steps in the biosynthesis of pyrimidine nucleotides) and is used as a selectable marker in exactly the same way as *LEU2*. A YIp cannot replicate independently as it does not contain any parts of the 2 μm plasmid, and instead depends for its survival on integration into yeast chromosomal DNA. Integration occurs just as described for a YEp (Figure 7.5).
- **Yeast replicative plasmids (YRps)** are able to multiply as independent plasmids because they carry a chromosomal DNA sequence that includes an origin of replication. Replication origins are known to be located very close to several yeast genes, including one or two which can be used as selectable markers. YRp7 (Figure 7.6b) is an example of a replicative plasmid. It is made up of pBR322 plus the yeast gene *TRP1*. This gene, which is involved in tryptophan biosynthesis, is located adjacent to a chromosomal origin of replication. The yeast DNA fragment present in YRp7 contains both *TRP1* and the origin.

Three factors come into play when deciding which type of yeast vector is most suitable for a particular cloning experiment. The first of these, **transformation frequency**, is a measure of the number of transformants that can be obtained per microgram of plasmid DNA. A high transformation frequency is necessary if a large number of recombinants are needed, or if the starting DNA is in short supply. YEps have the highest transformation frequency, providing between 10 000 and 100 000 transformed cells per microgram. YRps are also quite productive, giving between 1000 and 10 000 transformants per microgram, but a YIp yields less than 1000 transformants per microgram, and only 1–10 unless special procedures are used. The low transformation frequency of a YIp reflects the fact that the rather rare chromosomal integration event is necessary before the vector can be retained in a yeast cell.

The second important factor is **copy number**. YEps and YRps have the highest copy numbers of 20–50 and 5–100, respectively, whereas a YIp is usually present at just one copy per cell. These figures are important if the objective is to obtain protein from the

Figure 7.7
Chromosome structure.



cloned gene, as the more copies there are of the gene, the greater the expected yield of the protein product.

So why would one ever wish to use a YIp? The answer is because YIps produce very stable recombinants, as the loss of a YIp that has become integrated into a chromosome occurs at only a very low frequency. In contrast, YRp recombinants are extremely unstable, the plasmids tending to congregate in the mother cell when a daughter cell buds off, so the daughter cell is non-recombinant. YEp recombinants suffer from similar problems, though an improved understanding of the biology of the 2 μm plasmid has enabled more stable YEps to be developed in recent years. Nevertheless, a YIp is the vector of choice if the needs of the experiment dictate that the recombinant yeast cells must retain the cloned gene for long periods in culture.

7.1.5 Artificial chromosomes can be used to clone long pieces of DNA in yeast

The final type of yeast cloning vector to consider is the **yeast artificial chromosome (YAC)**, which presents a totally different approach to gene cloning. The development of YACs was a spin-off from fundamental research into the structure of eukaryotic chromosomes, work that has identified the key components of a chromosome as being (Figure 7.7):

- The centromere, which is required for the chromosome to be distributed correctly to daughter cells during cell division.
- Two telomeres, the structures at the ends of a chromosome, which are needed in order for the ends to be replicated correctly and which also prevent the chromosome from being nibbled away by exonucleases.
- The origins of replication, which are the positions along the chromosome at which DNA replication initiates, similar to the origin of replication of a plasmid.

Once chromosome structure had been defined in this way, the possibility arose that the individual components might be isolated by recombinant DNA techniques and then joined together again in the test tube, creating an artificial chromosome. As the DNA molecules present in natural yeast chromosomes are several hundred kilobases in length, it might be possible with an artificial chromosome to clone long pieces of DNA.

The structure and use of a YAC vector

Although several YAC vectors have been developed, each one is constructed along the same lines, with pYAC3 being a typical example (Figure 7.8a). At first glance, pYAC3 does not resemble an artificial chromosome but on closer examination its unique features become apparent. pYAC3 is essentially a pBR322 plasmid into which a number of

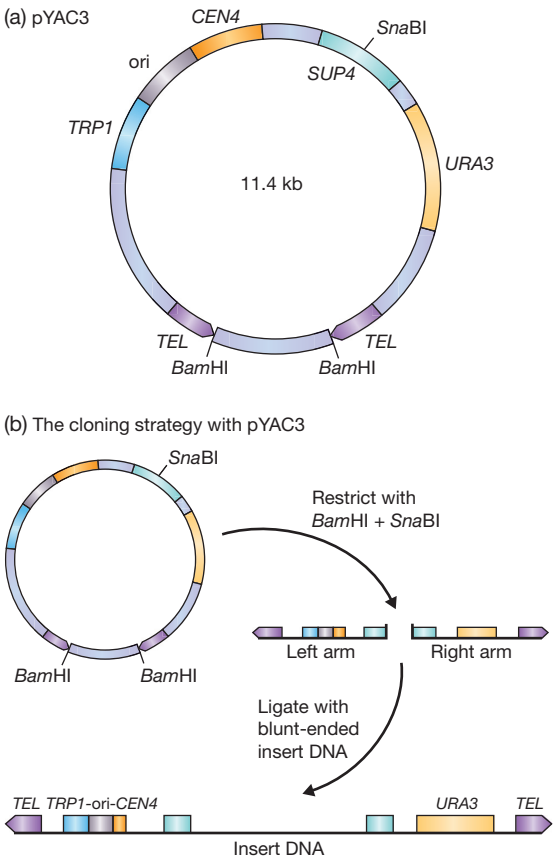


Figure 7.8
A YAC vector and the way it is used to clone large pieces of DNA.

yeast genes have been inserted. Two of these genes, *URA3* and *TRP1*, have been encountered already as the selectable markers for YIp5 and YRp7, respectively. As in YRp7, the DNA fragment that carries *TRP1* also contains an origin of replication, but in pYAC3 this fragment is extended even further to include the sequence called *CEN4*, which is the DNA from the centromere region of chromosome 4. The *TRP1*–origin–*CEN4* fragment therefore contains two of the three components of the artificial chromosome.

The third component, the telomeres, is provided by the two sequences called *TEL*. These are not themselves complete telomere sequences, but once inside the yeast nucleus they act as seeding sequences onto which telomeres will be built. This just leaves one other part of pYAC3 that has not been mentioned: *SUP4*, which is the selectable marker into which new DNA is inserted during the cloning experiment.

The cloning strategy with pYAC3 is as follows (Figure 7.8b). The vector is first restricted with a combination of *Bam*HI and *Sna*BI, cutting the molecule into three fragments. The fragment flanked by *Bam*HI sites is discarded, leaving two arms, each bounded by one *TEL* sequence and one *Sna*BI site. The DNA to be cloned, which must have blunt ends (*Sna*BI is a blunt end cutter, recognizing the sequence TACGTA), is ligated between the two arms, producing the artificial chromosome. Protoplast transformation (p. 88) is then used to introduce the artificial chromosome into *S. cerevisiae*. The yeast strain that is used is a double auxotrophic mutant, $trp1^-$ $ura3^-$, which is converted to $trp1^+$ $ura3^+$ by the two markers on the artificial chromosome. Transformants

are therefore selected by plating onto minimal medium, on which only cells containing a correctly constructed artificial chromosome are able to grow. Any cell transformed with an incorrect artificial chromosome, containing two left or two right arms rather than one of each, is not able to grow on minimal medium as one of the markers is absent. The presence of the insert DNA in the vector can be checked by testing for insertional inactivation of *SUP4*, which is carried out using a simple colour test: white colonies are recombinants, red colonies are not.

Applications for YAC vectors

The initial stimulus in designing artificial chromosomes originated from yeast geneticists who wished to use them for studying various aspects of chromosome structure and behaviour, for instance to examine the segregation of chromosomes during meiosis. These experiments established that artificial chromosomes are stable during propagation in yeast cells, and raised the possibility that they might be used as vectors for genes that are too long to be cloned as a single fragment in an *E. coli* vector. Several important mammalian genes are greater than 100 kb in length (e.g., the human cystic fibrosis gene is 250 kb), which is beyond the capacity of all but the most sophisticated *E. coli* cloning systems (p. 106) but well within the range of a YAC vector. Yeast artificial chromosomes therefore opened the way to studies of the functions and modes of expression of genes that previously had been intractable to analysis by recombinant DNA techniques. A new dimension to these experiments was provided by the discovery that, under some circumstances, YACs can be propagated in mammalian cells, enabling the functional analysis to be carried out in the organism in which the gene normally resides.

Yeast artificial chromosomes are equally important in the production of gene libraries. Recall that with fragments of 300 kb, the maximum insert size for the highest capacity *E. coli* vector, some 30 000 clones are needed for a human gene library (p. 106). YAC vectors are routinely used to clone 600 kb fragments, and special types are able to handle DNA up to 1400 kb in length, the latter bringing the size of a human gene library down to just 6500 clones. Unfortunately, these 'mega-YACs' have run into problems of insert stability, the cloned DNA sometimes becoming rearranged by intramolecular recombination. Nevertheless, YACs have been of immense value in providing long pieces of cloned DNA for use in large-scale DNA-sequencing projects.

7.1.6 Vectors for other yeasts and fungi

Cloning vectors for other species of yeast and fungi are needed for basic studies of the molecular biology of these organisms, and to extend the possible uses of yeasts and fungi in biotechnology. Episomal plasmids based on the *S. cerevisiae* 2 μm plasmid are able to replicate in a few other types of yeast, but the range of species is not broad enough for 2 μm vectors to be of general value. In any case, the requirements of biotechnology are better served by integrative plasmids, equivalent to YIps, as these provide stable recombinants that can be grown for long periods. Efficient integrative vectors are now available for a number of species, including yeasts such as *Pichia pastoris* and *Kluveromyces lactis*, and the filamentous fungi such as *Aspergillus nidulans* and *Neurospora crassa*. We will look at these vectors in more detail when we study the use of yeasts and fungi in recombinant protein production (p. 260).

7.2 Cloning vectors for higher plants

Cloning vectors for higher plants were developed during the 1980s, and their use has led to the **genetically modified (GM) crops** that are in the headlines today. We will examine the genetic modification of crops and other plants in Chapter 15. Here we look at the cloning vectors and how they are used.

Three types of vector system have been used with varying degrees of success with higher plants:

- Vectors based on naturally occurring plasmids of *Agrobacterium*.
- Direct gene transfer using various types of plasmid DNA.
- Vectors based on plant viruses.

7.2.1 Agrobacterium tumefaciens: *nature's smallest genetic engineer*

Although no naturally occurring plasmids are known in higher plants, one bacterial plasmid – the Ti plasmid of *Agrobacterium tumefaciens* – is of great importance.

A. tumefaciens is a soil microorganism that causes crown gall disease in many species of dicotyledonous plants. Crown gall occurs when a wound on the stem allows *A. tumefaciens* bacteria to invade the plant. After infection, the bacteria cause a cancerous proliferation of the stem tissue in the region of the crown (Figure 7.9).

The ability to cause crown gall disease is associated with the presence of the Ti (tumour-inducing) plasmid within the bacterial cell. This is a large (>200 kb) plasmid that carries numerous genes involved in the infective process (Figure 7.10a). A remarkable feature of the Ti plasmid is that, after infection, part of the molecule is integrated into the plant chromosomal DNA (Figure 7.10b). This segment, called the **T-DNA**, is between 15 and 30 kb in size, depending on the strain. The T-DNA is maintained in a stable form in the plant cell, and is passed on to daughter cells as an integral part of the chromosomes. The most remarkable feature of the Ti plasmid is that the T-DNA contains eight or so genes that are expressed in the plant cell and are responsible for the cancerous properties of the transformed cells. These genes also direct the synthesis of unusual compounds, called opines, that the bacteria use as nutrients (Figure 7.10c). In short, *A. tumefaciens* genetically engineers the plant cell for its own purposes.

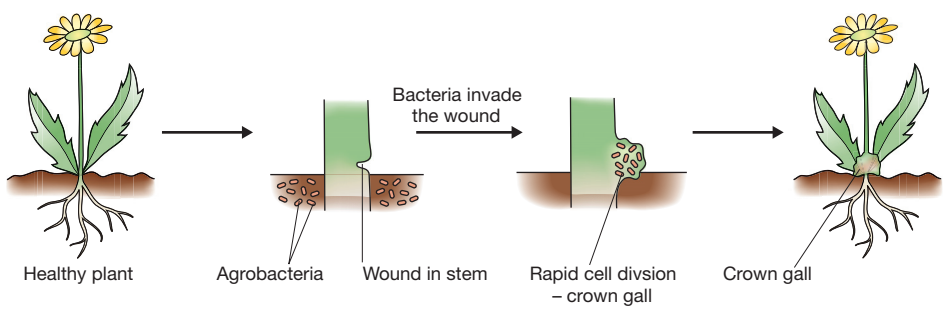
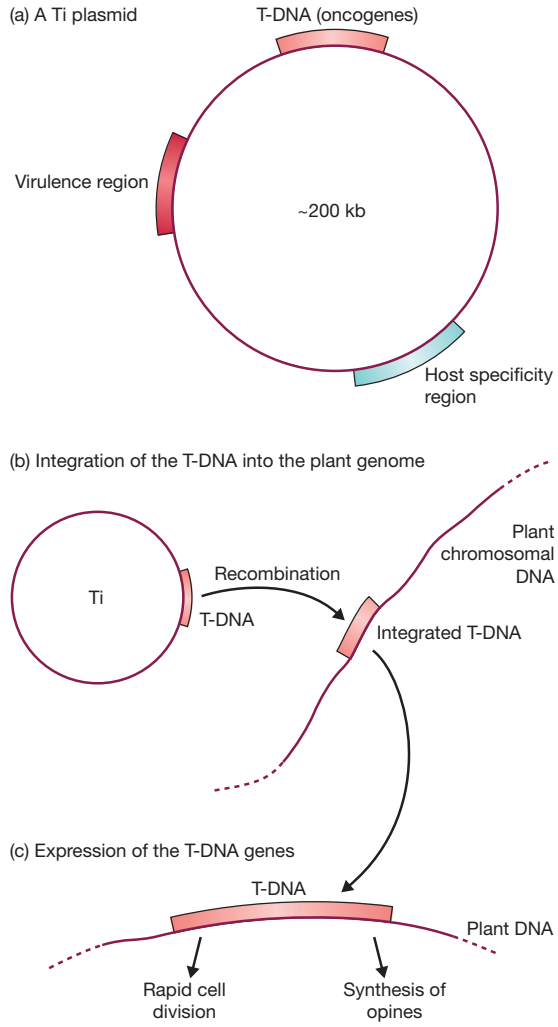


Figure 7.9
Crown gall disease.

Figure 7.10
The Ti plasmid and its integration into the plant chromosomal DNA after *A. tumefaciens* infection.



Using the Ti plasmid to introduce new genes into a plant cell

It was realized very quickly that the Ti plasmid could be used to transport new genes into plant cells. All that would be necessary would be to insert the new genes into the T-DNA, after which the bacterium could do the 'hard work' of integrating them into the plant chromosomal DNA. In practice this has proved a tricky proposition, mainly because the large size of the Ti plasmid makes manipulation of the molecule very difficult.

The main problem is, of course, that it is very unlikely that there would be any unique restriction sites in a plasmid 200 kb in size, and modifying the plasmid to change all but one of the, for example, *Bam*HI sites would be impracticable. Novel strategies therefore had to be developed for inserting new DNA into the plasmid. Two are in general use:

- The **binary vector strategy** (Figure 7.11) is based on the observation that the T-DNA does not need to be physically attached to the rest of the Ti plasmid.

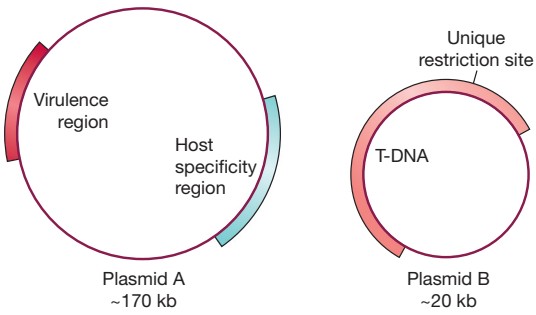


Figure 7.11

The binary vector strategy. Plasmids A and B complement each other when present together in the same *A. tumefaciens* cell. The T-DNA carried by plasmid B is transferred to the plant chromosomal DNA by proteins coded by genes carried by plasmid A.

A two-plasmid system, with the T-DNA on a relatively small molecule, and the rest of the plasmid in normal form, is just as effective at transforming plant cells. In fact, some strains of *A. tumefaciens*, and related agrobacteria, have natural binary plasmid systems. The T-DNA plasmid is small enough to have a unique restriction site and to be manipulated using standard techniques.

- The **co-integration strategy** (Figure 7.12) uses an entirely new plasmid, based on an *E. coli* vector, but carrying a small portion of the T-DNA. The homology between the new molecule and the Ti plasmid means that if both are present in the same *A. tumefaciens* cell, recombination can integrate the *E. coli* plasmid into the T-DNA

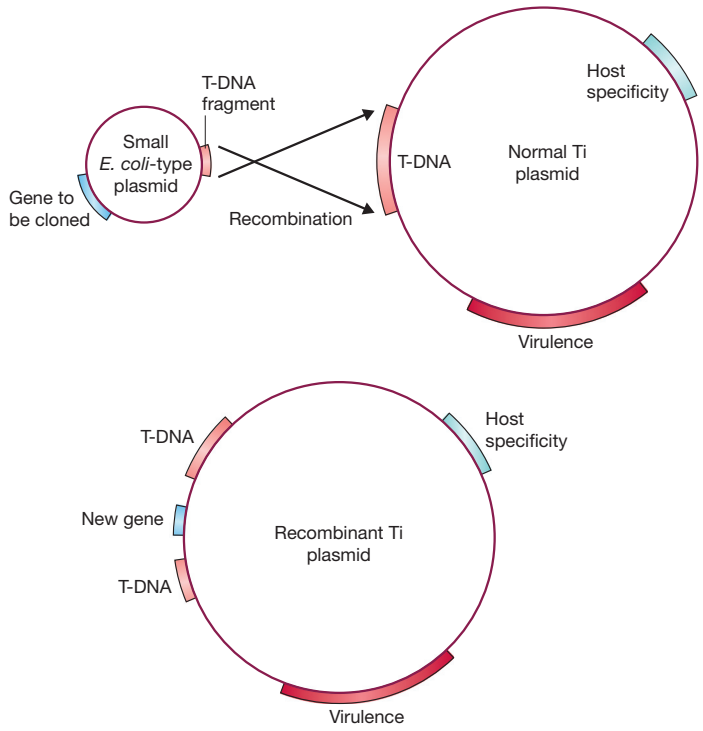


Figure 7.12

The co-integration strategy.

region. The gene to be cloned is therefore inserted into a unique restriction site on the small *E. coli* plasmid, introduced into *A. tumefaciens* cells carrying a Ti plasmid, and the natural recombination process left to integrate the new gene into the T-DNA. Infection of the plant leads to insertion of the new gene, along with the rest of the T-DNA, into the plant chromosomes.

