Chapter 4

Manipulation of Purified DNA

Chapter contents
CHAPTER CONTENTS

4.1 The range of DNA manipulative enzymes
4.2 Enzymes for cutting DNA—restriction endonucleases
4.3 Ligation—joining DNA molecules together

Once pure samples of DNA have been prepared, the next step in a gene cloning experiment is construction of the recombinant DNA molecule (see Figure 1.1). To produce this recombinant molecule, the vector, as well as the DNA to be cloned, must be cut at specific points and then joined together in a controlled manner. Cutting and joining are two examples of DNA manipulative techniques, a wide variety of which have been developed over the past few years. As well as being cut and joined, DNA molecules can be shortened, lengthened, copied into RNA or into new DNA molecules, and modified by the addition or removal of specific chemical groups. These manipulations, all of which can be carried out in the test tube, provide the foundation not only for gene cloning, but also for studies of DNA biochemistry, gene structure, and the control of gene expression.

Almost all DNA manipulative techniques make use of purified enzymes. Within the cell these enzymes participate in essential processes such as DNA replication and transcription, breakdown of unwanted or foreign DNA (e.g., invading virus DNA), repair of mutated DNA, and **recombination** between different DNA molecules. After purification from cell extracts, many of these enzymes can be persuaded to carry out their natural reactions, or something closely related to them, under artificial conditions. Although these enzymatic reactions are often straightforward, most are absolutely impossible to perform by standard chemical methods. Purified enzymes are therefore crucial to genetic engineering and an important industry has sprung up around their preparation, characterization, and marketing. Commercial suppliers of high purity enzymes provide an essential service to the molecular biologist.

Gene Cloning and DNA Analysis: An Introduction. 6th edition. By T.A. Brown. Published 2010 by Blackwell Publishing.

The cutting and joining manipulations that underlie gene cloning are carried out by enzymes called restriction endonucleases (for cutting) and ligases (for joining). Most of this chapter will be concerned with the ways in which these two types of enzyme are used. First, however, we must consider the whole range of DNA manipulative enzymes, to see exactly what types of reaction can be performed. Many of these enzymes will be mentioned in later chapters when procedures that make use of them are described.

4.1 The range of DNA manipulative enzymes

DNA manipulative enzymes can be grouped into four broad classes, depending on the type of reaction that they catalyze:

- Nucleases are enzymes that cut, shorten, or degrade nucleic acid molecules.
- Ligases join nucleic acid molecules together.
- Polymerases make copies of molecules.
- Modifying enzymes remove or add chemical groups.

Before considering in detail each of these classes of enzyme, two points should be made. The first is that, although most enzymes can be assigned to a particular class, a few display multiple activities that span two or more classes. Most importantly, many polymerases combine their ability to make new DNA molecules with an associated DNA degradative (i.e., nuclease) activity.

Second, it should be appreciated that, as well as the DNA manipulative enzymes, many similar enzymes able to act on RNA are known. The ribonuclease used to remove contaminating RNA from DNA preparations (p. 30) is an example of such an enzyme. Although some RNA manipulative enzymes have applications in gene cloning and will be mentioned in later chapters, we will in general restrict our thoughts to those enzymes that act on DNA.

4.1.1 Nucleases

Nucleases degrade DNA molecules by breaking the phosphodiester bonds that link one nucleotide to the next in a DNA strand. There are two different kinds of nuclease (Figure 4.1):

- Exonucleases remove nucleotides one at a time from the end of a DNA molecule.
- Endonucleases are able to break internal phosphodiester bonds within a DNA molecule.

The main distinction between different exonucleases lies in the number of strands that are degraded when a double-stranded molecule is attacked. The enzyme called Bal31 (purified from the bacterium *Alteromonas espejiana*) is an example of an exonuclease that removes nucleotides from both strands of a double-stranded molecule (Figure 4.2a). The greater the length of time that Bal31 is allowed to act on a group of DNA molecules, the shorter the resulting DNA fragments will be. In contrast, enzymes such as *E. coli* exonuclease III degrade just one strand of a double-stranded molecule, leaving single-stranded DNA as the product (Figure 4.2b).

The same criterion can be used to classify endonucleases. S1 endonuclease (from the fungus *Aspergillus oryzae*) only cleaves single strands (Figure 4.3a), whereas deoxyribonuclease I (DNase I), which is prepared from cow pancreas, cuts both single- and double-stranded molecules (Figure 4.3b). DNase I is non-specific in that it attacks DNA at any internal phosphodiester bond, so the end result of prolonged DNase I

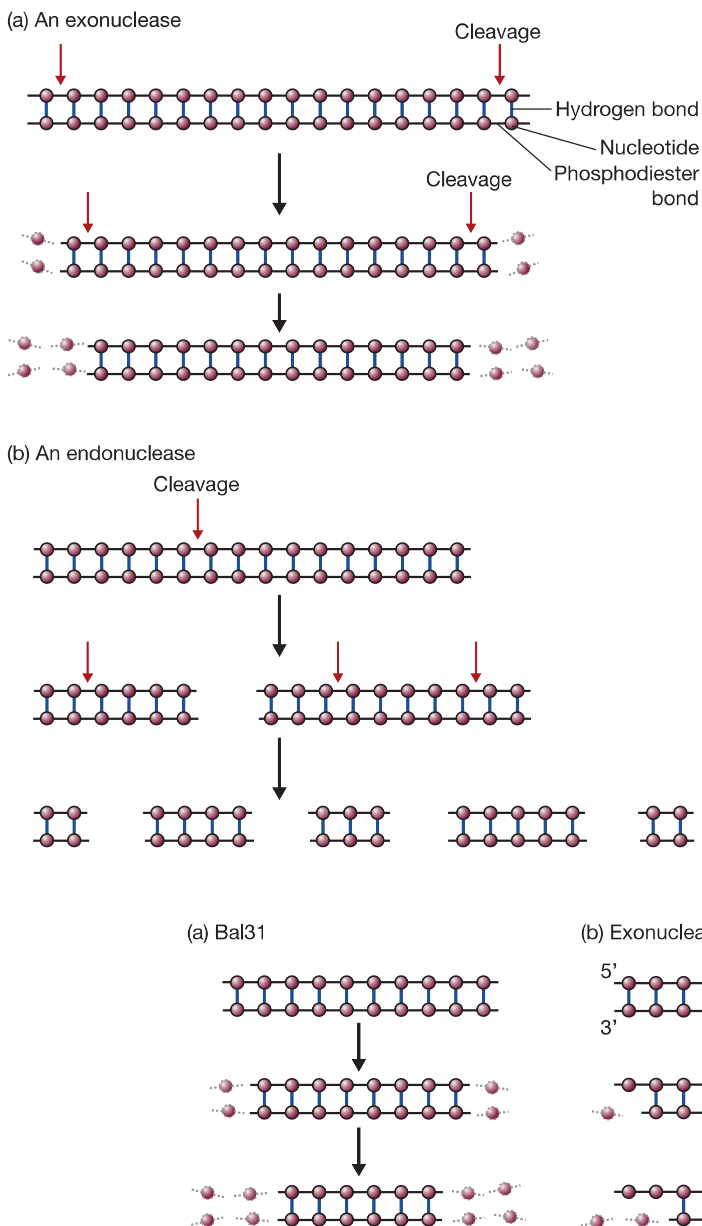


Figure 4.1

The reactions catalyzed by the two different kinds of nuclease. (a) An exonuclease, which removes nucleotides from the end of a DNA molecule. (b) An endonuclease, which breaks internal phosphodiester bonds.

Figure 4.2

The reactions catalyzed by different types of exonuclease. (a) Bal31, which removes nucleotides from both strands of a double-stranded molecule. (b) Exonuclease III, which removes nucleotides only from the 3′ terminus.

action is a mixture of mononucleotides and very short oligonucleotides. On the other hand, the special group of enzymes called restriction endonucleases cleave double-stranded DNA only at a limited number of specific recognition sites (Figure 4.3c). These important enzymes are described in detail on p. 50.

4.1.2 Ligases

In the cell the function of DNA ligase is to repair single-stranded breaks ("discontinuities") that arise in double-stranded DNA molecules during, for example, DNA replication. DNA ligases from most organisms can also join together two individual fragments of

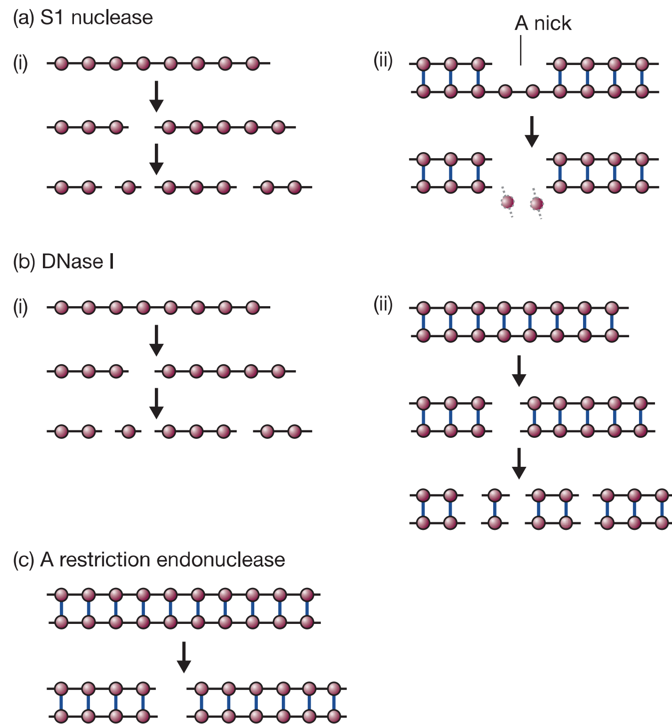


Figure 4.3

The reactions catalyzed by different types of endonuclease. (a) S1 nuclease, which cleaves only single-stranded DNA, including single-stranded nicks in mainly double-stranded molecules. (b) DNase I, which cleaves both single- and double-stranded DNA. (c) A restriction endonuclease, which cleaves double-stranded DNA, but only at a limited number of sites.

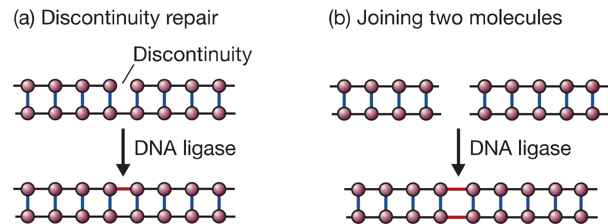


Figure 4.4

The two reactions catalyzed by DNA ligase. (a) Repair of a discontinuity—a missing phosphodiester bond in one strand of a double-stranded molecule. (b) Joining two molecules together.

double-stranded DNA (Figure 4.4). The role of these enzymes in construction of recombinant DNA molecules is described on p. 63.

4.1.3 Polymerases

DNA polymerases are enzymes that synthesize a new strand of DNA complementary to an existing DNA or RNA template (Figure 4.5a). Most polymerases can function only if the **template** possesses a double-stranded region that acts as a **primer** for initiation of polymerization.

Four types of DNA polymerase are used routinely in genetic engineering. The first is DNA polymerase I, which is usually prepared from *E. coli*. This enzyme attaches to a short single-stranded region (or **nick**) in a mainly double-stranded DNA molecule, and then synthesizes a completely new strand, degrading the existing strand as it proceeds (Figure 4.5b). DNA polymerase I is therefore an example of an enzyme with a dual activity—DNA polymerization and DNA degradation.

The polymerase and nuclease activities of DNA polymerase I are controlled by different parts of the enzyme molecule. The nuclease activity is contained in the first 323 amino acids of the polypeptide, so removal of this segment leaves a modified enzyme that retains the polymerase function but is unable to degrade DNA. This modified

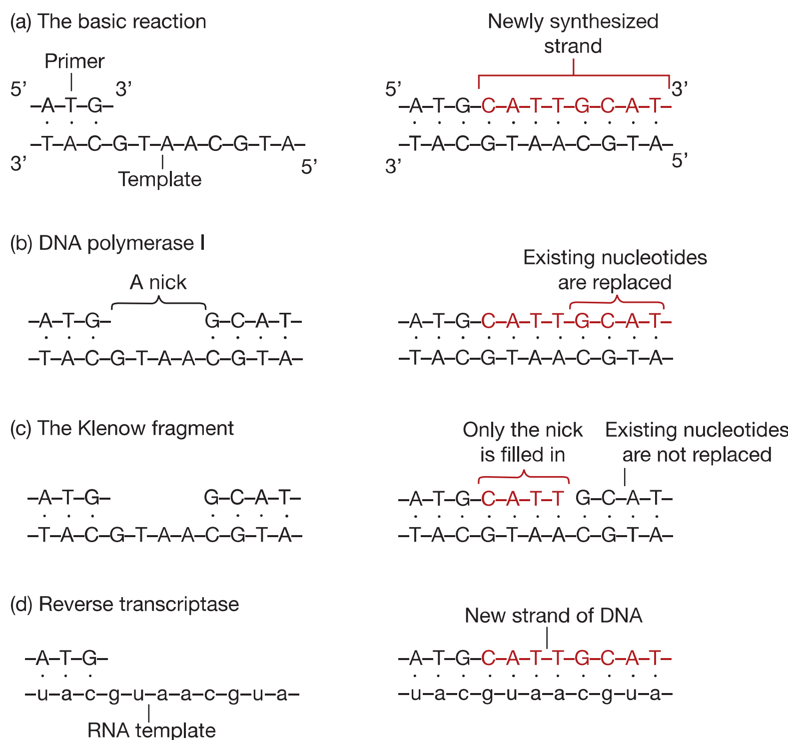


Figure 4.5

The reactions catalyzed by DNA polymerases. (a) The basic reaction: a new DNA strand is synthesized in the 5′ to 3′ direction. (b) DNA polymerase I, which initially fills in nicks but then continues to synthesize a new strand, degrading the existing one as it proceeds. (c) The Klenow fragment, which only fills in nicks. (d) Reverse transcriptase, which uses a template of RNA.