Production of transformed plants with the Ti plasmid

If *A. tumefaciens* bacteria that contain an engineered Ti plasmid are introduced into a plant in the natural fashion, by infection of a wound in the stem, then only the cells in the resulting crown gall will possess the cloned gene (Figure 7.13a). This is obviously of little value to the biotechnologist, and an alternative method of introducing the new gene into every cell in the plant is clearly needed.

There are several solutions to this problem, the simplest being to infect not the mature plant but rather a culture of plant cells or protoplasts (p. 88) in liquid medium (Figure 7.13b). Plant cells and protoplasts whose cell walls have re-formed can be treated in the same way as microorganisms. In particular, they can be plated onto a selective medium

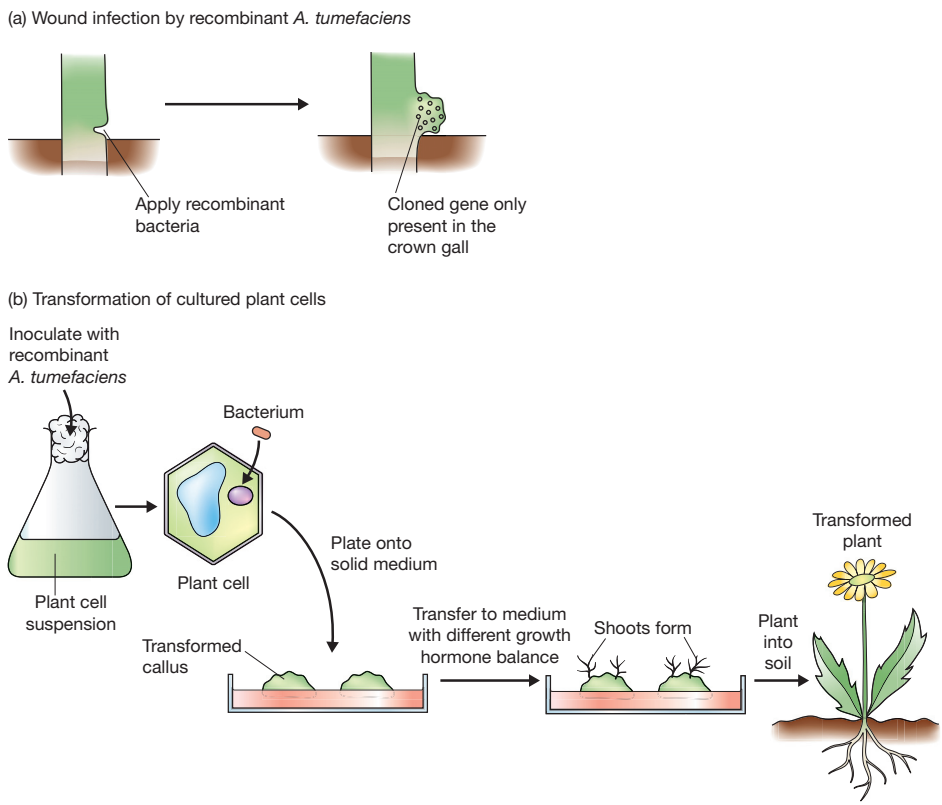


Figure 7.13

Transformation of plant cells by recombinant *A. tumefaciens*. (a) Infection of a wound: transformed plant cells are present only in the crown gall. (b) Transformation of a cell suspension: all the cells in the resulting plant are transformed.

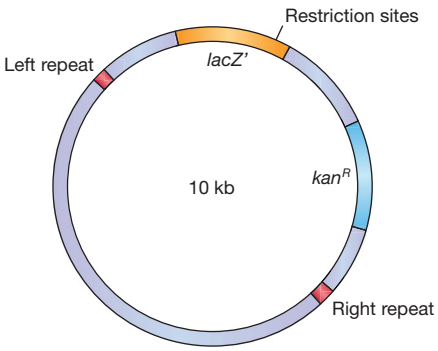


Figure 7.14
The binary Ti vector pBIN19. kan^R = kanamycin resistance gene.

in order to isolate transformants. A mature plant regenerated from transformed cells will contain the cloned gene in every cell and will pass the cloned gene to its offspring. However, the regeneration of a transformed plant can occur only if the Ti vector has been '**disarmed**' so that the transformed cells do not display cancerous properties. Disarming is possible because the cancer genes, all of which lie in the T-DNA, are not needed for the infection process, infectivity being controlled mainly by the virulence region of the Ti plasmid. In fact, the only parts of the T-DNA that are involved in infection are two 25 bp repeat sequences found at the left and right borders of the region integrated into the plant DNA. Any DNA placed between these two repeat sequences will be treated as 'T-DNA' and transferred to the plant. It is therefore possible to remove all the cancer genes from the normal T-DNA, and to replace them with an entirely new set of genes, without disturbing the infection process.

Today, a number of disarmed Ti cloning vectors are available, a typical example being the binary vector pBIN19 (Figure 7.14). The left and right T-DNA borders present in this vector flank a copy of the *lacZ′* gene, containing a number of cloning sites, and a kanamycin resistance gene that functions after integration of the vector sequences into the plant chromosome. As with a yeast shuttle vector (p. 112), the initial manipulations that result in insertion of the gene to be cloned into pBIN19 are carried out in *E. coli*, the correct recombinant pBIN19 molecule then being transferred to *A. tumefaciens* and thence into the plant. Transformed plant cells are selected by plating onto agar medium containing kanamycin.

The Ri plasmid

Over the years there has also been much interest in developing plant cloning vectors based on the **Ri plasmid** of *Agrobacterium rhizogenes*. The Ri and Ti plasmids are very similar, the main difference being that transfer of the T-DNA from an Ri plasmid to a plant results not in a crown gall but in hairy root disease, which is typified by a massive proliferation of a highly branched root system. The possibility of growing transformed roots at high density in liquid culture has been explored by biotechnologists as a potential means of obtaining large amounts of protein from genes cloned in plants.

Limitations of cloning with Agrobacterium *plasmids*

Higher plants are divided into two broad categories, the monocotyledons (monocots) and the dicotyledons (dicots). Several factors have combined to make it much easier to

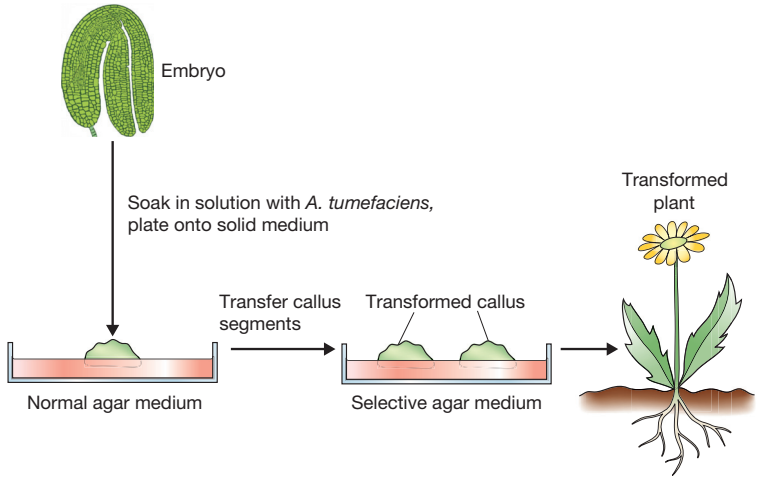


Figure 7.15
Embryo transformation.

clone genes in dicots such as tomato, tobacco, potato, peas and beans, but much more difficult to obtain the same results with monocots. This has been frustrating because monocots include wheat, barley, rice and maize, which are the most important crop plants and hence the most desirable targets for genetic engineering projects.

One difficulty stems from the fact that, in nature, *A. tumefaciens* and *A. rhizogenes* infect only dicotyledonous plants, and monocots are therefore outside of the normal host range. For some time this natural barrier was thought to be insurmountable, and that monocots were totally resistant to transformation with Ti and Ri vectors, but eventually techniques for achieving T-DNA transfer were eventually devised. But this was not the end of the story, as transformation with an *Agrobacterium* vector normally involves the regeneration of an intact plant from a transformed protoplast, cell, or callus culture. The ease with which a plant can be regenerated depends very much on the particular species involved and, once again, the most difficult plants are the monocots. With monocots, the focus has been on callus derived from embryos, as it is easier to regenerate plants from embryogenic callus than it is from callus derived from somatic tissues. An embryo is first soaked in a solution containing recombinant *A. tumefaciens*, and then placed on an agar medium that induces callus formation (Figure 7.15). As not all of the cells in the embryo become transformed, pieces of callus are cultured on a selective medium on which only the transformants can grow. Mature plants are then regenerated from the transformed callus. This approach has been used successfully with maize, barley, and several other important monocots.

7.2.2 Cloning genes in plants by direct gene transfer

Although the Ti plasmid has always been looked on as the most promising system for cloning genes in plants, limitations with the natural host range of *A. tumefaciens* have stimulated the search for alternative methods for introducing new DNA into plant cells; one such method is **direct gene transfer**.

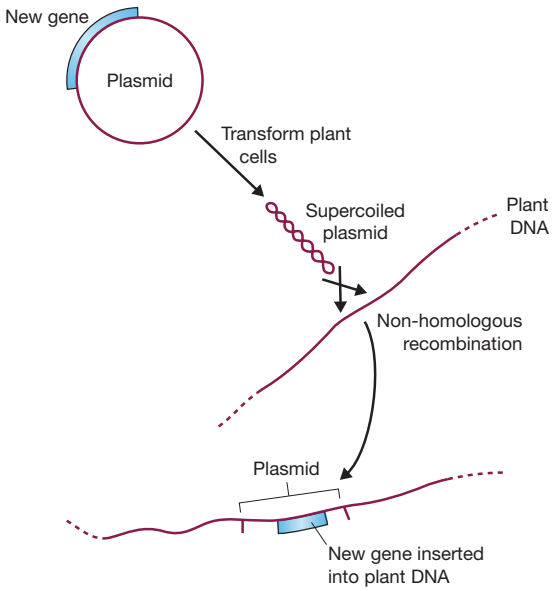


Figure 7.16
Direct gene transfer.

Direct gene transfer into the nucleus

Direct gene transfer is based on the observation, first made in 1984, that a supercoiled bacterial plasmid, although unable to replicate in a plant cell on its own, can become integrated by recombination into one of the plant chromosomes. The recombination event is poorly understood but is almost certainly distinct from the processes responsible for T-DNA integration. It is also distinct from the chromosomal integration of a yeast vector (p. 113), as there is no requirement for a region of similarity between the bacterial plasmid and the plant DNA. Integration appears to occur randomly at any position in any of the plant chromosomes (Figure 7.16).

Direct gene transfer therefore makes use of supercoiled plasmid DNA, possibly a simple bacterial plasmid, into which an appropriate selectable marker (e.g., a kanamycin resistance gene) and the gene to be cloned have been inserted. Biolistics is frequently used to introduce the plasmid DNA into plant embryos, which are then regenerated as described above for *Agrobacterium* transformation. Alternatively, if the species being engineered can be regenerated from protoplasts or single cells, then other strategies, possibly more efficient than biolistics, are possible.

One method involves resuspending protoplasts in a viscous solution of polyethylene glycol (PEG), a polymeric, negatively charged compound that is thought to precipitate DNA onto the surfaces of the protoplasts and to induce uptake by endocytosis (Figure 7.17). Electroporation (p. 89) is also sometimes used to increase transformation frequency. After treatment, the protoplasts are left for a few days in a solution that encourages regeneration of the cell walls. The cells are then spread onto selective medium to identify transformants and to provide callus cultures from which intact plants can be grown.

Transfer of genes into the chloroplast genome

If biolistics is used to introduce DNA into a plant embryo, then some particles may penetrate one or more of the chloroplasts present in the cells. Chloroplasts contain their

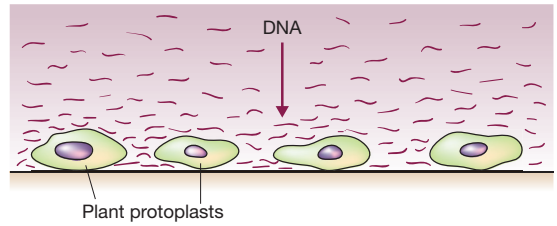


Figure 7.17
Direct gene transfer by precipitation of DNA onto the surfaces of protoplasts.

own genomes, distinct from (and much shorter) than the DNA molecules in the nucleus, and under some circumstances plasmid DNA can become integrated into this chloroplast genome. Unlike the integration of DNA into nuclear chromosomes, integration into the chloroplast genome will not occur randomly. Instead, the DNA to be cloned must be flanked by sequences similar to the region of the chloroplast genome into which the DNA is to be inserted, so that insertion can take place by homologous recombination (see p. 114). Each of these flanking sequences must be 500 bp or so in length. A low level of chloroplast transformation can also be achieved after PEG-induced DNA delivery into protoplasts if the plasmid that is taken up carries these flanking sequences.

As a plant cell contains tens of chloroplasts, and probably only one per cell becomes transformed, the inserted DNA must carry a selectable marker such as the kanamycin resistance gene, and the embryos must be treated with the antibiotic for a considerable period to ensure that the transformed genomes propagate within the cell. Although this means that chloroplast transformation is a difficult method to carry out successfully, it is becoming an important adjunct to the more traditional methods for obtaining GM crops. As each cell has many chloroplasts, but only one nucleus, a gene inserted into the chloroplast genome is likely to be expressed at a higher level than one placed in the nucleus. This is particularly important when the engineered plants are to be used for the production of pharmaceutical proteins (see Chapter 14). To date, this approach has been most successful with tobacco, though chloroplast transformation has also been achieved with more useful crops such as soybean and cotton.

7.2.3 Attempts to use plant viruses as cloning vectors

Modified versions of λ and M13 bacteriophages are important cloning vectors for *E. coli* (Chapter 6). Most plants are subject to viral infection, so could viruses be used to clone genes in plants? If they could, then they would be much more convenient to use than other types of vector, because with many viruses transformation can be achieved simply by rubbing the virus nucleic acid onto the surface of a leaf. The natural infection process then spreads the virus throughout the plant.

The potential of plant viruses as cloning vectors has been explored for several years, but without great success. One problem here is that the vast majority of plant viruses have genomes not of DNA but of RNA. RNA viruses are less useful as potential cloning vectors because manipulations with RNA are more difficult to carry out. Only two classes of DNA virus are known to infect higher plants – the **caulimoviruses** and **geminiviruses** – and neither is ideally suited to gene cloning.

Caulimovirus vectors

Although one of the first successful plant genetic engineering experiments, performed back in 1984, used a caulimovirus vector to clone a new gene into turnip plants, two general difficulties with these viruses have limited their usefulness.

The first problem is that the total size of a caulimovirus genome is, like that of λ, constrained by the need to package it into its protein coat. Even after the deletion of non-essential sections of the virus genome the capacity for carrying inserted DNA is still very limited. Recent studies have shown that it might be possible to circumvent this problem by adopting a helper virus strategy, similar to that used with phagemids (p. 108). In this strategy, the cloning vector is a **cauliflower mosaic virus (CaMV)** genome that lacks several of the essential genes, which means that it can carry a large DNA insert but cannot, by itself, direct infection. Plants are inoculated with the vector DNA along with a normal CaMV genome. The normal viral genome then provides the genes needed for the cloning vector to be packaged into virus proteins and spread through the plant.

Although this approach has considerable potential it does not solve the second problem, which is the extremely narrow host range of caulimoviruses. This restricts cloning experiments to just a few plants, mainly brassicas such as turnips, cabbages and cauliflowers. Caulimoviruses have, however, been important in genetic engineering as the source of highly active promoters that function in all plants and that are used to obtain expression of genes introduced by Ti plasmid cloning or direct gene transfer.

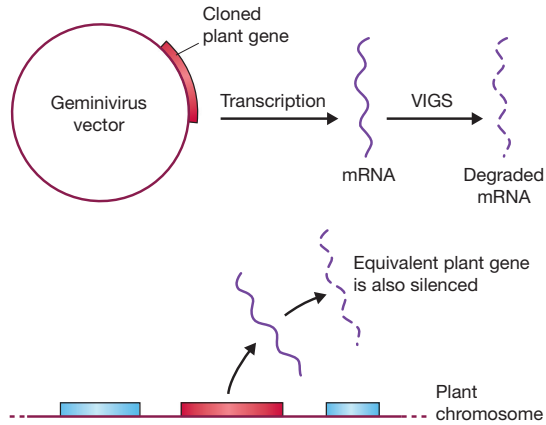
Geminivirus vectors

The geminiviruses are particularly interesting because their natural hosts include plants such as maize and wheat, and they could therefore serve as potential vectors for these and other monocots. But geminiviruses have presented their own set of difficulties, one problem being that during the infection cycle the genomes of some geminiviruses undergo rearrangements and deletions, which would scramble up any additional DNA that has been inserted – an obvious disadvantage for a cloning vector. Investigations performed over the years have addressed these problems and geminiviruses are now beginning to find some specialist applications in plant gene cloning. One such application is **virus-induced gene silencing (VIGS)**, a technique used to investigate the functions of individual plant genes. This method exploits one of the natural defence mechanisms that plants use to protect themselves against viral attack. This method, termed **RNA silencing**, results in the degradation of viral mRNAs. If one of the viral RNAs is transcribed from a cloned gene contained within a geminivirus genome, then not only are the viral transcripts degraded but also the cellular mRNAs derived from the plant's copy of the gene (Figure 7.18). Consequently, the plant gene becomes silenced and the effect of its inactivation on the phenotype of the plant can be studied.

7.3 Cloning vectors for animals

During recent years, considerable effort has been expended into the development of vector systems for cloning genes in animal cells. These vectors are required in biotechnology for the synthesis of recombinant protein from genes that are not expressed correctly when cloned in *E. coli* or yeast (see Chapter 13), and methods for cloning in humans are being sought by clinical molecular biologists attempting to devise techniques

Figure 7.18
The use of a geminivirus vector to silence a plant gene via virus-induced gene silencing.



for **gene therapy** (p. 286), in which a disease is treated by the introduction of a cloned gene into a patient.

The clinical aspect of this has led to most attention being directed at cloning systems for mammals, although important progress has also been made with insects. Cloning in insects is interesting because it makes use of a novel type of vector that we have not met so far. We will therefore examine insect vectors before concluding the chapter with an overview of the cloning methods used with mammals.