enzyme, called the **Klenow fragment**, can still synthesize a complementary DNA strand on a single-stranded template, but as it has no nuclease activity it cannot continue the synthesis once the nick is filled in (Figure 4.5c). Several other enzymes—natural polymerases and modified versions—have similar properties to the Klenow fragment. The major application of these polymerases is in DNA sequencing (p. 166).

The *Taq* DNA polymerase used in the polymerase chain reaction (PCR) (see Figure 1.2) is the DNA polymerase I enzyme of the bacterium *Thermus aquaticus*. This organism lives in hot springs, and many of its enzymes, including the *Taq* DNA polymerase, are thermostable, meaning that they are resistant to denaturation by heat treatment. This is the special feature of *Taq* DNA polymerase that makes it suitable for PCR, because if it was not thermostable it would be inactivated when the temperature of the reaction is raised to 94°C to denature the DNA.

The final type of DNA polymerase that is important in genetic engineering is **reverse transcriptase**, an enzyme involved in the replication of several kinds of virus. Reverse transcriptase is unique in that it uses as a template not DNA but RNA (Figure 4.5d). The ability of this enzyme to synthesize a DNA strand complementary to an RNA template is central to the technique called complementary DNA (cDNA) cloning (p. 133).

4.1.4 DNA modifying enzymes

There are numerous enzymes that modify DNA molecules by addition or removal of specific chemical groups. The most important are as follows:

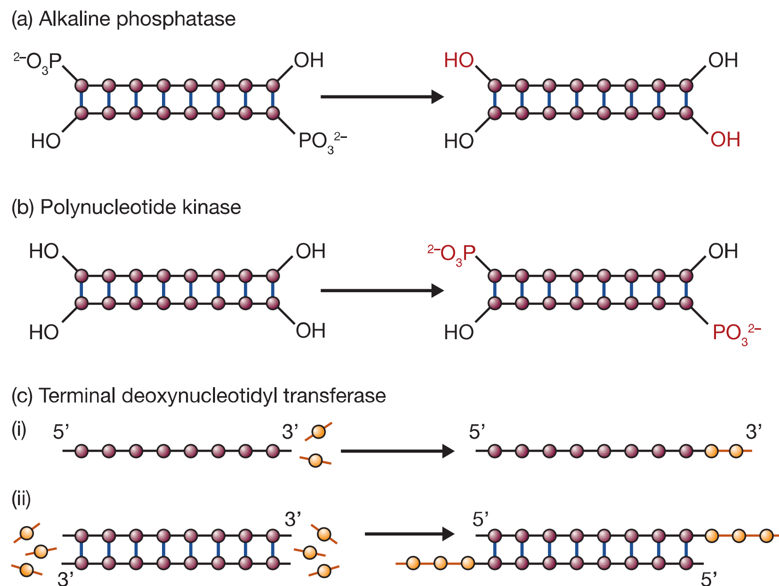


Figure 4.6

The reactions catalyzed by DNA modifying enzymes. (a) Alkaline phosphatase, which removes 5′-phosphate groups. (b) Polynucleotide kinase, which attaches 5′-phosphate groups. (c) Terminal deoxynucleotidyl transferase, which attaches deoxyribonucleotides to the 3′ termini of polynucleotides in either (i) single-stranded or (ii) double-stranded molecules.

- **Alkaline phosphatase** (from *E. coli*, calf intestinal tissue, or arctic shrimp), which removes the phosphate group present at the **5′ terminus** of a DNA molecule (Figure 4.6a).
- **Polynucleotide kinase** (from *E. coli* infected with T4 phage), which has the reverse effect to alkaline phosphatase, adding phosphate groups onto free 5′ termini (Figure 4.6b).
- **Terminal deoxynucleotidyl transferase** (from calf thymus tissue), which adds one or more deoxyribonucleotides onto the **3′ terminus** of a DNA molecule (Figure 4.6c).

4.2 Enzymes for cutting DNA—restriction endonucleases

Gene cloning requires that DNA molecules be cut in a very precise and reproducible fashion. This is illustrated by the way in which the vector is cut during construction of a recombinant DNA molecule (Figure 4.7a). Each vector molecule must be cleaved at a single position, to open up the circle so that new DNA can be inserted: a molecule that is cut more than once will be broken into two or more separate fragments and will be of no use as a cloning vector. Furthermore, each vector molecule must be cut at exactly the same position on the circle—as will become apparent in later chapters, random cleavage is not satisfactory. It should be clear that a very special type of nuclease is needed to carry out this manipulation.

Often it is also necessary to cleave the DNA that is to be cloned (Figure 4.7b). There are two reasons for this. First, if the aim is to clone a single gene, which may consist of only 2 or 3 kb of DNA, then that gene will have to be cut out of the large (often greater than 80 kb) DNA molecules produced by skilfull use of the preparative techniques

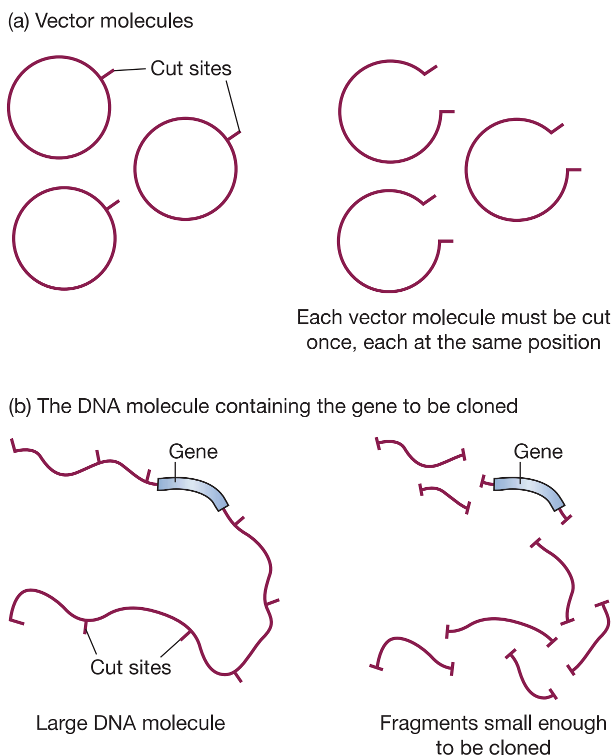


Figure 4.7

The need for very precise cutting manipulations in a gene cloning experiment.

described in Chapter 3. Second, large DNA molecules may have to be broken down simply to produce fragments small enough to be carried by the vector. Most cloning vectors exhibit a preference for DNA fragments that fall into a particular size range: most plasmid-based vectors, for example, are very inefficient at cloning DNA molecules more than 8 kb in length.

Purified restriction endonucleases allow the molecular biologist to cut DNA molecules in the precise, reproducible manner required for gene cloning. The discovery of these enzymes, which led to Nobel Prizes for W. Arber, H. Smith, and D. Nathans in 1978, was one of the key breakthroughs in the development of genetic engineering.

4.2.1 The discovery and function of restriction endonucleases

The initial observation that led to the eventual discovery of restriction endonucleases was made in the early 1950s, when it was shown that some strains of bacteria are immune to bacteriophage infection, a phenomenon referred to as **host-controlled restriction**.

The mechanism of restriction is not very complicated, even though it took over 20 years to be fully understood. Restriction occurs because the bacterium produces an enzyme that degrades the phage DNA before it has time to replicate and direct synthesis of new phage particles (Figure 4.8a). The bacterium's own DNA, the destruction of which would of course be lethal, is protected from attack because it carries additional methyl groups that block the degradative enzyme action (Figure 4.8b).

These degradative enzymes are called restriction endonucleases and are synthesized by many, perhaps all, species of bacteria: over 2500 different ones have been isolated and more than 300 are available for use in the laboratory. Three different classes of restriction endonuclease are recognized, each distinguished by a slightly different mode of action. Types I and III are rather complex and have only a limited role in genetic

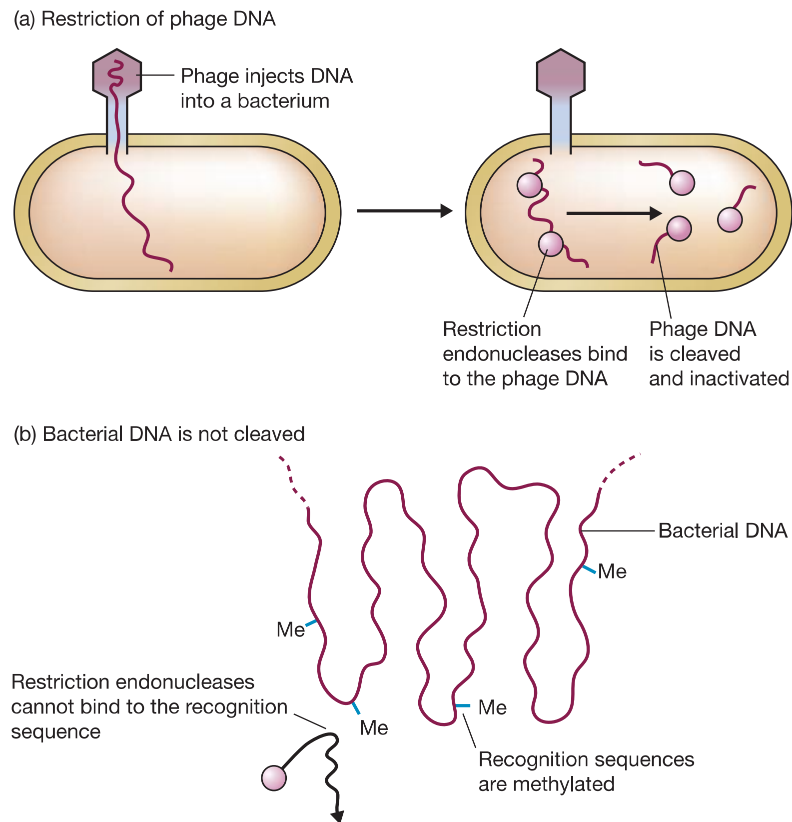


Figure 4.8
The function of a restriction endonuclease in a bacterial cell: (a) phage DNA is cleaved, but (b) bacterial DNA is not.

engineering. Type II restriction endonucleases, on the other hand, are the cutting enzymes that are so important in gene cloning.

4.2.2 Type II restriction endonucleases cut DNA at specific nucleotide sequences

The central feature of type II restriction endonucleases (which will be referred to simply as "restriction endonucleases" from now on) is that each enzyme has a specific recognition sequence at which it cuts a DNA molecule. A particular enzyme cleaves DNA at the recognition sequence and nowhere else. For example, the restriction endonuclease called *Pvu*I (isolated from *Proteus vulgaris*) cuts DNA only at the hexanucleotide CGATCG. In contrast, a second enzyme from the same bacterium, called *Pvu*II, cuts at a different hexanucleotide, in this case CAGCTG.

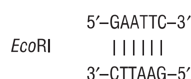
Many restriction endonucleases recognize hexanucleotide target sites, but others cut at four, five, eight, or even longer nucleotide sequences. *Sau*3A (from *Staphylococcus aureus* strain 3A) recognizes GATC, and *Alu*I (*Arthrobacter luteus*) cuts at AGCT. There are also examples of restriction endonucleases with degenerate recognition sequences, meaning that they cut DNA at any one of a family of related sites. *Hin*fI (*Haemophilus influenzae* strain R_f), for instance, recognizes GANTC, so cuts at GAATC, GATTC, GAGTC, and GACTC. The recognition sequences for some of the most frequently used restriction endonucleases are listed in Table 4.1.

Table 4.1

The recognition sequences for some of the most frequently used restriction endonucleases.