7.3.1 Cloning vectors for insects

The fruit fly, *Drosophila melanogaster*, has been – and still is – one of the most important model organisms used by biologists. Its potential was first recognized by the famous geneticist Thomas Hunt Morgan, who in 1910 started to carry out genetic crosses between fruit flies with different eye colours, body shapes, and other inherited characteristics. These experiments led to the techniques still used today for gene mapping in insects and other animals. More recently, the discovery that the homeotic selector genes of *Drosophila* – the genes that control the overall body plan of the fly – are closely related to equivalent genes in mammals, has led to *D. melanogaster* being used as a model for the study of human developmental processes. The importance of the fruit fly in modern biology makes it imperative that vectors for cloning genes in this organism are available.

P elements as cloning vectors for Drosophila

The development of cloning vectors for *Drosophila* has taken a different route to that followed with bacteria, yeast, plants and mammals. No plasmids are known in *Drosophila*, and although fruit flies are, like all organisms, susceptible to infection with viruses, these have not been used as the basis for cloning vectors. Instead, cloning in *Drosophila* makes use of a **transposon** called the **P element.**

Transposons, which are common in all types of organism, are short pieces of DNA (usually less than 10 kb in length) that can move from one position to another in the chromosomes of a cell. P elements, which are one of several types of transposon in *Drosophila*, are 2.9 kb in length and contain three genes flanked by short, inverted repeat sequences at either end of the element (Figure 7.19a). The genes code for

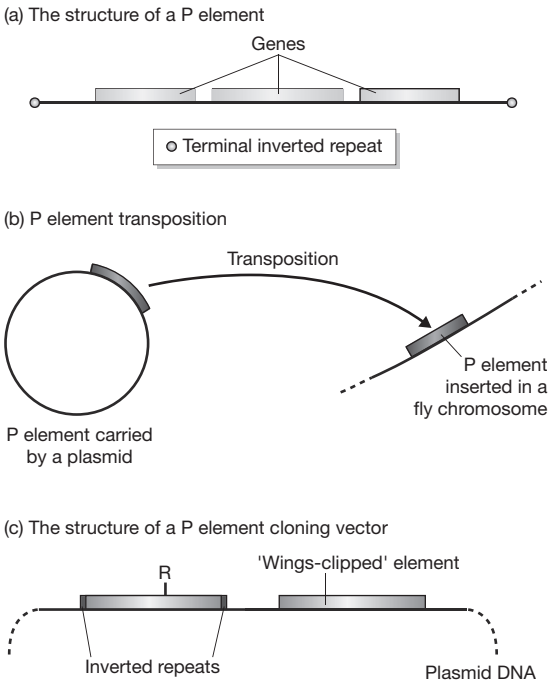


Figure 7.19

Cloning in *Drosophila* with a P element vector. (a) The structure of a P element. (b) Transposition of a P element from a plasmid to a fly chromosome. (c) The structure of a P element cloning vector. The left-hand P element contains a cloning site (R) that disrupts its transposase gene. The right-hand P element has an intact transposase gene but cannot itself transpose because it is 'wings-clipped' – it lacks terminal inverted repeats.

transposase, the enzyme that carries out the transposition process, and the inverted repeats form the recognition sequences that enable the enzyme to identify the two ends of the inserted transposon.

As well as moving from one site to another within a single chromosome, P elements can also jump between chromosomes, or between a plasmid carrying a P element and one of the fly's chromosomes (Figure 7.19b). The latter is the key to the use of P elements as cloning vectors. The vector is a plasmid that carries two P elements, one of which contains the insertion site for the DNA that will be cloned. Insertion of the new DNA into this P element results in disruption of its transposase gene, so this element is inactive. The second P element carried by the plasmid is therefore one that has an intact version of the transposase gene. Ideally, this second element should not itself be transferred to the *Drosophila* chromosomes, so it has its 'wings clipped', which means that its inverted repeats are removed so that the transposase does not recognize it as being a real P element (Figure 7.19c). Once the gene to be cloned has been inserted into the vector, the plasmid DNA is microinjected into fruit fly embryos. The transposase from the wings-clipped P element directs transfer of the engineered P element into one of the fruit fly chromosomes. If this happens within a germline nucleus, then the adult fly that develops from the embryo will carry copies of the cloned gene in all its cells. P element cloning was first developed during the 1980s and has made a number of important contributions to *Drosophila* genetics.

Cloning vectors based on insect viruses

Although virus vectors have not been developed for cloning genes in *Drosophila*, one type of virus – the **baculovirus** – has played an important role in gene cloning with

other insects. The main use of baculovirus vectors is in the production of recombinant protein, and we will return to them when we consider this topic in Chapter 13.

7.3.2 Cloning in mammals

At present, gene cloning in mammals is carried out for one of three reasons:

- To achieve a **gene knockout**, which is an important technique used to help determine the function of an unidentified gene (p. 233). These experiments are usually carried out with rodents, such as mice.
- For the production of recombinant protein in a mammalian cell culture, and in the related technique of **pharming**. The latter method involves the genetic engineering of a farm animal so that it synthesizes an important protein such as a pharmaceutical, often in its milk (p. 263).
- In **gene therapy**, in which human cells are engineered in order to treat a disease (p. 286).

Viruses as cloning vectors for mammals

For many years it was thought that viruses would prove to be the key to cloning in mammals, but this expectation has only partially been realized. The first cloning experiment involving mammalian cells was carried out in 1979 with a vector based on simian virus 40 (SV40). This virus is capable of infecting several mammalian species, following a lytic cycle in some hosts and a lysogenic cycle in others. The genome is 5.2 kb in size (Figure 7.20a) and contains two sets of genes, the 'early' genes, which are expressed early in the infection cycle and code for proteins involved in viral DNA replication, and the 'late' genes, which code for the viral capsid proteins. SV40 suffers from the same problem as λ and the plant caulimoviruses, in that packaging constraints limit the amount of new DNA that can be inserted into the genome. Cloning with SV40 therefore involves replacing one or more of the existing genes with the DNA to be cloned. In the original experiment, a segment of the late gene region was replaced (Figure 7.20b), but early gene replacement is also an option.

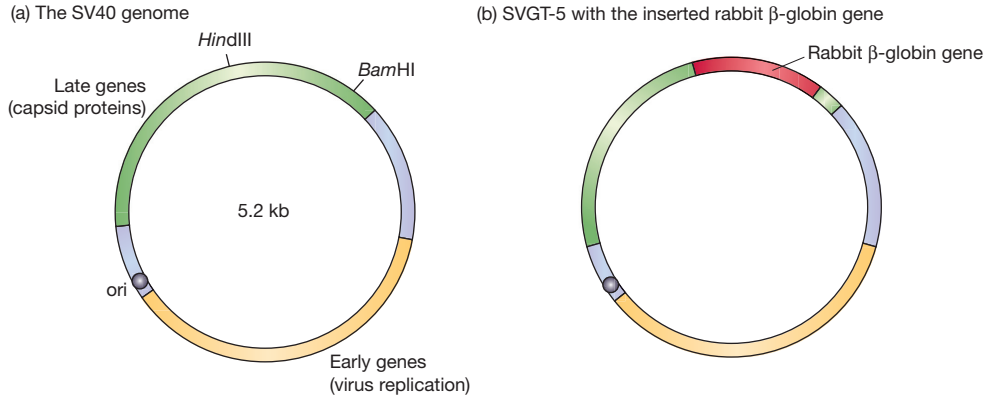


Figure 7.20

SV40 and an example of its use as a cloning vector. To clone the rabbit β-globin gene, the *Hin*dIII to *Bam*HI restriction fragment was deleted (resulting in SVGT-5) and replaced with the rabbit gene.

During the years since 1979, a number of other types of virus have been used to clone genes in mammals, including:

- **Adenoviruses**, which enable DNA fragments of up to 8 kb to be cloned, longer than is possible with an SV40 vector, though adenoviruses are more difficult to handle because their genomes are bigger.
- **Papillomaviruses**, which also have a relatively high capacity for inserted DNA. Bovine papillomavirus (BPV), which causes warts on cattle, is particularly attractive because it has an unusual infection cycle in mouse cells, taking the form of a multicopy plasmid with about 100 molecules present per cell. It does not cause death of the mouse cell, and BPV molecules are passed to daughter cells on cell division, giving rise to a permanently transformed cell line. Shuttle vectors consisting of BPV and *E. coli* sequences, and capable of replication in both mouse and bacterial cells, have been used for the production of recombinant proteins in mouse cell lines.
- **Adeno-associated virus (AAV)**, which is unrelated to adenovirus but often found in the same infected tissues, mainly because AAV utilizes some of the proteins synthesized by adenovirus in order to complete its replication cycle. In the absence of this helper virus, the AAV genome inserts into its host's DNA. With most integrative viruses this is a random event, but AAV has the unusual property of always inserting at the same position, within human chromosome 19. Knowing exactly where the cloned gene will be in the host genome is important if the outcome of the cloning experiment must be checked rigorously, as is the case for applications such as gene therapy. AAV vectors are therefore considered to have major potential in this area.
- **Retroviruses**, which are the most commonly used vectors for gene therapy. Although they insert at random positions, the resulting integrants are very stable, which means that the therapeutic effects of the cloned gene will persist for some time. We will return to gene therapy in Chapter 14.

Gene cloning without a vector

One of the reasons why virus vectors have not become widespread in mammalian gene cloning is because it was discovered in the early 1990s that the most effective way of transferring new genes into mammalian cells is by microinjection. Although a difficult procedure to carry out, microinjection of bacterial plasmids or linear DNA copies of genes into mammalian nuclei results in the DNA being inserted into the chromosomes, possibly as multiple copies in a tandem, head-to-tail arrangement (Figure 7.21). This procedure is generally looked on as being more satisfactory than the use of a viral vector, because it avoids the possibility that viral DNA will infect the cells and cause defects of one type or another.

Microinjection of DNA forms the basis to creation of a **transgenic animal**, one that contains a cloned gene in all of its cells. A transgenic mouse can be generated by microinjection of a fertilized egg cell which is subsequently cultured *in vitro* for several cell

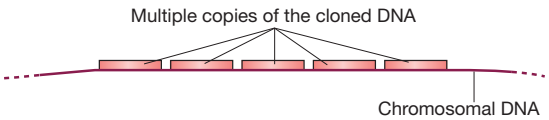


Figure 7.21
Multiple copies of cloned DNA molecules inserted as a tandem array in a chromosomal DNA molecule.

divisions and then implanted into a foster mother. Alternatively, an **embryonic stem (ES) cell** can be used. These are obtained from within an early embryo and, unlike most mammalian cells, are **totipotent**, which means that their developmental pattern is not preset and cells descended from them can form many different structures in the adult mouse. After microinjection, the ES cell is placed back in an embryo which is then implanted into the foster mother. The resulting mouse is a **chimera**, comprising a mixture of engineered and non-engineered cells, because the embryo that receives the ES cell also contains a number of ordinary cells that contribute, along with the ES cell, to the make-up of the adult mouse. Non-chimeric mice, which contain the cloned gene in all their cells, are obtained by allowing the chimera to reproduce, as some of the offspring will be derived from egg cells that contain the cloned gene.

Further reading

FURTHER READING

Brisson, N., Paszkowski, J., Penswick, J.R., *et al.* (1984) Expression of a bacterial gene in plants by using a viral vector. *Nature*, **310**, 511–114. [The first cloning experiment with a caulimovirus.]

Broach, J.R. (1982) The yeast 2 μm circle. *Cell*, **28**, 203–204.

Burke, D.T., Carle, G.F., and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, **236**, 806–812.

Carrillo-Tripp, J., Shimada-Beltrán, H., and Rivera-Bustamante, R. (2006) Use of geminiviral vectors for functional genomics. *Current Opinion in Plant Biology*, **9**, 209–215. [Virus-induced gene silencing.]

Chilton, M.D. (1983) A vector for introducing new genes into plants. *Scientific American*, **248** (June), 50–59. [The Ti plasmid.]

Clarke, J.L., Daniell, H., and Nugent, J.M. (2011) Chloroplast biotechnology, genomics and evolution: current status, challenges and future directions. *Plant Molecular Biology*, **76**, 207–209. [Information on chloroplast transformation.]

Colosimo, A., Goncz, K.K., Holmes, A.R., *et al.* (2000) Transfer and expression of foreign genes in mammalian cells. *Biotechniques*, **29**, 314–321.

Crystal, R.G. (2014) Adenovirus: the first effective *in vivo* gene delivery vector. *Human Gene Therapy*, **25**, 3–11.

Evans, M. (2011) Discovering pluripotency: 30 years of mouse embryonic stem cells. *Nature Reviews Molecular Cell Biology*, **12**, 680–686.

Guillon, S., Trémouillaux-Guiller, J., Pati, P.K., Rideau, M., and Gantet, P. (2006) Hairy root research: recent scenario and exciting prospects. *Current Opinion in Plant Biology*, **9**, 341–346. [Applications of the Ri plasmid.]

Hamer, D.H. and Leder, P. (1979) Expression of the chromosomal mouse β-maj-globin gene cloned in SV40. *Nature*, **281**, 35–40.

Lee, L.-Y. and Gelvin, S.B. (2008) T-DNA binary vectors and systems. *Plant Physiology*, **146**, 325–332.

Păcurar, D.I., Thordal-Christensen, H., Păcurer, M.L., *et al.* (2011) *Agrobacterium tumefaciens*: from crown gall tumors to genetic transformation. *Physiological and Molecular Plant Pathology*, **76**, 76–81.

Parent, S.A., Fenimore, C.M., and Bostian, K.A. (1985) Vector systems for the expression, analysis and cloning of DNA sequences in *S. cerevisiae*. *Yeast*, **1**, 83–138. [Details of different yeast cloning vectors.]

Paszkowski, J., Shillito, R.D., Saul, M., *et al*. (1984) Direct gene transfer to plants. *EMBO Journal*, **3**, 2717–2722.

Rubin, G.M. and Spradling, A.C. (1982) Genetic transformation of *Drosophila* with transposable element vectors. *Science*, **218**, 348–353. [Cloning with P elements.]

Viaplana, R., Turner, D.S., and Covey, S.N. (2001) Transient expression of a GUS reporter gene from cauliflower mosaic virus replacement vectors in the presence and absence of helper virus. *Journal of General Virology*, **82**, 59–65. [The helper virus approach to cloning with CaMV.]

Chapter 8

How to Obtain a Clone of a Specific Gene

Chapter contents

CHAPTER CONTENTS

In the preceding chapters we have examined the basic methodology used to clone genes, and surveyed the range of vector types that are used with bacteria, yeast, plants and animals. Now, we must look at the methods available for obtaining a clone of an individual, specified gene. This is the critical test of a gene cloning experiment, where success or failure often depends on whether or not a strategy can be devised by which clones of the desired gene can be selected directly, or alternatively, distinguished from other recombinants. Once this problem has been resolved, and a clone has been obtained, the molecular biologist is able to make use of a wide variety of different techniques that will extract information about the gene. The most important of these will be described in Chapters 10 and 11.

8.1 The problem of selection

The problem faced by the molecular biologist wishing to obtain a clone of a single, specified gene was illustrated in Figure 1.4. Even the simplest organisms, such as *E. coli*, contain several thousand genes, and a restriction digest of total cell DNA produces not only the fragment carrying the desired gene but also many other fragments carrying all the other genes (Figure 8.1a). During the ligation reaction there is no selection for an individual fragment, with numerous different recombinant DNA molecules being produced, all of which contain different pieces of DNA (Figure 8.1b). Consequently, a

Gene Cloning and DNA Analysis: An Introduction, Seventh Edition. T.A. Brown.

Companion Website: www.wiley.com/go/brown/cloning

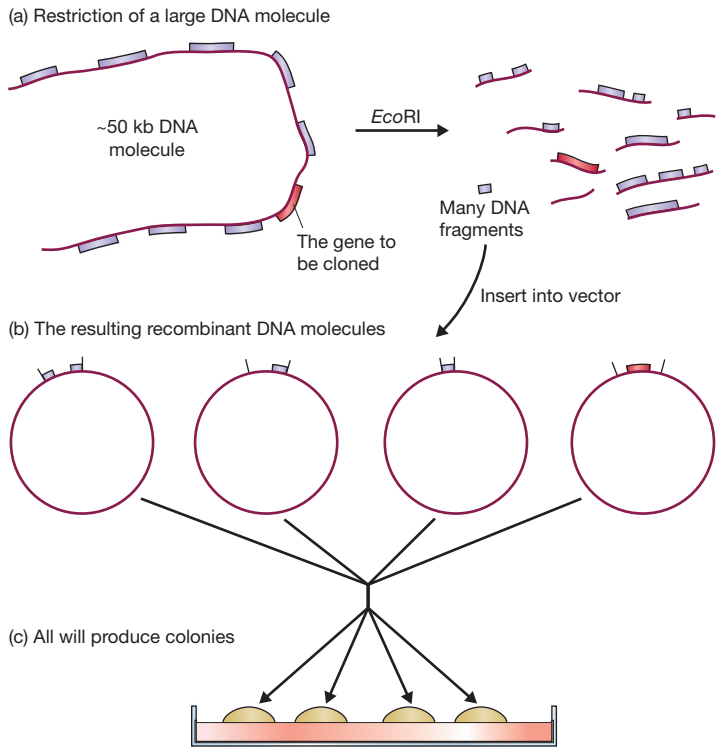


Figure 8.1
The problem of selection.

variety of recombinant clones are obtained after transformation and plating out (Figure 8.1c). Somehow, the correct one must be identified.

8.1.1 There are two basic strategies for obtaining the clone you want

Although there are many different procedures by which the desired clone can be obtained, all are variations on two basic themes:

- **Direct selection for the desired gene** (Figure 8.2a), which means that the cloning experiment is designed in such a way that the only clones that are obtained are those of the required gene. Almost invariably, selection occurs at the plating-out stage.
- **Identification of the clone from a gene library** (Figure 8.2b). This entails an initial **shotgun** cloning experiment to produce a clone library representing all or most of the genes present in the cell, followed by analysis of the individual clones to identify the correct one.

In general terms, direct selection is the preferred method, as it is quick and usually unambiguous. Unfortunately, as we shall see, it is not applicable to all genes and techniques for clone identification are therefore very important.

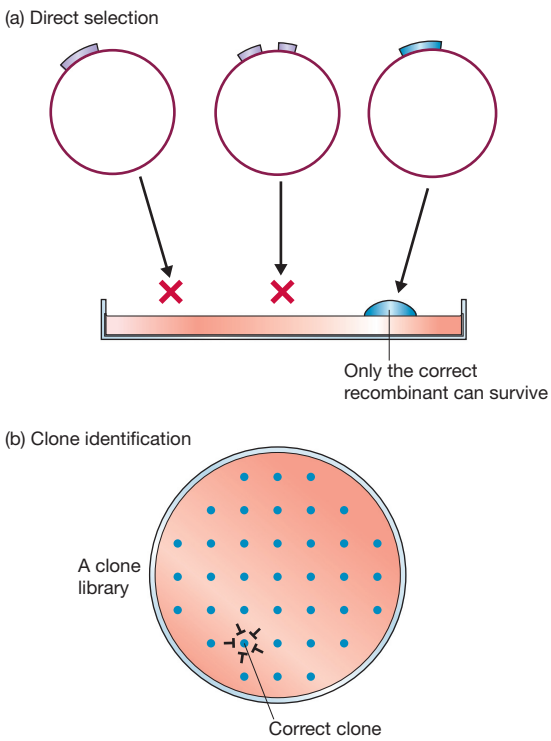


Figure 8.2
The basic strategies that can be used to obtain a particular clone. (a) Direct selection. (b) Identification of the desired recombinant from a clone library.