ENZYME	ORGANISM	RECOGNITION SEQUENCE*	BLUNT OR STICKY END
*Eco*RI	*Escherichia coli*	GAATTC	Sticky
*Bam*HI	*Bacillus amyloliquefaciens*	GGATCC	Sticky
*Bgl*II	*Bacillus globigii*	AGATCT	Sticky
*Pvu*I	*Proteus vulgaris*	CGATCG	Sticky
*Pvu*II	*Proteus vulgaris*	CAGCTG	Blunt
*Hind*III	*Haemophilus influenzae* R_d	AAGCTT	Sticky
*Hin*fI	*Haemophilus influenzae* R_f	GANTC	Sticky
*Sau*3A	*Staphylococcus aureus*	GATC	Sticky
*Alu*I	*Arthrobacter luteus*	AGCT	Blunt
*Taq*I	*Thermus aquaticus*	TCGA	Sticky
*Hae*III	*Haemophilus aegyptius*	GGCC	Blunt
*Not*I	*Nocardia otitidis-caviarum*	GCGGCCGC	Sticky
*Sfi*I	*Streptomyces fimbriatus*	GGCCNNNNNGGCC	Sticky

*The sequence shown is that of one strand, given in the 5′ to 3′ direction. "N" indicates any nucleotide. Note that almost all recognition sequences are palindromes: when both strands are considered they read the same in each direction, for example:



4.2.3 Blunt ends and sticky ends

The exact nature of the cut produced by a restriction endonuclease is of considerable importance in the design of a gene cloning experiment. Many restriction endonucleases make a simple double-stranded cut in the middle of the recognition sequence (Figure 4.9a), resulting in a **blunt end** or **flush end**. *Pvu*II and *Alu*I are examples of blunt end cutters.

Other restriction endonucleases cut DNA in a slightly different way. With these enzymes the two DNA strands are not cut at exactly the same position. Instead the cleavage is staggered, usually by two or four nucleotides, so that the resulting DNA fragments have short single-stranded overhangs at each end (Figure 4.9b). These are called sticky or cohesive ends, as base pairing between them can stick the DNA molecule back together again (recall that sticky ends were encountered on p. 20 during the description of λ phage replication). One important feature of sticky end enzymes is that restriction endonucleases with different recognition sequences may produce the same sticky ends. *Bam*HI (recognition sequence GGATCC) and *Bgl*II (AGATCT) are examples—both produce GATC sticky ends (Figure 4.9c). The same sticky end is also produced by *Sau*3A, which recognizes only the tetranucleotide GATC. Fragments of DNA produced by cleavage with either of these enzymes can be joined to each other, as each fragment carries a complementary sticky end.

4.2.4 The frequency of recognition sequences in a DNA molecule

The number of recognition sequences for a particular restriction endonuclease in a DNA molecule of known length can be calculated mathematically. A tetranucleotide sequence (e.g., GATC) should occur once every $4^4 = 256$ nucleotides, and a hexanucleotide

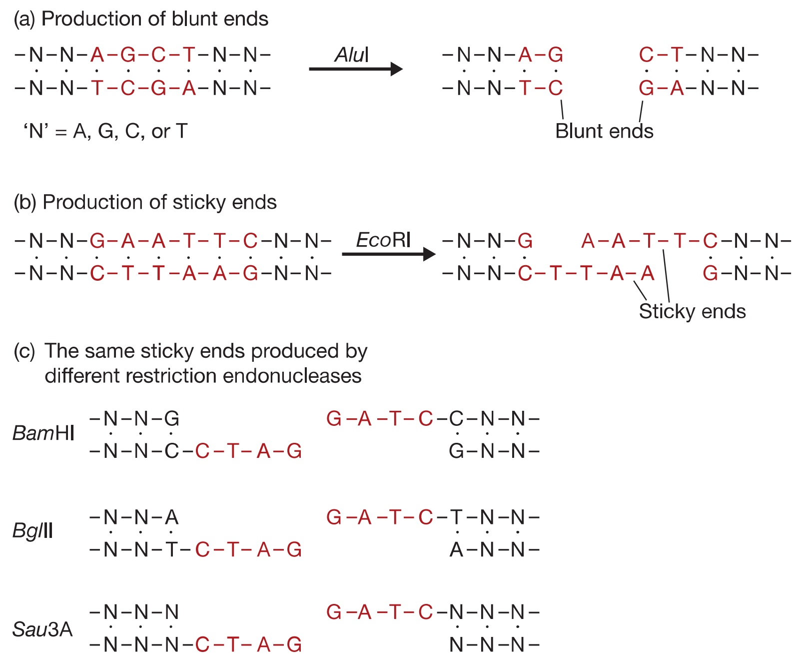


Figure 4.9

The ends produced by cleavage of DNA with different restriction endonucleases. (a) A blunt end produced by *Alu*I. (b) A sticky end produced by *Eco*RI. (c) The same sticky ends produced by *Bam*HI, *Bgl*II and *Sau*3A.

(e.g., GGATCC) once every $4^6 = 4096$ nucleotides. These calculations assume that the nucleotides are ordered in a random fashion and that the four different nucleotides are present in equal proportions (i.e., the GC content = 50%). In practice, neither of these assumptions is entirely valid. For example, the λ DNA molecule, at 49 kb, should contain about 12 sites for a restriction endonuclease with a hexanucleotide recognition sequence. In fact, many of these recognition sites occur less frequently (e.g., six for *Bgl*II, five for *Bam*HI, and only two for *Sal*I), a reflection of the fact that the GC content for λ is rather less than 50% (Figure 4.10a).

Furthermore, restriction sites are generally not evenly spaced out along a DNA molecule. If they were, then digestion with a particular restriction endonuclease would give fragments of roughly equal sizes. Figure 4.10b shows the fragments produced by cutting λ DNA with *Bgl*II, *Bam*HI, and *Sal*I. In each case there is a considerable spread of fragment sizes, indicating that in λ DNA the nucleotides are not randomly ordered.

The lesson to be learned from Figure 4.10 is that although mathematics may give an idea of how many restriction sites to expect in a given DNA molecule, only experimental analysis can provide the true picture. We must therefore move on to consider how restriction endonucleases are used in the laboratory.

4.2.5 Performing a restriction digest in the laboratory

As an example, we will consider how to digest a sample of λ DNA (concentration 125 mg/ml) with *Bgl*II.

First, the required amount of DNA must be pipetted into a test tube. The amount of DNA that will be restricted depends on the nature of the experiment. In this case we will digest 2 μg of λ DNA, which is contained in 16 μl of the sample (Figure 4.11a). Very accurate micropipettes will therefore be needed.

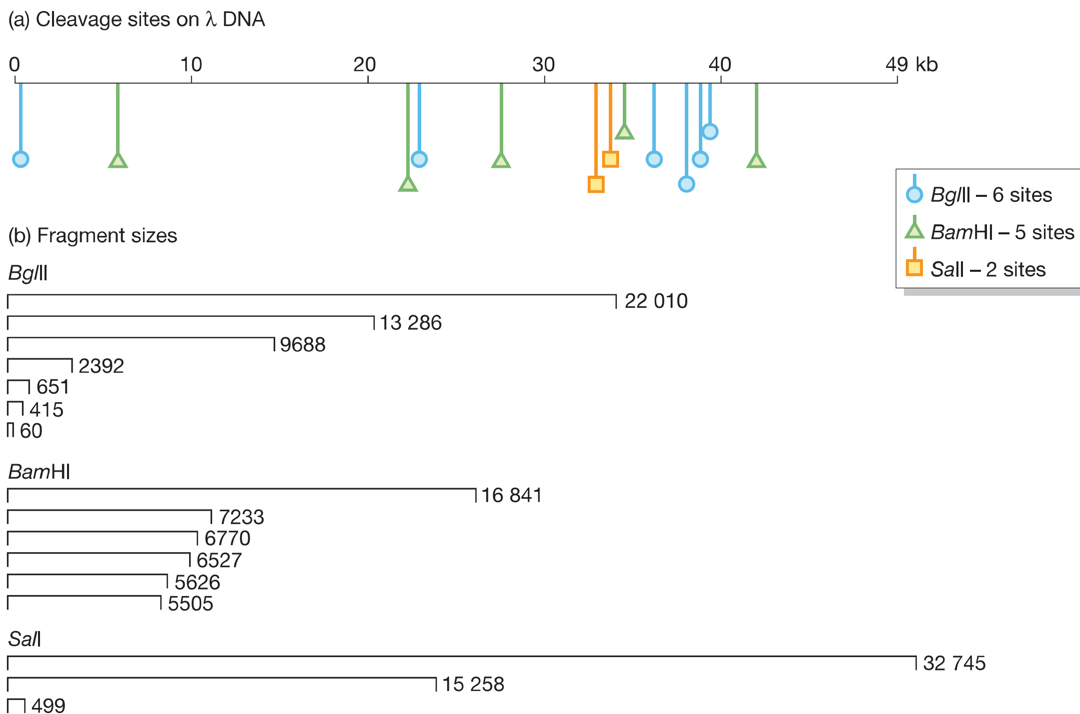


Figure 4.10

Restriction of the λ DNA molecule. (a) The positions of the recognition sequences for *Bgl*II, *Bam*HI, and *Sal*I. (b) The fragments produced by cleavage with each of these restriction endonucleases. The numbers are the fragment sizes in base pairs.

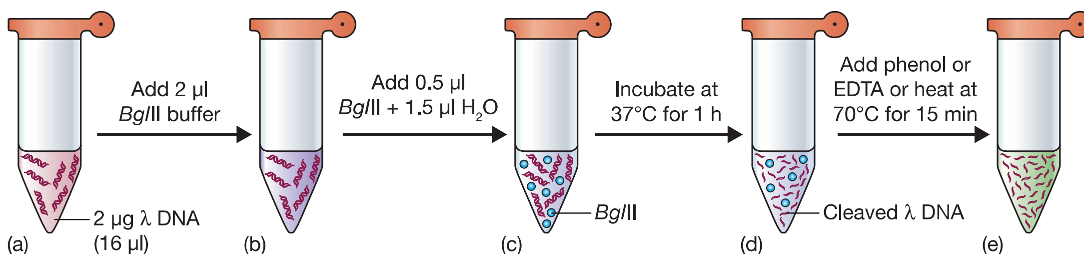


Figure 4.11

Performing a restriction digest in the laboratory.

The other main component in the reaction will be the restriction endonuclease, obtained from a commercial supplier as a pure solution of known concentration. But before adding the enzyme, the solution containing the DNA must be adjusted to provide the correct conditions to ensure maximal activity of the enzyme. Most restriction endonucleases function adequately at pH 7.4, but different enzymes vary in their requirements for ionic strength (usually provided by sodium chloride (NaCl)) and magnesium (Mg^{2+}) concentration (all type II restriction endonucleases require Mg^{2+} in order to function). It is also advisable to add a reducing agent, such as dithiothreitol (DTT), which stabilizes the enzyme and prevents its inactivation. Providing the right conditions for the enzyme is very important—incorrect NaCl or Mg^{2+} concentrations not only decrease the activity of the restriction endonuclease, they might also cause

Table 4.2

A 10 × buffer suitable for restriction of DNA with *Bgl*II.

COMPONENT	CONCENTRATION (mM)
Tris–HCl, pH 7.4	500
$MgCl_2$	100
NaCl	500
Dithiothreitol	10

changes in the specificity of the enzyme, so that DNA cleavage occurs at additional, non-standard recognition sequences.

The composition of a suitable buffer for *Bgl*II is shown in Table 4.2. This buffer is ten times the working concentration, and is diluted by being added to the reaction mixture. In our example, a suitable final volume for the reaction mixture would be 20 μl, so we add 2 μl of 10 × *Bgl*II buffer to the 16 μl of DNA already present (Figure 4.11b).

The restriction endonuclease can now be added. By convention, 1 unit of enzyme is defined as the quantity needed to cut 1 μg of DNA in 1 hour, so we need 2 units of *Bgl*II to cut 2 μg of λ DNA. *Bgl*II is frequently obtained at a concentration of 4 units/μl, so 0.5 μl is sufficient to cleave the DNA. The final ingredients in the reaction mixture are therefore 0.5 μl *Bgl*II + 1.5 μl water, giving a final volume of 20 μl (Figure 4.11c).

The last factor to consider is incubation temperature. Most restriction endonucleases, including *Bgl*II, work best at 37°C, but a few have different requirements. *Taq*I, for example, is a restriction enzyme from *Thermus aquaticus* and, like *Taq* DNA polymerase, has a high working temperature. Restriction digests with *Taq*I must be incubated at 65°C to obtain maximum enzyme activity.

After 1 hour the restriction should be complete (Figure 4.11d). If the DNA fragments produced by restriction are to be used in cloning experiments, the enzyme must somehow be destroyed so that it does not accidentally digest other DNA molecules that may be added at a later stage. There are several ways of "killing" the enzyme. For many a short incubation at 70°C is sufficient, for others phenol extraction or the addition of ethylenediamine tetraacetate (EDTA), which binds Mg^{2+} ions preventing restriction endonuclease action, is used (Figure 4.11e).

4.2.6 Analyzing the result of restriction endonuclease cleavage

A restriction digest results in a number of DNA fragments, the sizes of which depend on the exact positions of the recognition sequences for the endonuclease in the original molecule (see Figure 4.10). A way of determining the number and sizes of the fragments is needed if restriction endonucleases are to be of use in gene cloning. Whether or not a DNA molecule is cut at all can be determined fairly easily by testing the viscosity of the solution. Larger DNA molecules result in a more viscous solution than smaller ones, so cleavage is associated with a decrease in viscosity. However, working out the number and sizes of the individual cleavage products is more difficult. In fact, for several years this was one of the most tedious aspects of experiments involving DNA. Eventually the problems were solved in the early 1970s when the technique of gel electrophoresis was developed.