8.2 Direct selection

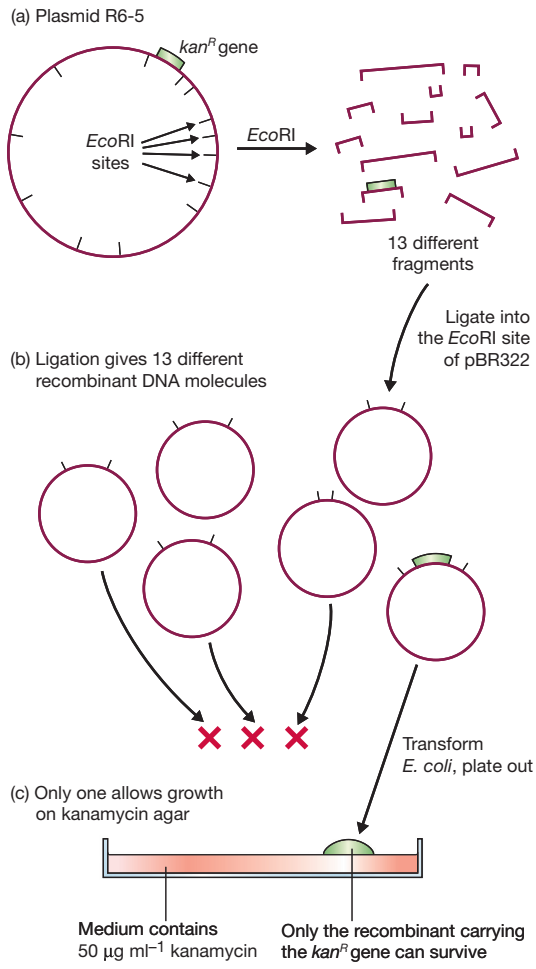
To be able to select for a cloned gene it is necessary to plate the transformants onto an agar medium on which only the desired recombinants, and no others, can grow. The only colonies that are obtained will therefore be ones that comprise cells containing the desired recombinant DNA molecule.

The simplest example of direct selection occurs when the desired gene specifies resistance to an antibiotic. As an example, we will consider an experiment to clone the gene for kanamycin resistance from plasmid R6-5. This plasmid carries genes for resistances to four antibiotics: kanamycin, chloramphenicol, streptomycin, and sulphonamide. The kanamycin resistance gene lies within one of the 13 fragments produced when R6-5 is digested with *Eco*RI (Figure 8.3a).

To clone this gene, the *Eco*RI fragments of R6-5 could be inserted into the *Eco*RI site of a vector such as pBR322. The ligated mix will comprise many copies of 13 different recombinant DNA molecules, one set of which carries the gene for kanamycin resistance (Figure 8.3b).

Insertional inactivation cannot be used to select recombinants when the *Eco*RI site of pBR322 is used. This is because this site does not lie in either the ampicillin or the tetracycline resistance genes of this plasmid (see Figure 6.1). But this is immaterial for cloning the kanamycin resistance gene because in this case the cloned gene can be used as the selectable marker. Transformants are plated onto kanamycin agar, on which the only cells able to survive and produce colonies are those recombinants that contain the cloned kanamycin resistance gene (Figure 8.3c).

Figure 8.3
Direct selection for the cloned R6-5 kanamycin resistance (*kan*^{R}) gene.



8.2.1 Marker rescue extends the scope of direct selection

Direct selection would be very limited indeed if it could be used only for cloning antibiotic resistance genes. Fortunately, the technique can be extended by making use of mutant strains of *E. coli* as the hosts for transformation.

As an example, consider an experiment to clone the gene *trpA* from *E. coli*. This gene codes for the enzyme tryptophan synthase, which is involved in biosynthesis of the essential amino acid tryptophan. A mutant strain of *E. coli* that has a non-functional *trpA* gene is called *trpA*$^-$, and is able to survive only if tryptophan is added to the growth medium. *E. coli trpA*$^-$ is therefore another example of an auxotroph (p. 112).

The *E. coli trpA*$^-$ auxotroph can be used to clone the correct version of the *trpA* gene. Total DNA is first purified from a normal (wild-type) strain of the bacterium. Digestion with a restriction endonuclease, followed by ligation into a vector, produces numerous recombinant DNA molecules, one of which may (with luck) carry an intact copy of the *trpA* gene (Figure 8.4a). This is, of course, the functional gene, as it has been obtained from the wild-type strain.

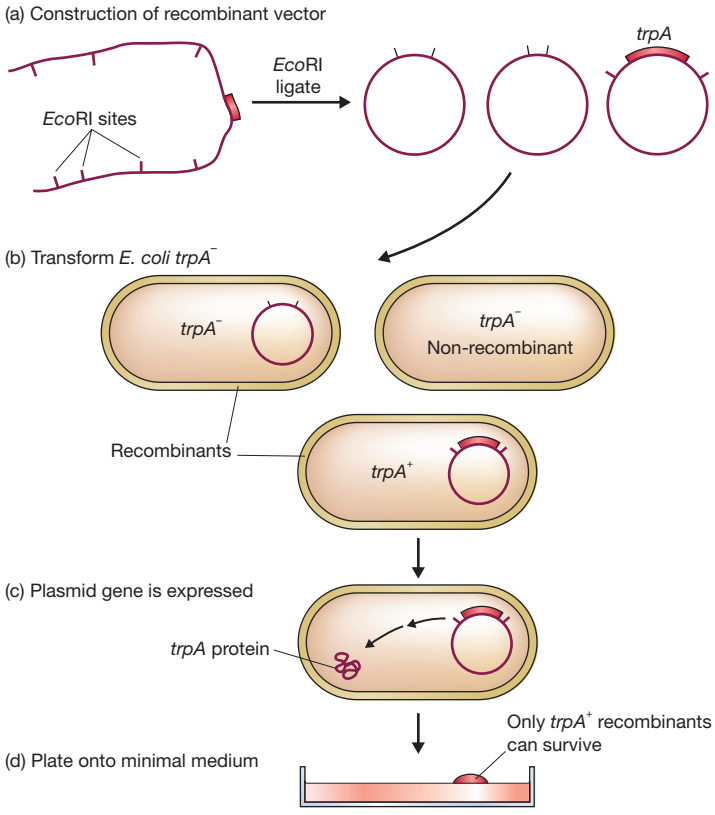


Figure 8.4

Direct selection for the *trpA* gene cloned in a *trpA*⁻ strain of *E. coli*.

The ligation mixture is now used to transform the auxotrophic *E. coli trpA*⁻ cells (Figure 8.4b). The vast majority of the resulting transformants will be auxotrophic, but a few will now have the plasmid-borne copy of the correct *trpA* gene. These recombinants are non-auxotrophic; they no longer require tryptophan as the cloned gene is able to direct the production of tryptophan synthase (Figure 8.4c). Direct selection is therefore performed by plating transformants onto minimal medium, which lacks any added supplements, and in particular has no tryptophan (Figure 8.4d). As auxotrophs cannot grow on minimal medium, the only colonies to appear are recombinants that contain the cloned *trpA* gene.

8.2.2 The scope and limitations of marker rescue

Although marker rescue can be used to obtain clones of many genes, the technique is subject to two limitations:

- A mutant strain must be available for the gene in question.
- A medium on which only the wild-type can survive is needed.

Marker rescue is applicable for most genes that code for biosynthetic enzymes, as clones of these genes can be selected on minimal medium in the manner described for

trpA. The technique is not limited to *E. coli* or bacteria – auxotrophic strains of yeast and filamentous fungi are also available, and marker rescue has been used to select genes cloned into these organisms.

In addition, *E. coli* auxotrophs can be used as hosts for the selection of some genes from other organisms. Often, there is sufficient similarity between equivalent enzymes from different bacteria, or even from yeast, for the foreign enzyme to function in *E. coli*, so that the cloned gene is able to transform the host to wild-type.

8.3 Identification of a clone from a gene library

Although marker rescue is a powerful technique, it is not all-embracing and there are many important genes that cannot be selected using this method. Many bacterial mutants are not auxotrophs, so the mutant and wild-type strains cannot be distinguished by plating onto minimal or any other special medium. In addition, neither marker rescue nor any other direct selection method is of much use in providing bacterial clones of genes from animals or plants, as in these cases the differences are usually so great that the foreign enzymes do not function in the bacterial cell.

An alternative strategy must therefore be considered. This is where a large number of different clones are obtained and the desired one identified in some way.

8.3.1 Gene libraries

Before examining the methods used to identify individual clones, the library itself must be considered. A genomic library (p. 105) is a collection of clones sufficient in number to be likely to contain every single gene present in a particular organism. Genomic libraries are prepared by purifying total cell DNA, and then making a partial restriction digest, resulting in fragments that can be cloned into a suitable vector (Figure 8.5), usually a λ replacement vector, a cosmid, or possibly a yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC) or P1 vector.

Not all genes are expressed at the same time

For bacteria, yeast and fungi, the number of clones needed for a complete genomic library is not so large as to be unmanageable (see Table 6.1). For plants and animals, a complete library contains so many different clones that identification of the desired one may prove a mammoth task. With these multicellular organisms a second type of library, specific not to the whole organism but rather to a particular cell type, might be more useful. Each cell contains the same complement of genes, but in different cell types different sets of genes are switched on, while others are silent (Figure 8.6). Only those genes that are being expressed are transcribed into messenger RNA (mRNA), so if mRNA is used as the starting material then the resulting clones would comprise only a selection of the total number of genes in the cell.

A cloning method that uses mRNA would be particularly useful if the desired gene is expressed at a high rate in an individual cell type. For example, the gene for gliadin (one of the nutritionally important proteins present in wheat) is expressed at a very high level in the cells of developing wheat seeds. In these cells, over 30% of the total mRNA specifies gliadin. Clearly, if we could clone the mRNA from wheat seeds we would obtain a large number of clones specific for gliadin.

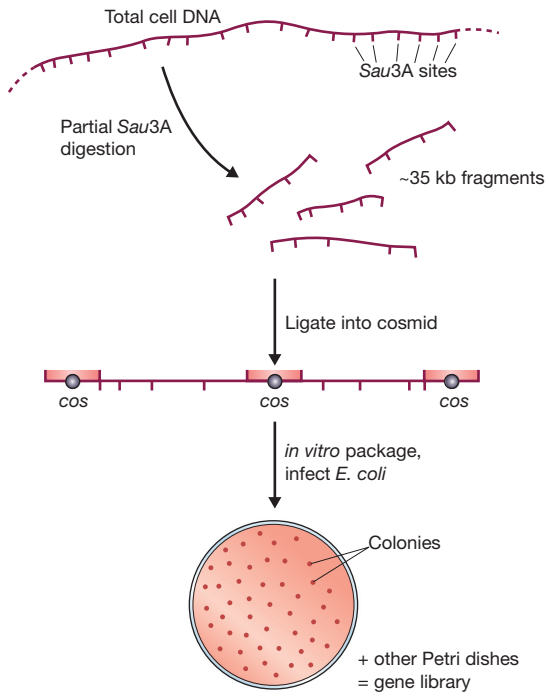


Figure 8.5

Preparation of a gene library in a cosmid vector.

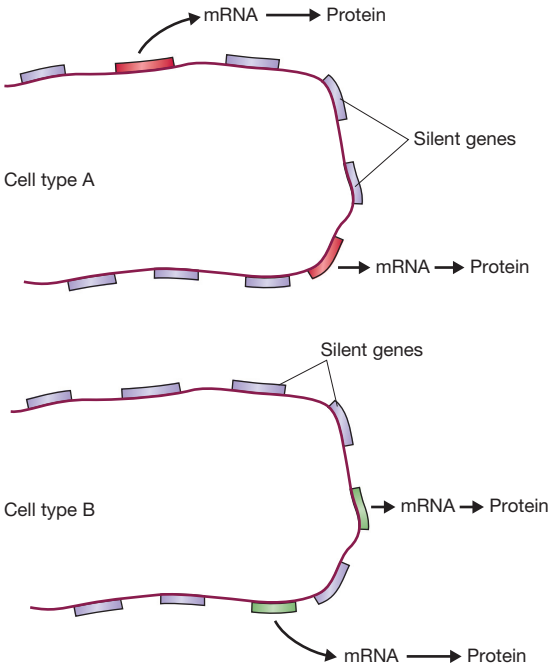


Figure 8.6

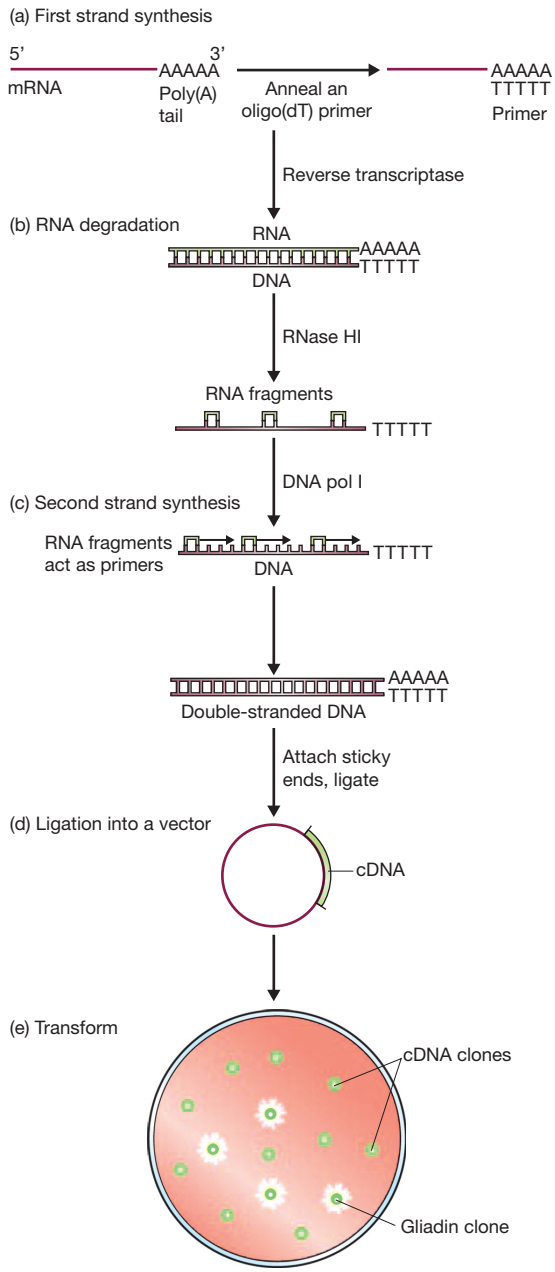
Different genes are expressed in different types of cell.

mRNA can be cloned as complementary DNA

Messenger RNA cannot itself be ligated into a cloning vector, but it can be converted into DNA by **complementary DNA (cDNA)** synthesis. The key to this method is the enzyme reverse transcriptase (p. 52), which synthesizes a DNA polynucleotide complementary to an existing RNA strand (Figure 8.7a). Once the cDNA strand has been synthesized, the RNA member of the hybrid molecule can be partially degraded by treating with ribonuclease HI (Figure 8.7b). The remaining RNA fragments then serve as primers

Figure 8.7

One possible scheme for cDNA cloning. First strand synthesis is primed by oligo(dT), a short oligonucleotide made up entirely of deoxythymidine nucleotides, which base pairs to the poly(A) tail present at the 3′ end of a eukaryotic mRNA.



(p. 51) for DNA polymerase I, which synthesizes the second cDNA strand (Figure 8.7c). The result is a double-stranded DNA fragment that can be ligated into a vector and cloned (Figure 8.7d).

The resulting cDNA clones are representative of the mRNA present in the original preparation. In the case of mRNA prepared from wheat seeds, the cDNA library would contain a large proportion of clones representing gliadin mRNA (Figure 8.7e). Other clones will also be present, but locating the cloned gliadin cDNA is a much easier process than identifying the equivalent gene from a complete wheat genomic library.

8.4 Methods for clone identification

Once a suitable library has been prepared, a number of procedures can be employed to attempt identification of the desired clone. Although a few of these procedures are based on detecting the translation product of the cloned gene, it is usually easier to identify directly the correct recombinant DNA molecule. This can be achieved by the important technique of **hybridization probing**.

8.4.1 Complementary nucleic acid strands hybridize to each other

Any two single-stranded nucleic acid molecules have the potential to form base pairs with one another. With most pairs of molecules, the resulting hybrid structures are unstable as only a small number of individual interstrand bonds are formed (Figure 8.8a). If the polynucleotides are complementary, extensive base pairing can occur to form a stable double-stranded molecule (Figure 8.8b). Not only can this occur between single-stranded DNA molecules to form the DNA double helix, but also between a pair of single-stranded RNA molecules or between combinations of one DNA strand and one RNA strand (Figure 8.8c).

Nucleic acid hybridization can be used to identify a particular recombinant clone if a DNA or RNA probe, complementary to the desired gene, is available. The exact nature of the probe will be discussed later in this chapter. First we must consider the technique itself.

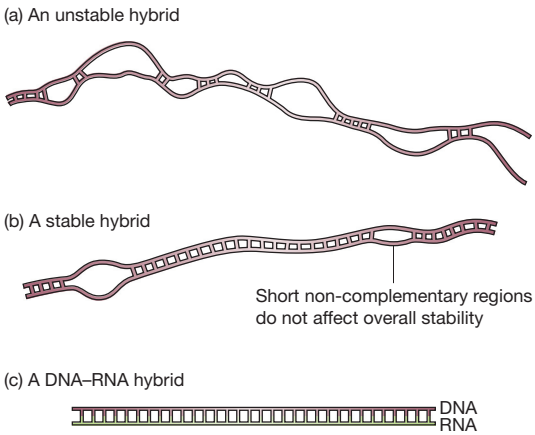


Figure 8.8

Nucleic acid hybridization. (a) An unstable hybrid molecule formed between two non-homologous DNA strands. (b) A stable hybrid formed between two complementary strands. (c) A DNA–RNA hybrid, such as may be formed between a gene and its transcript.

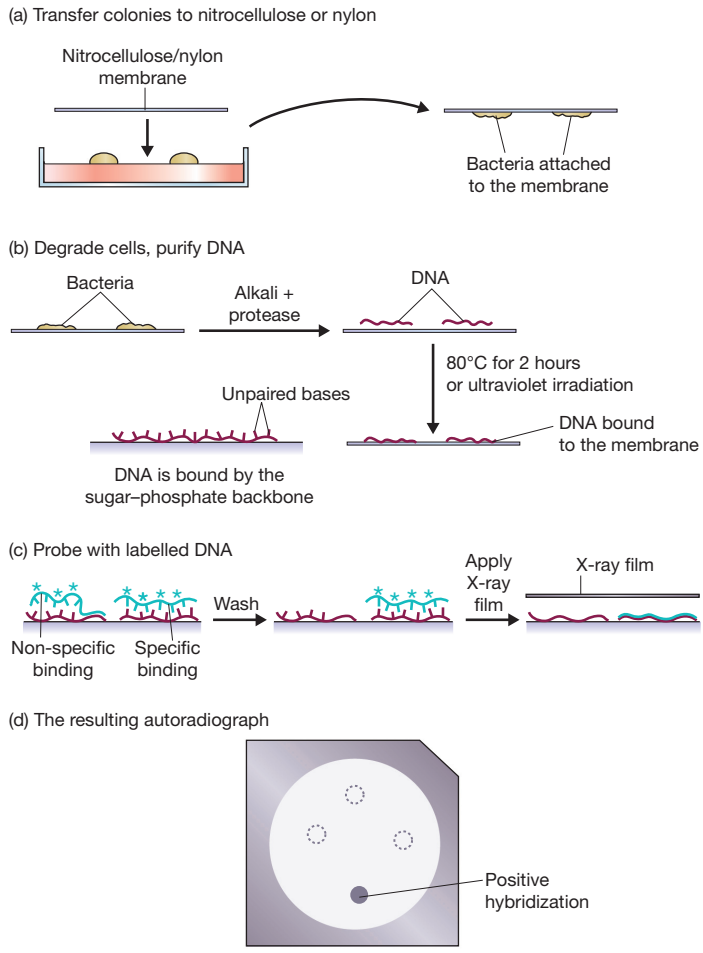


Figure 8.9

Colony hybridization probing. In this example, the probe is labelled with a radioactive marker and hybridization detected by autoradiography, but other types of label and detection system can also be used.

8.4.2 Colony and plaque hybridization probing

Hybridization probing can be used to identify recombinant DNA molecules contained in either bacterial colonies or bacteriophage plaques. First, the colonies or plaques are transferred to a nitrocellulose or nylon membrane (Figure 8.9a), and then treated to remove all contaminating material, leaving just DNA (Figure 8.9b). Usually, this treatment also results in denaturation of the DNA molecules, so that the hydrogen bonds between individual strands in the double helix are broken. These single-stranded molecules can then be bound tightly to the membrane by heating for a short period at 80 °C if a nitrocellulose membrane is being used, or with a nylon membrane by ultraviolet (UV) irradiation. The molecules become attached to the membrane through their sugar–phosphate backbones, which means that the bases are free to pair with complementary nucleic acid molecules.

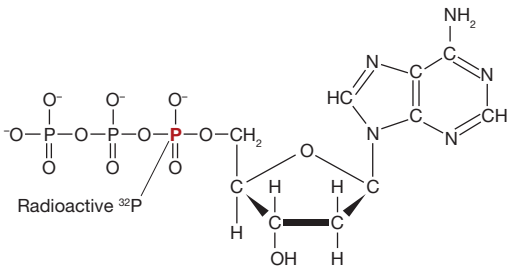


Figure 8.10
The structure of α-^{32}P-deoxyadenosine triphosphate ([α-^{32}P]dATP).

The probe must now be **labelled** with a radioactive or other type of marker, denatured by heating, and applied to the membrane in a solution of chemicals that promote nucleic acid hybridization (Figure 8.9c). After a period to allow hybridization to take place, the filter is washed to remove any unbound probe, dried, and the label detected in order to identify the colonies or plaques to which the probe has become bound (Figure 8.9d).

Labelling with a radioactive marker

A DNA molecule is often labelled by incorporating nucleotides that carry a radioactive isotope of phosphorus, ^{32}P (Figure 8.10). Several methods are available:

- **Nick translation**. Most purified samples of DNA contain some nicked molecules, however carefully the preparation has been carried out, which means that DNA polymerase I is able to attach to the DNA and catalyse a strand replacement reaction (Figure 8.11a). This reaction requires a supply of nucleotides; if one of these nucleotides is radioactively labelled the DNA molecule will itself become labelled. Nick translation can be used to label any DNA molecule, but might under some circumstances also cause DNA cleavage.
- **End filling** is a gentler method than nick translation and rarely causes breakage of the DNA, but unfortunately it can only be used to label DNA molecules that have sticky ends. The enzyme used is the Klenow fragment (p. 51), which 'fills in' a sticky end by synthesizing the complementary strand (Figure 8.11b)). As with nick

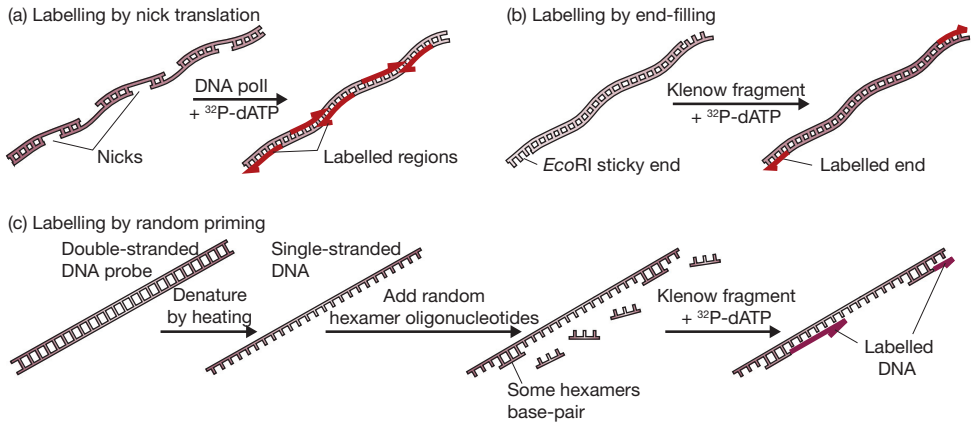


Figure 8.11
Methods for labelling DNA.

translation, if the end-filling reaction is carried out in the presence of labelled nucleotides, the DNA becomes labelled.

- **Random priming** results in a probe with higher activity and is therefore able to detect smaller amounts of membrane-bound DNA. The denatured DNA is mixed with a set of hexameric oligonucleotides of random sequence. By chance, these random hexamers will contain a few molecules that will base pair with the probe and prime new DNA synthesis. The Klenow fragment is used as this enzyme lacks the nuclease activity of DNA polymerase I (p. 51) and so only fills in the gaps between adjacent primers (Figure 8.11c). Labelled nucleotides are incorporated into the new DNA that is synthesized.

After hybridization, the location of the bound probe is detected by **autoradiography**. For this, a sheet of X-ray-sensitive photographic film is placed over the membrane. The radioactive DNA then exposes the film, which is developed to reveal the positions of the colonies or plaques to which the probe has hybridized (see Figure 8.9d).

Non-radioactive labelling

Today, radioactive labelling methods are beginning to fall out of favour, partly because of health hazards to the research workers and partly because of problems associated with the disposal of radioactive waste. Alternatively, the hybridization probe can be labelled in a non-radioactive manner. A number of methods have been developed for this purpose, two of which are illustrated in Figure 8.12. The first method utilizes deoxyuridine triphosphate (dUTP) nucleotides which have been modified by reaction with **biotin**, an organic molecule that has a high affinity for a protein called **avidin**. After hybridization, the positions of the bound biotinylated probe can be determined by washing with avidin coupled to a fluorescent marker (Figure 8.12a). This method is as sensitive as radioactive probing and is becoming increasingly popular.

The same is true for a second method for non-radioactive hybridization probing, in which the probe DNA is first complexed with the enzyme **horseradish peroxidase**. The probe DNA is then detected through the enzyme's ability to degrade luminol, with the emission of chemiluminescence (Figure 8.12b). The signal can be recorded on normal photographic film in a manner analogous to autoradiography.

8.4.3 Examples of the practical use of hybridization probing

Clearly, the success of colony or plaque hybridization as a means of identifying a particular recombinant clone depends on the availability of a DNA molecule that can be used as a probe. This probe must share at least a part of the sequence of the cloned gene. If the gene itself is not available (which presumably is the case if the aim of the experiment is to provide a clone of it), then what can be used as the probe?

In practice, the nature of the probe is determined by the information available about the desired gene. We will consider three possibilities:

- Where the desired gene is expressed at a high level in a cell type from which a cDNA clone library has been prepared.
- Where the amino acid sequence of the protein coded by the gene is completely or partially known.
- Where the gene is a member of a family of related genes.

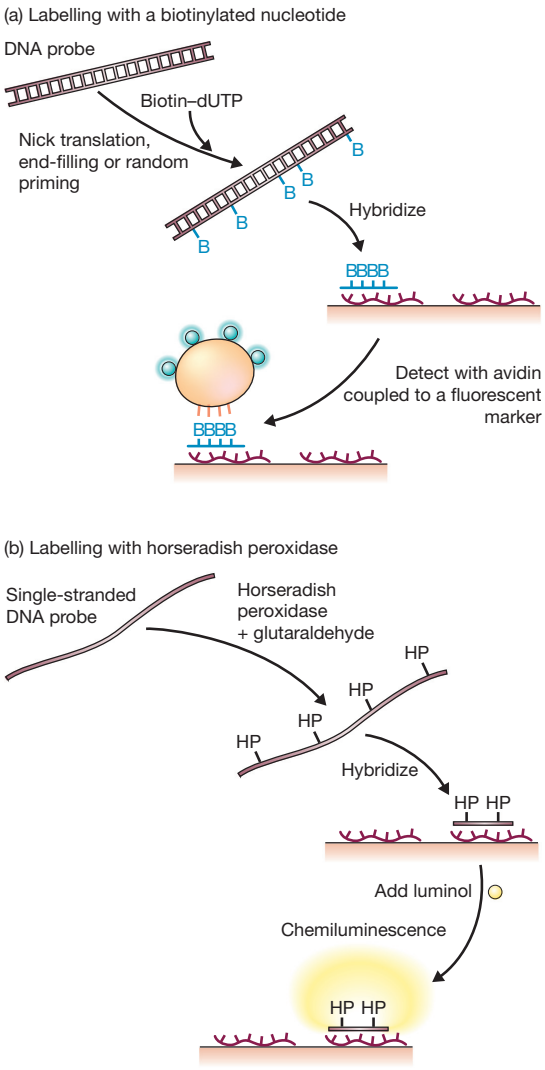


Figure 8.12

Two methods for the non-radioactive labelling of DNA probes.

Abundancy probing to analyse a cDNA library

As described earlier in this chapter, a cDNA library is often prepared in order to obtain a clone of a gene expressed at a relatively high level in a particular cell type. In the example of a cDNA library from developing wheat seeds, a large proportion of the clones are copies of the mRNA transcripts of the gliadin gene (see Figure 8.7e).

Identification of the gliadin clones is simply a case of using individual cDNA clones from the library to probe all the other members of the library (Figure 8.13). A clone is selected at random and the recombinant DNA molecule purified, labelled, and used to probe the remaining clones. This is repeated with different clones as probes until one that hybridizes to a large proportion of the library is obtained. This abundant cDNA is considered a possible gliadin clone and analysed in greater detail (e.g., by DNA sequencing and isolation of the translation product) to confirm the identification.

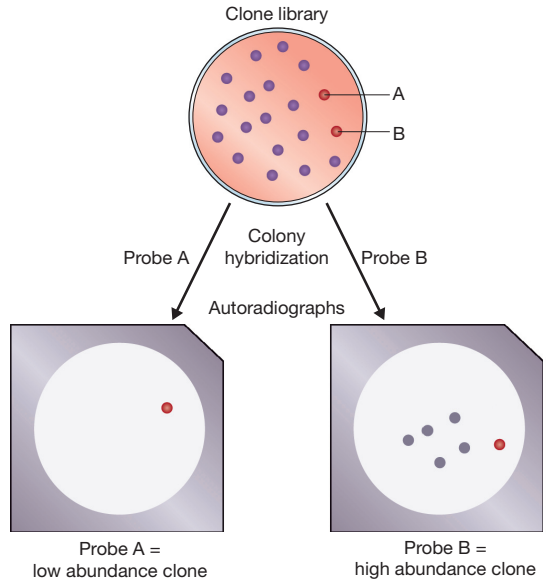


Figure 8.13
Probing within a library to identify an abundant clone.

Oligonucleotide probes for genes whose translation products have been characterized

Often, the gene to be cloned codes for a protein that has already been studied in some detail. In particular, the amino acid sequence of the protein might have been determined. If the amino acid sequence is known, then it is possible to use the genetic code to predict the nucleotide sequence of the relevant gene. This prediction is always an approximation, as only methionine and tryptophan can be assigned unambiguously to triplet codons, all other amino acids being coded by at least two codons each. Nevertheless, in most cases, the different codons for an individual amino acid are related. Alanine, for example, is coded by GCA, GCC, GCG, and GCT, so two out of the three nucleotides of the triplet coding for alanine can be predicted with certainty.

As an example to clarify how these predictions are made, consider cytochrome *c*, a protein that plays an important role in the respiratory chain of all aerobic organisms. The cytochrome *c* protein from yeast was sequenced in 1963, with the result shown in Figure 8.14. This sequence contains a segment, starting at amino acid 59, that runs Trp–Asp–Glu–Asn–Asn–Met. The genetic code states that this hexapeptide is coded

```
                                                                  15
GLY–SER–ALA–LYS–LYS–GLY–ALA–THR–LEU–PHE–LYS–THR–ARG–CYS–GLU–
                                                                  30
LEU–CYS–HIS–THR–VAL–GLU–LYS–GLY–GLY–PRO–HIS–LYS–VAL–GLY–PRO–
                                                                  45
ASN–LEU–HIS–GLY–ILE–PHE–GLY–ARG–HIS–SER–GLY–GLN–ALA–GLN–GLY–
                                                                  60
TYR–SER–TYR–THR–ASP–ALA–ASN–ILE–LYS–LYS–ASN–VAL–LEU–TRP–ASP–
                                                                  75
GLU–ASN–ASN–MET–SER–GLU–TYR–LEU–THR–ASN–PRO–LYS–LYS–TYR–ILE–
                                                                  90
PRO–GLY–THR–LYS–MET–ALA–PHE–GLY–GLY–LEU–LYS–LYS–GLU–LYS–ASP–
                                                                 103
ARG–ASN–ASP–LEU–ILE–THR–TYR–LEU–LYS–LYS–ALA–CYS–GLU
```

Figure 8.14
The amino acid sequence of yeast cytochrome *c*. The hexapeptide that is highlighted red is the one used to illustrate how a nucleotide sequence can be predicted from an amino acid sequence.

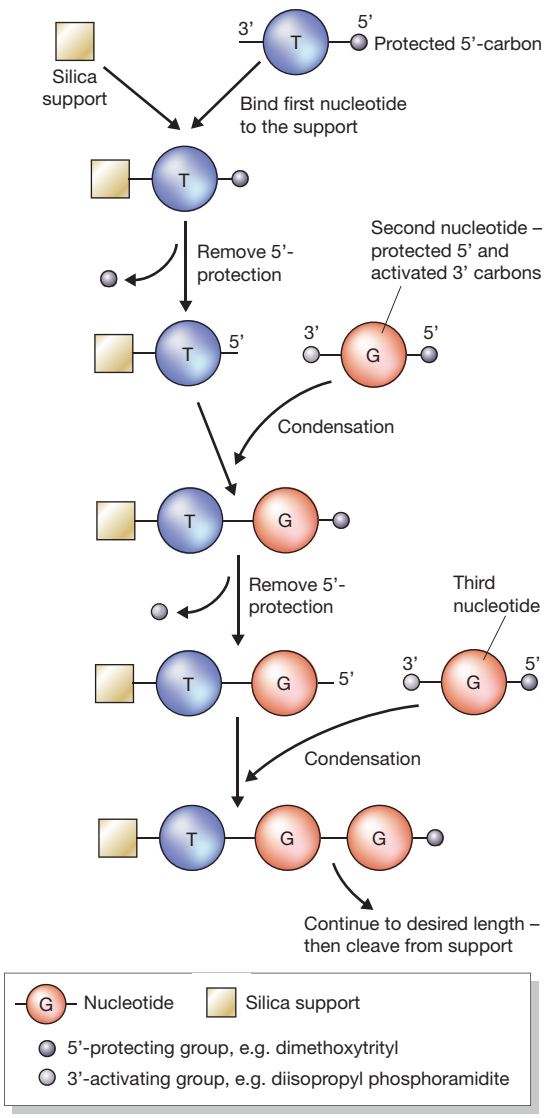


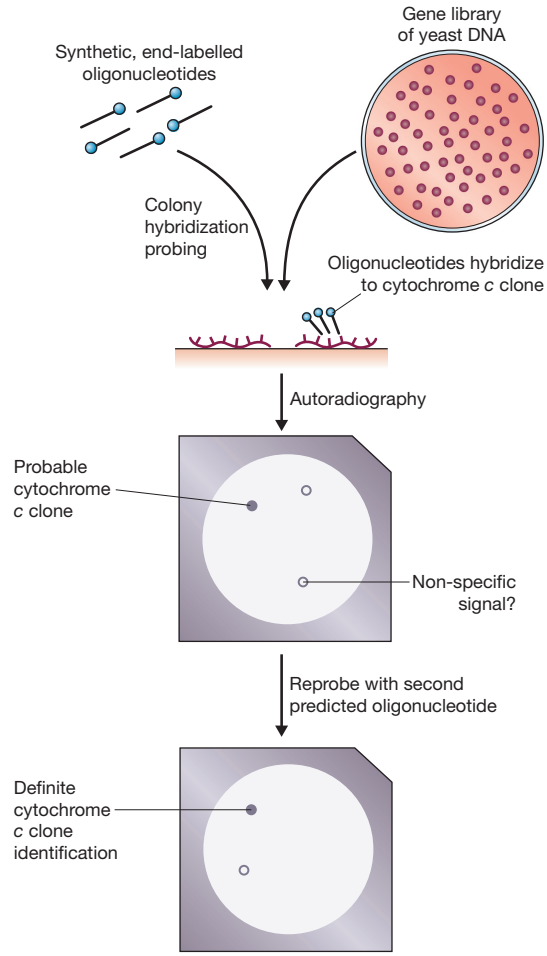
Figure 8.15

A simplified scheme for oligonucleotide synthesis. Each nucleotide is modified by attachment of an activating group to the 3′ carbon and a protecting group to the 5′ carbon. The activating group enables the normally inefficient process of nucleotide joining to proceed much more rapidly. The protecting group ensures that individual nucleotides cannot attach to one another, and instead react only with the terminal 5′ group of the growing oligonucleotide, this 5′ group being deprotected by chemical treatment at the appropriate point in each synthesis cycle.

by TGG–GA$^{T}/_{C}$–GA$^{A}/_{G}$–AA$^{T}/_{C}$–AA$^{T}/_{C}$–ATG. Although this represents a total of 16 different possible sequences, 14 of the 18 nucleotides can be predicted with certainty.