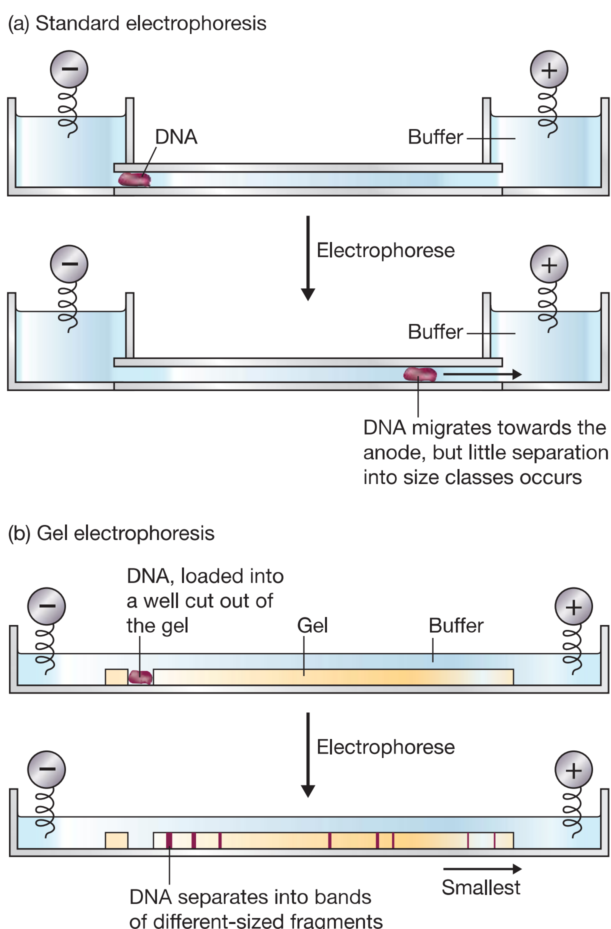


Figure 4.12

(a) Standard electrophoresis does not separate DNA fragments of different sizes, whereas (b) gel electrophoresis does.

Separation of molecules by gel electrophoresis

Electrophoresis, like ion-exchange chromatography (see p. 30), is a technique that uses differences in electrical charge to separate the molecules in a mixture. DNA molecules have negative charges, and so when placed in an electric field they migrate toward the positive pole (Figure 4.12a). The rate of migration of a molecule depends on two factors, its shape and its charge-to-mass ratio. Unfortunately, most DNA molecules are the same shape and all have very similar charge-to-mass ratios. Fragments of different sizes cannot therefore be separated by standard electrophoresis.

The size of the DNA molecule does, however, become a factor if the electrophoresis is performed in a gel. A gel, which is usually made of agarose, polyacrylamide, or a mixture of the two, comprises a complex network of pores, through which the DNA molecules must travel to reach the positive electrode. The smaller the DNA molecule, the faster it can migrate through the gel. **Gel electrophoresis** therefore separates DNA molecules according to their size (Figure 4.12b).

In practice the composition of the gel determines the sizes of the DNA molecules that can be separated. A 0.5 cm thick slab of 0.5% agarose, which has relatively large pores, would be used for molecules in the size range 1–30 kb, allowing, for example, molecules of 10 and 12 kb to be clearly distinguished. At the other end of the scale, a very thin (0.3 mm) 40% polyacrylamide gel, with extremely small pores, would be used to separate much smaller DNA molecules, in the range 1–300 bp, and could distinguish molecules differing in length by just a single nucleotide.

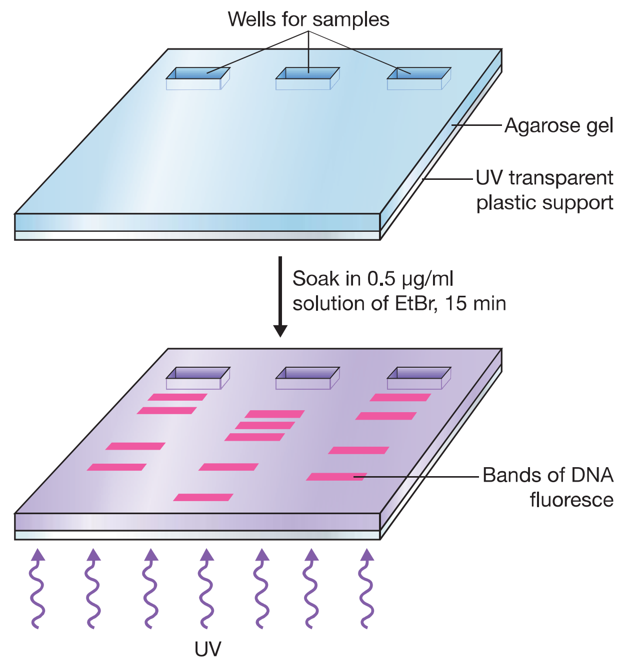


Figure 4.13

Visualizing DNA bands in an agarose gel by EtBr staining and ultraviolet (UV) irradiation.

Visualizing DNA molecules in an agarose gel

The easiest way to see the results of a gel electrophoresis experiment is to stain the gel with a compound that makes the DNA visible. Ethidium bromide (EtBr), already described on p. 37 as a means of visualizing DNA in caesium chloride gradients, is also routinely used to stain DNA in agarose and polyacrylamide gels (Figure 4.13). Bands showing the positions of the different size classes of DNA fragment are clearly visible under ultraviolet irradiation after EtBr staining, so long as sufficient DNA is present. Unfortunately, the procedure is very hazardous because ethidium bromide is a powerful mutagen. EtBr staining also has limited sensitivity, and if a band contains less than about 10 ng of DNA then it might not be visible after staining.

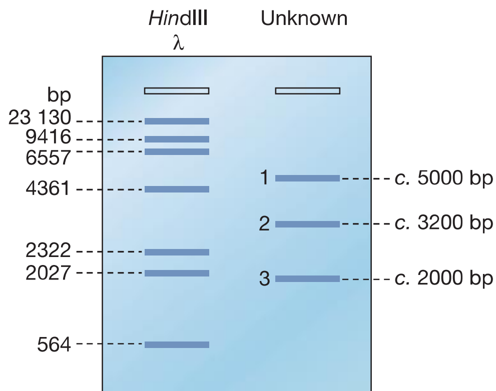
For this reason, non-mutagenic dyes that stain DNA green, red, or blue are now used in many laboratories. Most of these dyes can be used either as a post-stain after electrophoresis, as illustrated in Figure 4.13 for EtBr, or alternatively, because they are non-hazardous, they can be included in the buffer solution in which the agarose or polyacrylamide is dissolved when the gel is prepared. Some of these dyes require ultraviolet irradiation in order to make the bands visible, but others are visualized by illumination at other wavelengths, for example under blue light, removing a second hazard as ultraviolet radiation can cause severe burns. The most sensitive dyes are able to detect bands that contain less than 1 ng DNA.