Oligonucleotides of up to about 150 nucleotides in length can easily be synthesized in the laboratory (Figure 8.15). An oligonucleotide probe could therefore be constructed according to the predicted nucleotide sequence, and this probe might be able to identify the gene coding for the protein in question. In the example of yeast cytochrome *c*, the 16 possible oligonucleotides that can code for Trp–Asp–Glu–Asn–Asn–Met would be synthesized, either separately or as a pool, and then used to probe a yeast genomic or cDNA library (Figure 8.16). One of the oligonucleotides in the probe will have the correct sequence for this region of the cytochrome *c* gene, and its hybridization signal will indicate which clones carry this gene.

Figure 8.16
The use of synthetic, end-labelled oligonucleotides to identify a clone of the yeast cytochrome *c* gene.



The result can be checked by carrying out a second probing with a mixture of oligonucleotides whose sequences are predicted from a different segment of the cytochrome *c* protein (Figure 8.16). However, the segment of the protein used for nucleotide sequence prediction must be chosen with care. The hexapeptide Ser–Glu–Tyr–Leu–Thr–Asn, which immediately follows our first choice, could be coded by several thousand different 18-nucleotide sequences, clearly is an unsuitable choice for a synthetic probe.

Heterologous probing allows related genes to be identified

Often, a substantial amount of nucleotide similarity is seen when two genes for the same protein, but from different organisms, are compared, this being a reflection of the conservation of gene structure during evolution. Frequently, two genes from related organisms are sufficiently similar for a single-stranded probe prepared from one gene to form a stable hybrid with the second gene. Although the two molecules are not entirely complementary, enough base pairs are formed to produce a stable structure (Figure 8.17a).

Heterologous probing makes use of hybridization between related sequences for clone identification. For example, the yeast cytochrome *c* gene, as identified in the

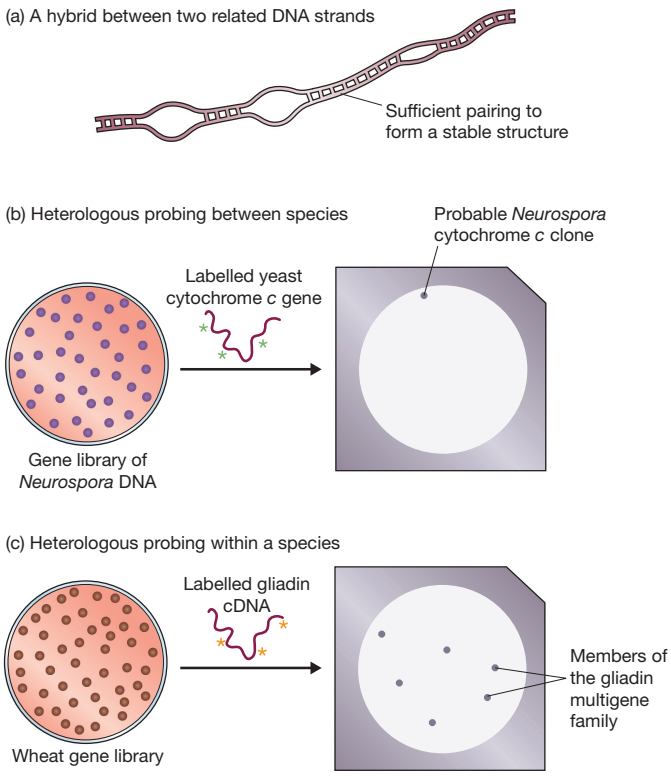


Figure 8.17
Heterologous probing.

previous section by oligonucleotide probing, could itself be used as a hybridization probe to identify cytochrome *c* genes in clone libraries of other organisms. A probe prepared from the yeast gene would not be entirely complementary to the gene from, say, the fungus *Neurospora crassa*, but sufficient base pairing should occur for a hybrid to be formed (Figure 8.17b).

Heterologous probing can also identify related genes in the same organism. If the wheat gliadin cDNA clone, as identified earlier in the chapter by abundancy probing, is used to probe a genomic library, it will hybridize not only to its own gene but also to a variety of other genes (Figure 8.17c). These genes are all related to the gliadin cDNA, but have slightly different nucleotide sequences. This is because the wheat gliadins form a complex group of related proteins that are coded by the members of a **multigene family**. Once one gene in the family has been cloned, all the other members can be isolated by heterologous probing.

Southern hybridization enables a specific restriction fragment containing a gene to be identified

As well as colony and plaque hybridization analysis, there are also occasions when it is necessary to use hybridization probing to identify which of a series of restriction fragments contains a gene of interest. As an example, we will return to the genomic clone of

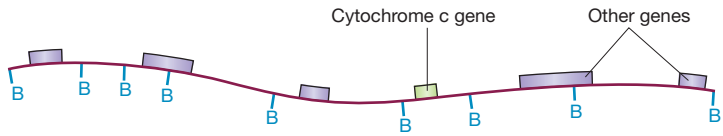


Figure 8.18

A long cloned DNA fragment may contain several genes in addition to the one in which we are interested. B = *Bam*HI restriction site.

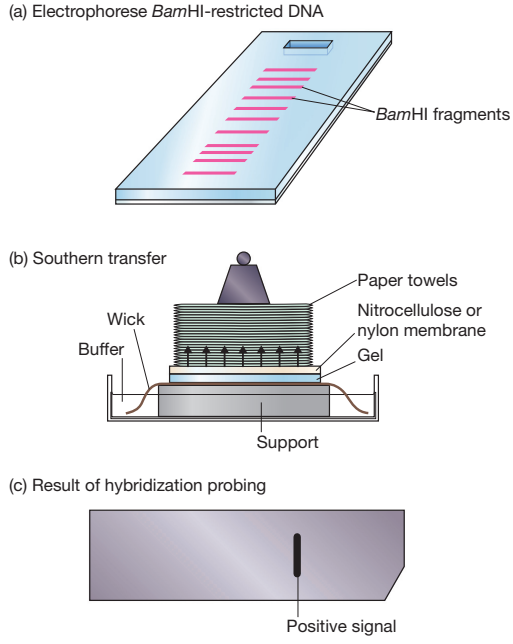
the yeast cytochrome *c* gene, which we identified by oligonucleotide hybridization probing. Let us imagine that this particular genomic library was prepared by partial restriction of yeast DNA with *Bam*HI followed by cloning in the cosmid vector pJB8 (p. 103). The cloned fragment containing the cytochrome *c* gene will therefore be approximately 40 kb in length, and will probably contain about ten *Bam*HI fragments, remembering that the hexanucleotide recognition site for this enzyme will be present, on average, once every $4^6 = 4096$ bp.

The cytochrome *c* gene, on the other hand, is predicted to be just 309 bp in length (we know that the protein has 103 amino acids; see Figure 8.14). The gene therefore comprises less than 1% of the cloned DNA fragment, and it is quite possible that other genes that we are not interested in are also present in this insert (Figure 8.18). The method called **Southern hybridization** enables the individual restriction fragment containing the cytochrome *c* gene to be identified.

The first step in using Southern hybridization for this purpose would be to digest the clone with *Bam*HI and to then separate the restriction fragments by electrophoresis in an agarose gel (Figure 8.19a). The aim is to use the oligonucleotide probe for cytochrome *c* to identify the fragment that contains the gene. This can be attempted

Figure 8.19

Southern hybridization.



while the restriction fragments are still contained in the electrophoresis gel, although the results are usually not very good as the gel matrix causes a lot of spurious background hybridization that obscures the specific hybridization signal. Instead, the DNA bands in the agarose gel are transferred to a nitrocellulose or nylon membrane, which provides a much cleaner environment for the hybridization experiment.

The transfer of DNA bands from an agarose gel to a membrane employs the technique perfected in 1975 by Professor E.M. Southern, and referred to as **Southern transfer**. The membrane is placed on the gel, through which buffer is allowed to soak, carrying the DNA from the gel to the membrane where it becomes bound. Whilst sophisticated pieces of apparatus can be purchased to assist this process, many molecular biologists prefer a homemade set-up incorporating many paper towels and considerable balancing skills (Figure 8.19b). The same method can also be used for the transfer of RNA molecules (**northern transfer**) or proteins (**western transfer**). So far, no one has devised eastern transfers.

Southern transfer results in a membrane that carries a replica of the DNA bands from the agarose gel. If the labelled probe is now applied, hybridization occurs and autoradiography (or the equivalent detection system for a non-radioactive probe) reveals which restriction fragment contains the cloned gene (Figure 8.19c).

8.4.4 Identification methods based on detection of the translation product of the cloned gene

Hybridization probing is usually the preferred method for the identification of a particular recombinant from a clone library. The technique is easy to perform and, with modifications introduced in recent years, can be used to check up to 10 000 recombinants per experiment, allowing large genomic libraries to be screened in a reasonably short time. Nevertheless, the requirement for a probe that is at least partly complementary to the desired gene sometimes makes it impossible to use hybridization in clone identification. On these occasions, a different strategy is needed.

The main alternative to hybridization probing is **immunological screening**. The distinction here is that, whereas with hybridization probing the cloned DNA fragment is itself directly identified, an immunological method detects the protein coded by the cloned gene. Immunological techniques therefore presuppose that the cloned gene is being expressed, so that the protein is being made, and that this protein is not normally present in the host cells.

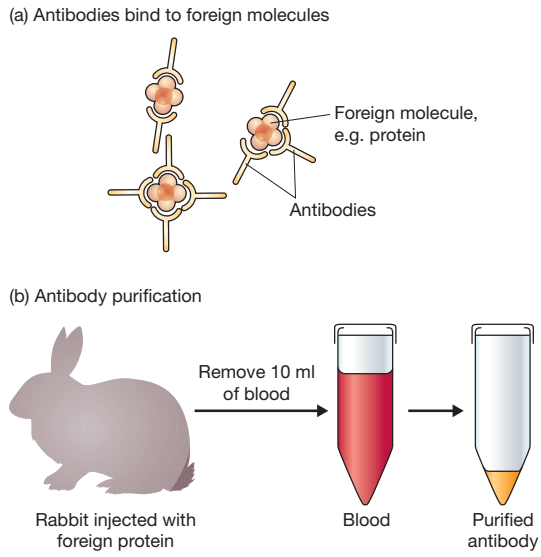
Antibodies are required for immunological detection methods

If a purified sample of a protein is injected into the bloodstream of a rabbit, the immune system of the animal responds by synthesizing antibodies that bind to and help degrade the foreign molecule (Figure 8.20a). This is a version of the natural defence mechanism that the animal uses to deal with invasion by bacteria, viruses and other infective agents.

Once a rabbit is challenged with a protein, the levels of antibody present in its bloodstream remain high enough over the next few days for substantial quantities to be purified. It is not necessary to kill the rabbit, because as little as 10 ml of blood will provide a considerable amount of antibody (Figure 8.20b). This purified antibody binds only to the protein with which the animal was originally challenged.

Figure 8.20

Antibodies. (a) Antibodies in the bloodstream bind to foreign molecules and help degrade them. (b) Purified antibodies can be obtained from a small volume of blood taken from a rabbit injected with the foreign protein.

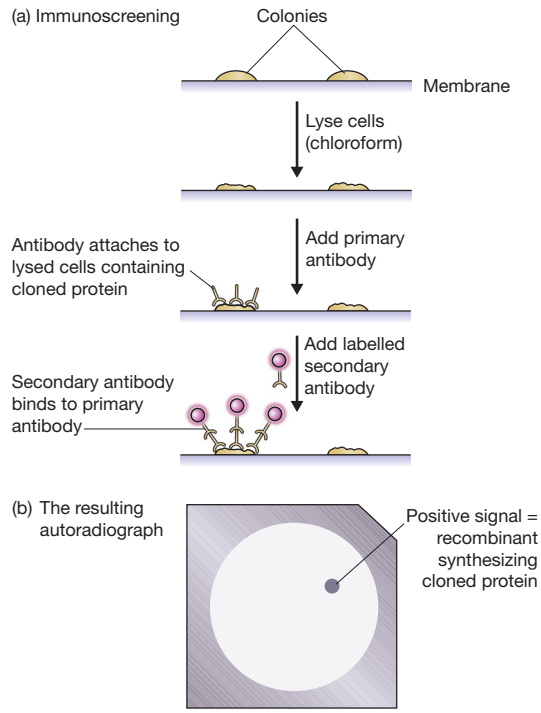


Using a purified antibody to detect protein in recombinant colonies

There are several versions of immunological screening, but the most useful method is a direct counterpart of colony hybridization probing. Recombinant colonies are transferred to a polyvinyl or nitrocellulose membrane, the cells are lysed, and a solution containing the specific antibody is added (Figure 8.21a). In the original methods, either the antibody itself was labelled, or the membrane was subsequently washed with a solution

Figure 8.21

Using a purified antibody to detect protein in recombinant colonies. Instead of using a secondary antibody, the primary antibody itself can be labelled, or the antibody can be detected with labelled protein A.



of labelled **protein A**, a bacterial protein that binds specifically to the immunoglobulins of which antibodies are made. In more modern methods, the bound antibody – the **primary antibody** – is detected by washing the membrane with a labelled **secondary antibody**, which binds specifically to the primary antibody. The secondary antibody is prepared by injecting the primary antibody into an animal, of a different species to the one from which the primary antibody was prepared. For example, if the primary antibody was prepared in a rabbit, the secondary antibody could be obtained by injecting a sample of the primary antibody into a goat. The goat's immune system will look on the primary antibody as a foreign protein antigen, and synthesize the secondary antibody to bind to it.

Several secondary antibody molecules can bind to a single primary antibody molecule, increasing the amount of signal that is produced and enabling a clearer detection of each positive colony. In all three methods, the label can be radioactive, in which case the colonies that bind the label can be detected using autoradiography (Figure 8.21b), or non-radioactive labels that result in a fluorescent or chemiluminescent signal can be used as an alternative.

The problem of gene expression

Immunological screening depends on the cloned gene being expressed so that the protein translation product is present in the recombinant cells. As will be discussed in greater detail in Chapter 13, a gene from one organism is often not expressed in a different organism. In particular, it is very unlikely that a cloned animal or plant gene will be expressed in *E. coli* cells. This problem can be circumvented by using a special type of vector, called an **expression vector** (p. 250), designed specifically to promote expression of the cloned gene in a bacterial host. Immunological screening of recombinant *E. coli* colonies carrying animal genes cloned into expression vectors has been very useful in identifying genes for several important hormones.

Further reading

Benton, W.D. and Davis, R.W. (1977) Screening λgt recombinant clones by hybridization to single plaques *in situ*. *Science*, **196**, 180–182.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for labelling DNA restriction fragments to high specific activity. *Analytical Biochemistry*, **132**, 6 –13. [Random priming labelling.]

Grunstein, M. and Hogness, D.S. (1975) Colony hybridization: a method for the isolation of cloned cDNAs that contain a specific gene. *Proceedings of the National Academy of Sciences of the United States of America*, **72**, 3961–3965.

Gubler, U. and Hoffman, B.J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene*, **25**, 263–269.

Southern, E.M. (2000) Blotting at 25. *Trends in Biochemical Science*, **25**, 585–588. [The origins of Southern hybridization.]

Thorpe, G.H.G., Kricka, L.J., Moseley, S.B., and Whitehead, T.P. (1985) Phenols as enhancers of the chemiluminescent horseradish peroxidase–luminol–hydrogen peroxide reaction: application in luminescence-monitored enzyme immunoassays. *Clinical Chemistry*, **31**, 1335–1341. [Describes the basis to a non-radioactive labelling method.]

Young, R.A. and Davis, R.W. (1983) Efficient isolation of genes by using antibody probes. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 1194–1198.

Chapter 9

The Polymerase Chain Reaction

Chapter contents

CHAPTER CONTENTS

As a result of the last seven chapters we have become familiar not only with the basic principles of gene cloning, but also with fundamental molecular biology techniques such as restriction analysis, gel electrophoresis, DNA labelling, and DNA–DNA hybridization. To complete our basic education in DNA analysis we must now return to the second major technique for studying genes, the polymerase chain reaction (PCR).

PCR is a very uncomplicated technique. All that happens is that a short region of a DNA molecule – a single gene for instance – is copied many times by a DNA polymerase enzyme (see Figure 1.2). This might seem a rather trivial exercise, but it has a multitude of applications in genetics research and in broader areas of biology.

We begin this chapter with an outline of PCR in order to understand exactly what it achieves. Then, we will look at the key issues that determine whether or not an individual PCR experiment is successful, before examining some of the methods that have been devised for studying the amplified DNA fragments that are obtained.

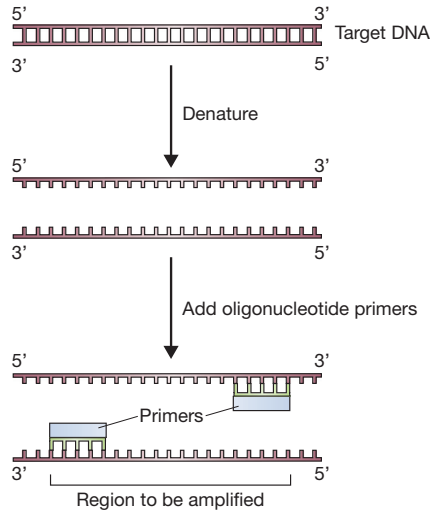
9.1 PCR in outline

PCR results in a selective amplification of a chosen region of a DNA molecule. Any region of any DNA molecule can be chosen, as long as the sequences at the borders of

Gene Cloning and DNA Analysis: An Introduction, Seventh Edition. T.A. Brown.

Companion Website: www.wiley.com/go/brown/cloning

Figure 9.1
Hybridization of the oligonucleotide primers to the template DNA at the beginning of a PCR.



the region are known. The border sequences must be known because, in order to carry out a PCR, two short oligonucleotides must hybridize to the DNA molecule, one to each strand of the double helix (Figure 9.1). These oligonucleotides, which act as primers for the DNA synthesis reactions, delimit the region that will be amplified.

Amplification is usually carried out by the DNA polymerase I enzyme from *Thermus aquaticus*. As mentioned elsewhere (p. 52), this organism lives in hot springs and many of its enzymes, including *Taq* polymerase, are thermostable, which means they are resistant to denaturation by heat treatment. As will be apparent very soon, the thermostability of *Taq* polymerase is an essential requirement in PCR methodology.

To carry out a PCR experiment, the target DNA is mixed with *Taq* polymerase, the two oligonucleotide primers, and a supply of nucleotides. The amount of target DNA can be very small because PCR is extremely sensitive and will operate with just a single starting molecule. The reaction is started by heating the mixture to 94 °C, at which temperature the hydrogen bonds holding together the two polynucleotides of the double helix are broken, so that the target DNA becomes denatured into single-stranded molecules (Figure 9.2). The temperature is then reduced to 50–60 °C, which results in some rejoining of the single strands of the target DNA, but also allows the primers to attach to their annealing positions. DNA synthesis can now begin, so the temperature is raised to 74 °C, just below the optimum for *Taq* polymerase. In this first stage of the PCR, a set of 'long products' is synthesized from each strand of the target DNA. These polynucleotides have identical 5′ ends but random 3′ ends, the latter representing positions where DNA synthesis terminates by chance.

The cycle of denaturation–annealing–synthesis is now repeated (Figure 9.3). The long products denature and the four resulting strands are copied during the DNA synthesis stage. This gives four double-stranded molecules, two of which are identical to the long products from the first cycle, and two of which are made entirely of new DNA. During the third cycle, the latter give rise to 'short products', the 5′ and 3′ ends of which are both set by the primer annealing positions. In subsequent cycles, the number of short products accumulates in an exponential fashion (doubling during each cycle) until one of the components of the reaction becomes depleted. This means that after 30 cycles, there will be over 130 million short products derived from each starting molecule.

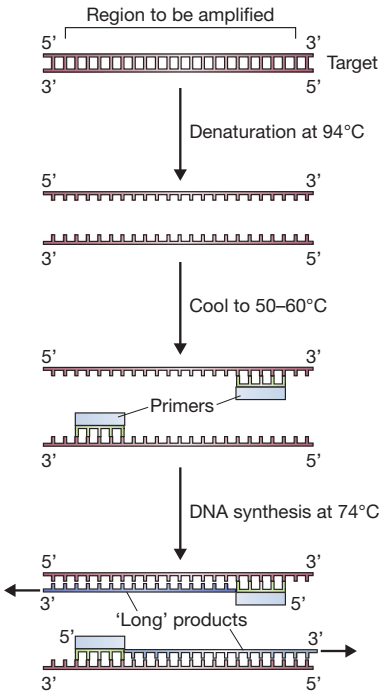


Figure 9.2
The first stage of a PCR, resulting in synthesis of the long products.

In real terms, this equates to several micrograms of PCR product from a few nanograms or less of target DNA.

At the end of a PCR, a sample of the reaction mixture is usually analysed using agarose gel electrophoresis, sufficient DNA having been produced for the amplified fragment to be visible as a discrete band after staining with ethidium bromide or some other DNA-binding dye. This may by itself provide useful information about the DNA region that has been amplified or, alternatively, the PCR product can be examined by techniques such as DNA sequencing.

9.2 PCR in more detail

Although PCR experiments are very easy to set up, they must be planned carefully if the results are to be of any value. The sequences of the primers are critical to the success of the experiment, as are the precise temperatures used in the heating and cooling stages of the reaction cycle.

9.2.1 Designing the oligonucleotide primers for a PCR

The primers are the key to the success or failure of a PCR experiment. If the primers are designed correctly the experiment results in amplification of a single DNA fragment, corresponding to the target region of the template molecule. If the primers are incorrectly designed the experiment will fail, possibly because no amplification occurs, or possibly because the wrong fragment, or more than one fragment, is amplified

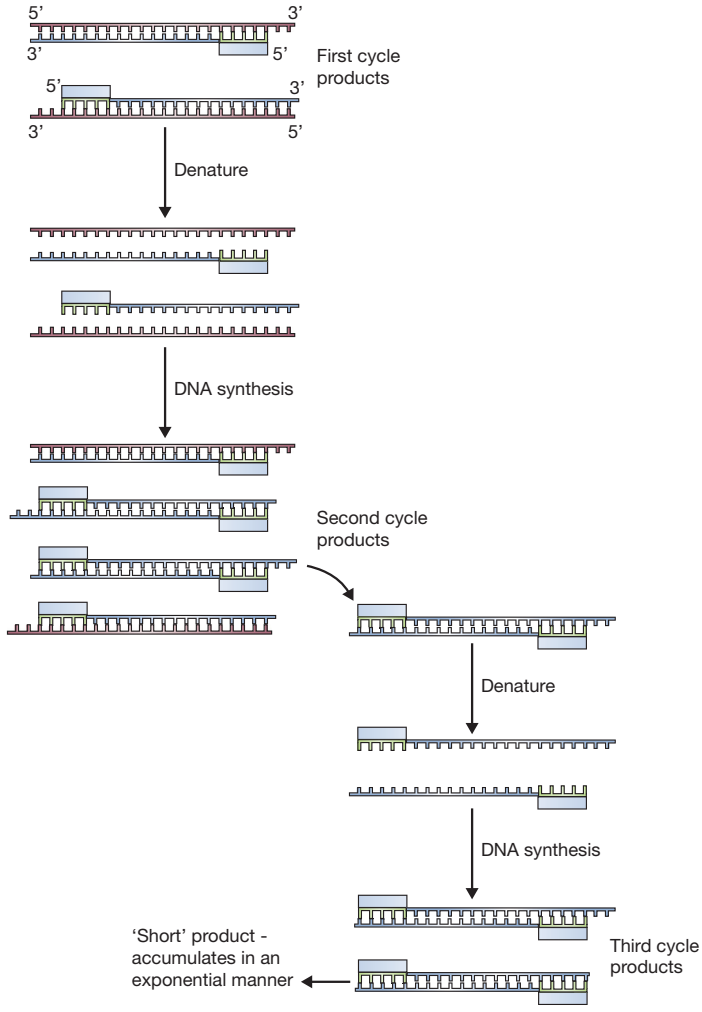


Figure 9.3

The second and third cycles of a PCR, during which the first short products are synthesized.

(Figure 9.4). Clearly, a great deal of thought must be put into the design of the primers.

Working out appropriate sequences for the primers is not a problem: they must correspond with the sequences flanking the target region on the template molecule. Each primer must, of course, be complementary (not identical) to its template strand in order for hybridization to occur, and the 3′ ends of the hybridized primers should point towards one another (Figure 9.5). The DNA fragment to be amplified should not be greater than about 3 kb in length and ideally less than 1 kb. Fragments up to 10 kb can be amplified using standard PCR techniques, but the longer the fragment the less efficient is the amplification and the more difficult it is to obtain consistent results. The amplification of very long fragments – up to 40 kb – is possible, but this requires special methods.

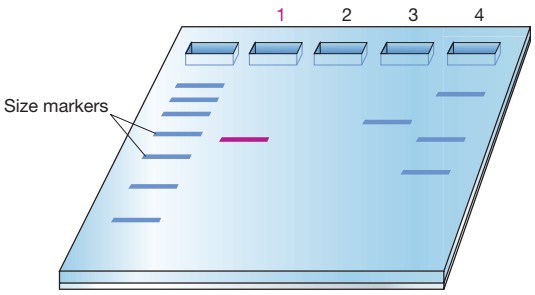


Figure 9.4

The results of PCRs with well-designed and poorly designed primers. Lane 1 shows a single amplified fragment of the expected size, the result of a well-designed experiment. In lane 2 there is no amplification product, suggesting that one or both of the primers were unable to hybridize to the template DNA. Lanes 3 and 4 show, respectively, an amplification product of the wrong size, and a mixture of products (the correct product plus two wrong ones). Both results are due to hybridization of one or both of the primers to non-target sites on the template DNA molecule.

The first important issue to address is the length of the primers. If the primers are too short they might hybridize to non-target sites and give undesired amplification products. To illustrate this point, imagine that total human DNA is used in a PCR experiment with a pair of primers eight nucleotides in length (in PCR jargon, these are called '8-mers'). The likely result is that a number of different fragments will be amplified. This is because attachment sites for these primers are expected to occur, on average, once every $4^8 = 65\,536$ bp, giving approximately 49 000 possible sites in the 3 200 000 kb of nucleotide sequence that makes up the human genome. This means that it would be very unlikely that a pair of 8-mer primers would give a single, specific amplification product with human DNA (Figure 9.6a).

What if the 17-mer primers shown in Figure 9.5 are used? The expected frequency of a 17-mer sequence is once every $4^{17} = 17\,179\,869\,184$ bp. This figure is more than fivefold greater than the length of the human genome, so a 17-mer primer would be expected to have just one hybridization site in total human DNA. A pair of 17-mer primers should therefore give a single, specific amplification product (Figure 9.6b).

Why not simply make the primers as long as possible? The length of the primer influences the rate at which it hybridizes to the template DNA, with long primers hybridizing at a slower rate. The efficiency of the PCR, measured by the number of amplified molecules produced during the experiment, is therefore reduced if the primers are too long, as complete hybridization to the template molecules cannot occur in the time allowed during the reaction cycle. In practice, primers longer than 30-mer are rarely used.

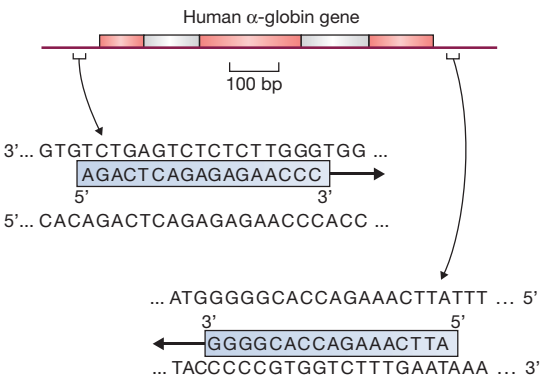
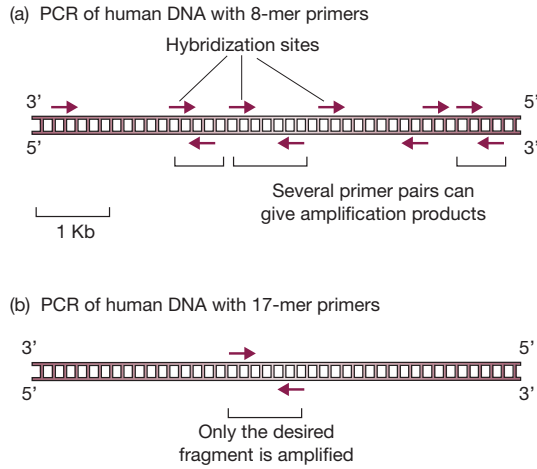


Figure 9.5

A pair of primers designed to amplify the human α_1-globin gene. The exons of the gene are shown as red boxes, the introns as grey boxes.

Figure 9.6
The lengths of the primers are critical for the specificity of the PCR.

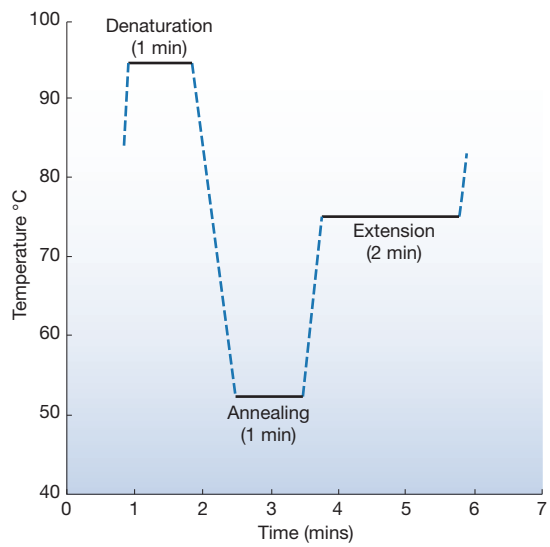


9.2.2 Working out the correct temperatures to use

During each cycle of a PCR, the reaction mixture is transferred between three temperatures (Figure 9.7):

- The denaturation temperature, usually 94 °C, which breaks the base pairs and releases single-stranded DNA to act as templates in the next round of DNA synthesis.
- The hybridization or annealing temperature, at which the primers attach to the templates.
- The extension temperature, at which DNA synthesis occurs. This is usually set at 74 °C, just below the optimum for *Taq* polymerase.

Figure 9.7
A typical temperature profile for a PCR.



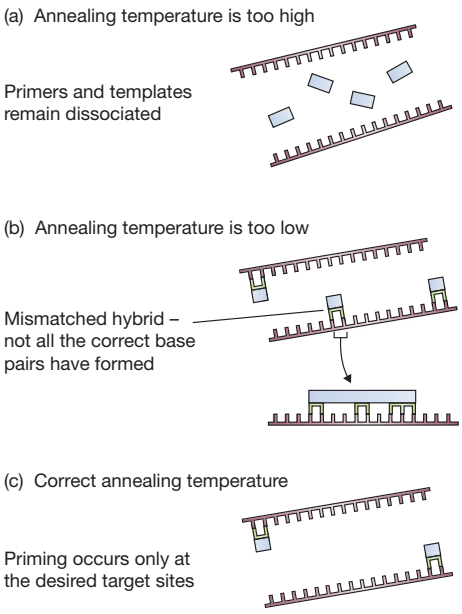


Figure 9.8
Temperature has an important effect on the hybridization of the primers to the template DNA.

The annealing temperature is the important one because, again, this can affect the specificity of the reaction. DNA–DNA hybridization is a temperature-dependent phenomenon. If the temperature is too high no hybridization takes place, and the primers and templates remain dissociated (Figure 9.8a). If the temperature is too low, mismatched hybrids – ones in which not all the correct base pairs have formed – are stable (Figure 9.8b). If this occurs, the earlier calculations regarding the appropriate lengths for the primers become irrelevant, as it was assumed in these calculations that only perfect primer–template hybrids are able to form. If mismatches are tolerated, the number of potential hybridization sites for each primer is greatly increased, and amplification is more likely to occur at non-target sites in the template molecule.

The ideal annealing temperature must be low enough to enable hybridization between primer and template, but high enough to prevent mismatched hybrids from forming (Figure 9.8c). This temperature can be estimated by determining the **melting temperature** or T_m of the primer–template hybrid. The T_m is the temperature at which the correctly base-paired hybrid dissociates ('melts'). A temperature 1–2 °C below this should be low enough to allow the correct primer–template hybrid to form, but too high for a hybrid with a single mismatch to be stable. The T_m can be determined experimentally, but is more usually calculated from the simple formula (Figure 9.9):

$$T_m = (4 \times [G + C]) + (2 \times [A + T])°C$$

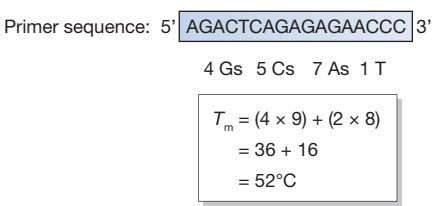


Figure 9.9
Calculating the T_m of a primer.

in which [G + C] is the number of G and C nucleotides in the primer sequence, and [A + T] is the number of A and T nucleotides.

The annealing temperature for a PCR experiment is therefore determined by calculating the T_m for each primer and using a temperature of 1–2 °C below this figure. Note that this means the two primers should be designed so that they have identical T_ms. If this is not the case, the appropriate annealing temperature for one primer may be too high or too low for the other member of the pair.

9.3 After the PCR: Studying PCR products

PCR is often the starting point for a longer series of experiments in which the amplification product is studied in various ways in order to obtain information about the DNA molecule that acted as the original template. We will encounter many studies of this type in Parts II and III of this book, when we examine the applications of gene cloning and PCR in research and biotechnology. Although a wide range of procedures have been devised for studying PCR products, three techniques are particularly important:

- Gel electrophoresis of PCR products.
- Cloning of PCR products.
- Sequencing of PCR products.

The first two of these techniques are dealt with in this chapter. The third technique is deferred until Chapter 10, when all aspects of DNA sequencing will be covered.

9.3.1 Gel electrophoresis of PCR products

The results of most PCR experiments are checked by running a portion of the amplified reaction mixture in an agarose gel. A band representing the amplified DNA may be visible after staining, or if the DNA yield is low the product can be detected using Southern hybridization (p. 151). If the expected band is absent, or if additional bands are present, then something has gone wrong and the experiment must be repeated.

In some cases, agarose gel electrophoresis is used not only to determine if a PCR experiment has worked, but also to obtain additional information. For example, the presence of restriction sites in the amplified region of the template DNA can be determined by treating the PCR product with a restriction endonuclease before running the sample in the agarose gel (Figure 9.10). This is a type of **restriction fragment length polymorphism (RFLP) analysis**.

Alternatively, the exact size of the PCR product can be used to establish if the template DNA contains an insertion or deletion mutation in the amplified region (Figure 9.10). Length mutations of this type form the basis of **DNA profiling**, a central technique in forensic science (Chapter 16).

In some experiments, the mere presence or absence of the PCR product is the diagnostic feature. An example of this occurs when PCR is used as the screening procedure to identify a desired gene from a genomic or cDNA library. Carrying out PCRs with every clone in a genomic library might seem a tedious task, but one of the main advantages of PCR is that individual experiments are quick to set up and many PCRs can be performed in parallel. The workload can also be reduced by **combinatorial screening**, an example of which is shown in Figure 9.11.

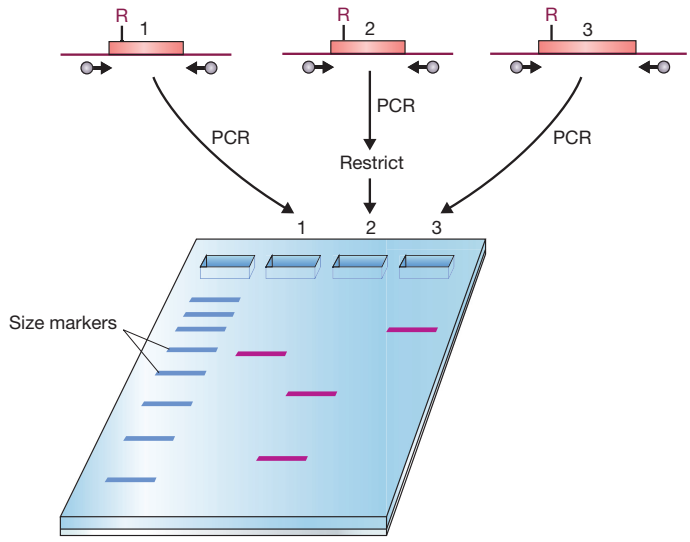


Figure 9.10

Gel electrophoresis of the PCR product can provide information on the template DNA molecule. Lanes 1 and 2 show, respectively, an unrestricted PCR product and a product restricted with the enzyme that cuts at site R. Lane 3 shows the result obtained when the template DNA contains an insertion in the amplified region.

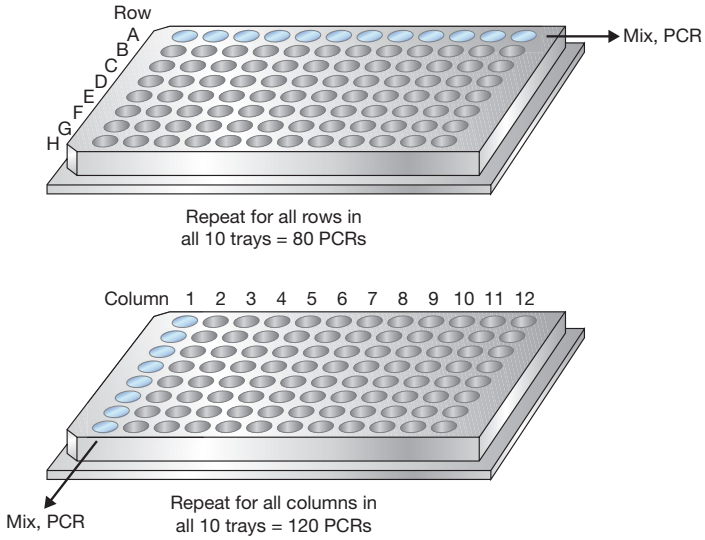


Figure 9.11

Combinatorial screening of clones in microtitre trays. A library of 960 clones, contained in ten microtitre trays, is screened by a series of PCRs, each with a combination of clones. The clone combinations that give positive results enable the well(s) containing positive clone(s) to be identified. For example, if positive PCRs are given with row A of tray 2, row D of tray 6, column 7 of tray 2, and column 9 of tray 6, then it can be deduced that there are positive clones in well A7 of tray 2 and well D9 of tray 6. Although there are 960 clones, unambiguous identification of the positive clones is therefore achieved after just 200 PCRs.

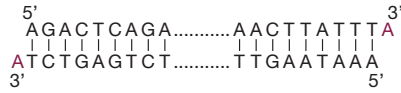


Figure 9.12
Polynucleotides synthesized by *Taq* polymerase often have an extra adenosine at their 3′ ends.

9.3.2 Cloning PCR products

Some applications require that after a PCR the resulting products are ligated into a vector and examined by any of the standard methods used for studying cloned DNA. This may sound easy, but there are complications.

The first problem concerns the ends of the PCR products. From an examination of Figure 9.3, it might be imagined that the short products resulting from PCR amplification are blunt-ended. If this were the case they could be inserted into a cloning vector by blunt-end ligation, or alternatively the PCR products could be provided with sticky ends by the attachment of linkers or adaptors (p. 68). Unfortunately, the situation is not so straightforward. *Taq* polymerase tends to add an additional nucleotide, usually an adenosine, to the end of each strand that it synthesizes. This means that a double-stranded PCR product is not blunt-ended, and instead most 3′ termini have a single nucleotide overhang (Figure 9.12). The overhangs could be removed by treatment with an exonuclease enzyme, resulting in PCR products with true blunt ends, but this is not a popular approach as it is difficult to prevent the exonuclease from becoming overactive and causing further damage to the ends of the molecules.

One solution is to use a special cloning vector which carries thymidine (T) overhangs and which can therefore be ligated to a PCR product (Figure 9.13). These vectors are usually prepared by restricting a standard vector at a blunt-end site, and then treating with *Taq* polymerase in the presence of just 2′-deoxythymidine 5′-triphosphate (dTTP). As no primer is present, all the polymerase can do is add a T nucleotide to the 3′ ends of the blunt-ended vector molecule, resulting in the T-tailed vector into which the PCR products can be inserted. Special vectors of this type have also been designed for use with the topoisomerase ligation method (p. 71), and this is currently the most popular way of cloning PCR products.

A second solution is to design primers that contain restriction sites. After PCR, the products are treated with the restriction endonuclease, which cuts each molecule within the primer sequence, leaving sticky-ended fragments that can be ligated efficiently into

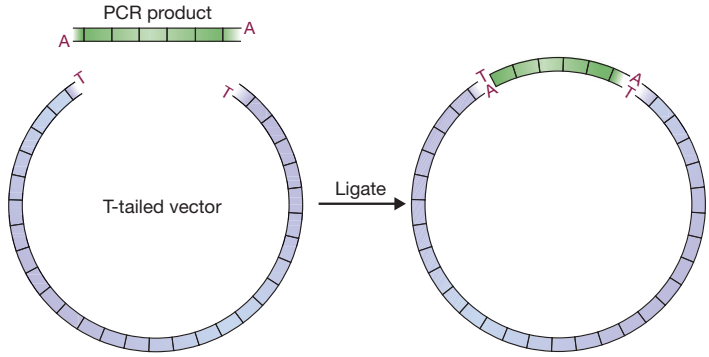


Figure 9.13
Using a special T-tailed vector to clone a PCR product.

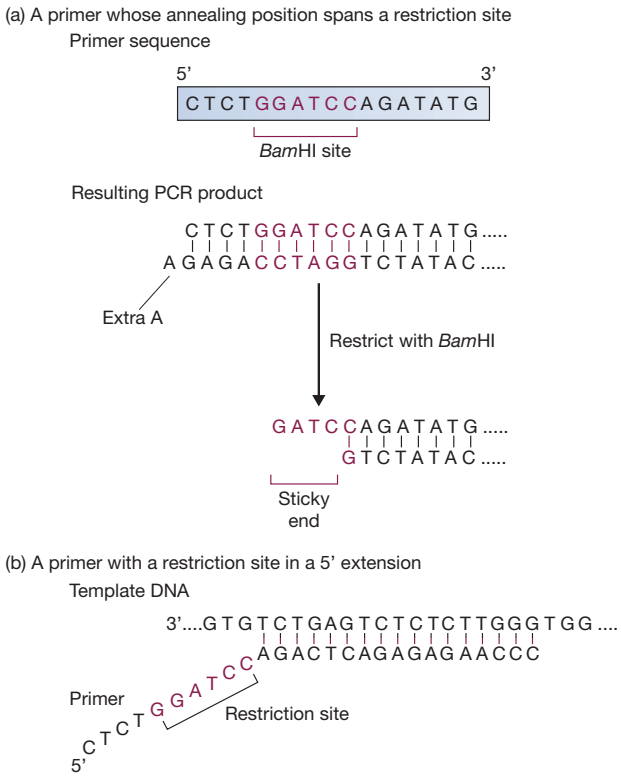


Figure 9.14

Obtaining a PCR product with a sticky end through use of a primer whose sequence includes a restriction site.

a standard cloning vector (Figure 9.14a). The approach is not limited to those instances where the primers span restriction sites that are present in the template DNA. Instead, the restriction site can be included within a short extension at the 5′ end of each primer (Figure 9.14b). These extensions cannot hybridize to the template molecule, but they are copied during the PCR, resulting in PCR products that carry terminal restriction sites.

9.3.3 Problems with the error rate of Taq polymerase

All DNA polymerases make mistakes during DNA synthesis, occasionally inserting an incorrect nucleotide into the growing DNA strand. Most polymerases are able to rectify these errors by reversing over the mistake and resynthesizing the correct sequence. This property, which is referred to as the **proofreading** function, depends on the polymerase possessing a 3′ to 5′ exonuclease activity (p. 178).

Taq polymerase lacks a proofreading activity and as a result is unable to correct its errors. This means that the DNA synthesized by *Taq* polymerase is not always an accurate copy of the template molecule. The error rate has been estimated at one mistake for every 9000 nucleotides of DNA that is synthesized, which might appear to be almost insignificant but translates to one error in every 300 bp for the PCR products obtained after 30 cycles. This is because PCR involves copies being made of copies of copies, so that the polymerase-induced errors gradually accumulate, and the fragments produced

at the end of a PCR will contain copies of earlier errors together with any new errors introduced during the final round of synthesis.

For many applications, this high error rate does not present a problem. In particular, sequencing of a PCR product provides the correct sequence of the template, even though the PCR products contain the errors introduced by *Taq* polymerase. This is because the errors are distributed randomly, so for every molecule that has an error at a particular nucleotide position, there will be many molecules with the correct sequence. The resulting sequence will be incorrect only if an error occurred during a very early cycle in the PCR, as repeated copying will result in these errors being present in many of the molecules in the final product.

Unfortunately, the errors become much more of a problem if the PCR products are cloned. As each resulting clone contains multiple copies of a single amplified fragment, the cloned DNA will not necessarily have the same sequence as the original template molecule used in the PCR (Figure 9.15). This possibility lends an uncertainty to all experiments carried out with cloned PCR products and dictates that, whenever possible, the amplified DNA should be studied directly rather than being cloned.

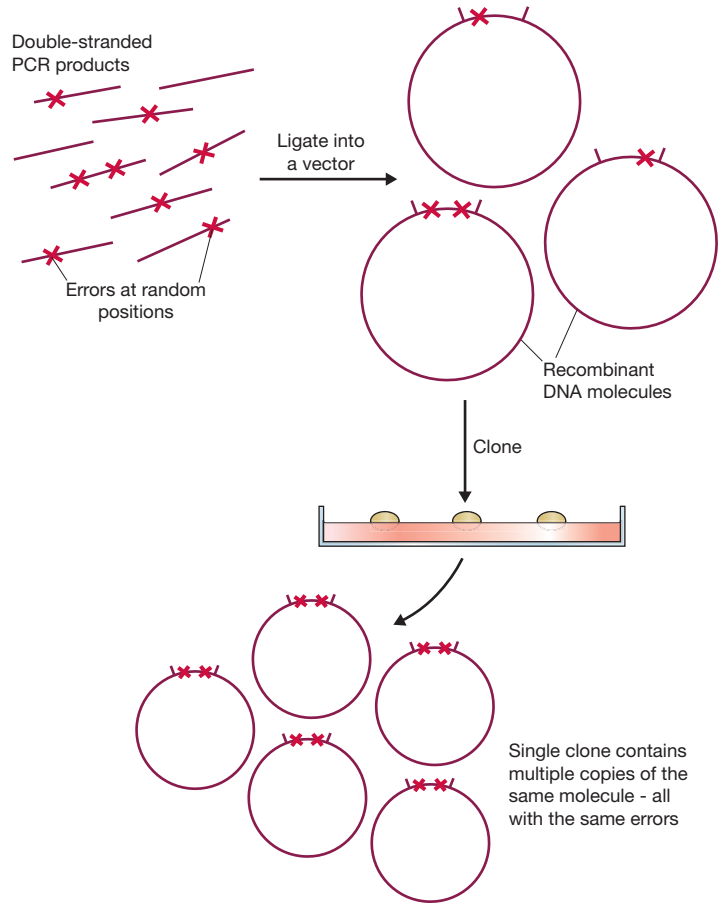


Figure 9.15

The high error rate of *Taq* polymerase becomes a factor when PCR products are cloned.

Table 9.1
Number of short products synthesized after 25 cycles of PCR with different numbers of starting molecule.

NUMBER OF STARTING MOLECULES	NUMBER OF SHORT PRODUCTS
1	4 194 304
2	8 388 608
5	20 971 520
10	41 943 040
25	104 857 600
50	209 715 200
100	419 430 400

The numbers assume that amplification is 100% efficient, and none of the reactants becoming limiting during the course of the PCR.

9.4 Real-time PCR enables the amount of starting material to be quantified

The amount of product synthesized during a set number of cycles of a PCR depends on the number of DNA molecules present in the starting mixture (Table 9.1). If only a few DNA molecules are present at the beginning of the PCR, then relatively little product will be made, but if there are many starting molecules then the product yield will be higher. This relationship enables PCR to be used to quantify the number of DNA molecules present in an extract.

9.4.1 Carrying out a quantitative PCR experiment

In **quantitative PCR (qPCR)** the amount of product synthesized during a test PCR is compared with the amounts synthesized during PCRs with known quantities of starting DNA. When PCR was first developed, agarose gel electrophoresis was used to make these comparisons. After staining the gel, the band intensities were examined to identify the control PCR whose product was most similar to that of the test (Figure 9.16). Although easy to perform, this type of qPCR is imprecise because large differences in the amount of starting DNA give relatively small differences in the band intensities of the resulting PCR products.

Today, quantification is carried out by **real-time PCR**, which is a modification of the standard PCR technique whereby synthesis of the product is measured over time, as

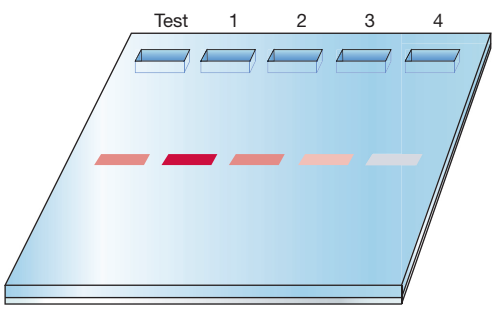
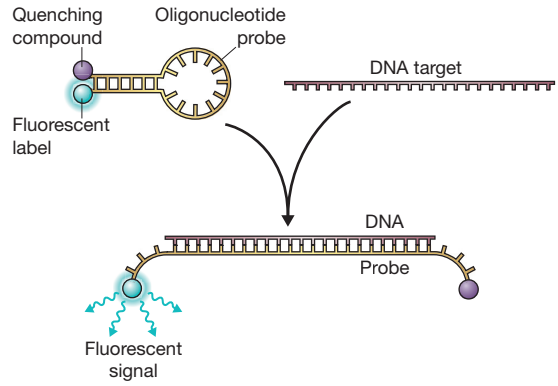


Figure 9.16
Using agarose gel electrophoresis to quantify the amount of DNA in a test PCR. Lanes 1 to 4 are control PCRs carried out with decreasing amounts of template DNA. The intensity of staining for the test band suggests that this PCR contained approximately the same amount of DNA as the control run in lane 2.

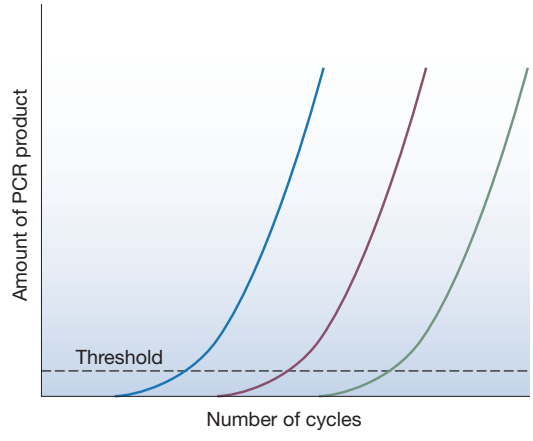
Figure 9.17
Hybridization of a reporter probe to its target DNA.



the PCR proceeds through its series of cycles. There are two ways of following product synthesis in real time:

- A dye that emits a fluorescent signal when it binds to double-stranded DNA can be included in the PCR mixture. This method is used to measure the total amount of double-stranded DNA in the PCR at any particular time, which may overestimate the actual amount of the product because the primers sometimes anneal to one another in various non-specific ways, increasing the amount of double-stranded DNA that is present.
- A short oligonucleotide called a **reporter probe**, which emits a fluorescent signal when it hybridizes to the PCR product, can be used. Because the probe only hybridizes to the PCR product, this method is less prone to inaccuracies caused by primer–primer annealing. Each probe molecule has a pair of labels. A fluorescent dye is attached to one end of the oligonucleotide, and a quenching compound, which inhibits the fluorescent signal, is attached to the other end (Figure 9.17). Normally, there is no fluorescence because the oligonucleotide is designed in such a way that its two ends base pair to one another, placing the quencher next to the dye. Hybridization between the oligonucleotide and the PCR product disrupts this base pairing, moving the quencher away from the dye and enabling the fluorescent signal to be generated.

Figure 9.18
Quantification by real-time PCR. The graph shows product synthesis during three PCRs, each with a different amount of starting DNA. During a PCR, product accumulates exponentially, the amount present at any particular cycle being proportional to the amount of starting DNA. The blue curve is therefore the PCR with the greatest amount of starting DNA, and the green curve is the one with the least starting DNA. If the amounts of starting DNA in these three PCRs are known, then the amount in a test PCR can be quantified by comparison with these controls. In practice, the comparison is made by identifying the cycle at which product synthesis moves above a threshold amount, indicated by the horizontal line on the graph.



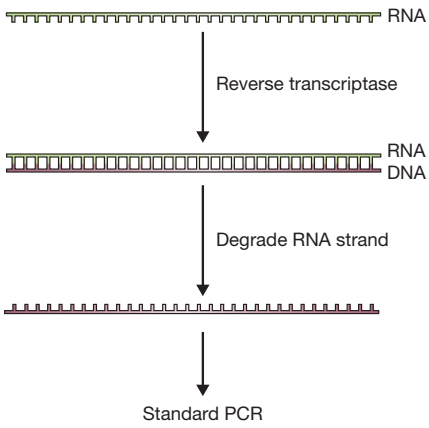


Figure 9.19
Reverse transcriptase PCR.

Both systems enable synthesis of the PCR product to be followed by measuring the fluorescent signal. Quantification again requires comparison between test and control PCRs, usually by identifying the stage in the PCR at which the amount of fluorescent signal reaches a pre-set threshold (Figure 9.18). The more rapidly the threshold is reached, the greater the amount of DNA in the starting mixture.

9.4.2 Real-time PCR can also quantify RNA

Real-time PCR is often used to quantify the amount of DNA in an extract, for example to follow the progression of a viral infection by measuring the amount of pathogen DNA that is present in a tissue. The method can also be used to measure RNA amounts, in particular to determine the extent of expression of a particular gene by quantifying its mRNA. The gene under study might be one that is switched on in cancerous cells, in which case quantifying its mRNA will enable the development of the cancer to be monitored and the effects of subsequent treatment to be assessed.

How do we carry out PCR if RNA is the starting material? The answer is to use **reverse transcriptase PCR**. The first step in this procedure is to convert the RNA molecules into single-stranded complementary DNA (cDNA) (Figure 9.19). Once this preliminary step has been carried out, the PCR primers and *Taq* polymerase are added and the experiment proceeds exactly as in the standard technique. Some thermostable polymerases are able to make DNA copies of both RNA and DNA molecules (i.e., they have both reverse transcriptase and DNA-dependent DNA polymerase activities) and so can carry out all the steps of this type of PCR in a single reaction.

Further reading

FURTHER READING

Higuchi, R., Dollinger, G., Walsh, P.S., and Griffith, R. (1992) Simultaneous amplification and detection of specific DNA sequences. *Biotechnology*, **10**, 413–417. [The first description of real-time PCR].

Marchuk, D., Drumm, M., Saulino, A., and Collins, F.S. (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research*, **19**, 1154.

Rychlik, W., Spencer, W.J., and Rhoads, R.E. (1990) Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Research*, **18**, 6409–6412.

Saiki, R.K., Gelfand, D.H., Stoffel, S., *et al*. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491. [The first description of PCR with *Taq* polymerase.]

Tindall, K.R. and Kunkel, T.A. (1988) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry*, **27**, 6008–6013. [Describes the error rate of *Taq* polymerase.]

VanGuilder, H.D., Vrana, K.E., and Freeman, W.M. (2008) Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, **44**, 619–624.