4.2.7 Estimation of the sizes of DNA molecules

Gel electrophoresis separates different sized DNA molecules, with the smallest molecules traveling the greatest distance toward the positive electrode. If several DNA fragments of varying sizes are present (the result of a successful restriction digest, for example), then a series of bands appears in the gel. How can the sizes of these fragments be determined?

The most accurate method is to make use of the mathematical relationship that links migration rate to molecular mass. The relevant formula is:

$$D = a - b(\log M)$$



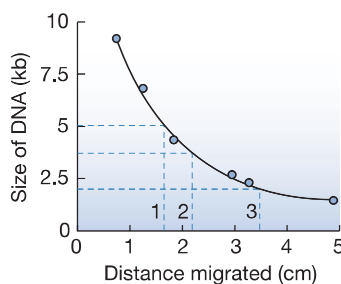


Figure 4.14

Estimation of the sizes of DNA fragments in an agarose gel. (a) A rough estimate of fragment size can be obtained by eye. (b) A more accurate measurement of fragment size is gained by using the mobilities of the *Hin*dIII–λ fragments to construct a calibration curve; the sizes of the unknown fragments can then be determined from the distances they have migrated.

where D is the distance moved, M is the molecular mass, and a and b are constants that depend on the electrophoresis conditions.

Because extreme accuracy in estimating DNA fragment sizes is not always necessary, a much simpler though less precise method is more generally used. A standard restriction digest, comprising fragments of known size, is usually included in each electrophoresis gel that is run. Restriction digests of λ DNA are often used in this way as size markers. For example, *Hin*dIII cleaves λ DNA into eight fragments, ranging in size from 125 bp for the smallest to over 23 kb for the largest. As the sizes of the fragments in this digest are known, the fragment sizes in the experimental digest can be estimated by comparing the positions of the bands in the two tracks (Figure 4.14). Special mixtures of DNA fragments called DNA ladders, whose sizes are multiples of 100 bp or of 1 kb, can also be used as size markers. Although not precise, size estimation by comparison with DNA markers can be performed with as little as a 5% error, which is satisfactory for most purposes.

4.2.8 Mapping the positions of different restriction sites in a DNA molecule

So far we have considered how to determine the number and sizes of the DNA fragments produced by restriction endonuclease cleavage. The next step in restriction analysis is to construct a map showing the relative positions in the DNA molecule of the recognition sequences for a number of different enzymes. Only when a restriction map is available can the correct restriction endonucleases be selected for the particular cutting manipulation that is required (Figure 4.15).

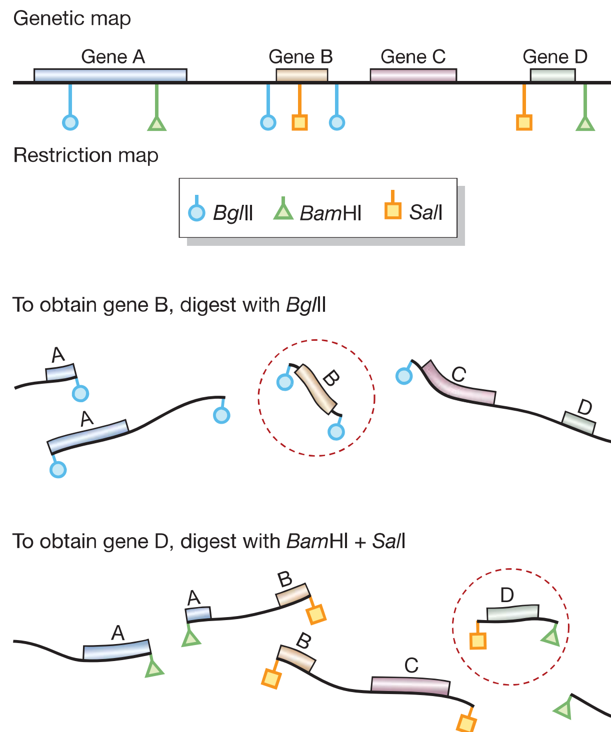


Figure 4.15

Using a restriction map to work out which restriction endonucleases should be used to obtain DNA fragments containing individual genes.

To construct a restriction map, a series of restriction digests must be performed. First, the number and sizes of the fragments produced by each restriction endonuclease must be determined by gel electrophoresis followed by comparison with size markers (Figure 4.16). This information must then be supplemented by a series of **double digestions**, in which the DNA is cut by two restriction endonucleases at once. It might be possible to perform a double digestion in one step if both enzymes have similar requirements for pH, Mg^{2+} concentration, etc. Alternatively, the two digestions may have to be carried out one after the other, adjusting the reaction mixture after the first digestion to provide a different set of conditions for the second enzyme.

Comparing the results of single and double digests will allow many, if not all, of the restriction sites to be mapped (Figure 4.16). Ambiguities can usually be resolved by **partial digestion**, carried out under conditions that result in cleavage of only a limited number of the restriction sites on any DNA molecule. Partial digestion is usually achieved by reducing the incubation period, so the enzyme does not have time to cut all the restriction sites, or by incubating at a low temperature (e.g., 4°C rather than 37°C), which limits the activity of the enzyme.

The result of a partial digestion is a complex pattern of bands in an electrophoresis gel. As well as the standard fragments, produced by total digestion, additional sizes are seen. These are molecules that comprise two adjacent restriction fragments, separated by a site that has not been cleaved. Their sizes indicate which restriction fragments in the complete digest are next to one another in the uncut molecule (Figure 4.16).

4.2.9 Special gel electrophoresis methods for separating larger molecules

During agarose gel electrophoresis, a DNA fragment migrates at a rate that is proportional to its size, but this relationship is not a direct one. The formula that links

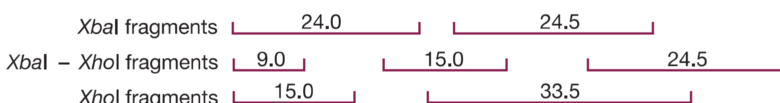
Single and double digestions

Enzyme	Number of fragments	Sizes (kb)
*Xba*I	2	24.0, 24.5
*Xho*I	2	15.0, 33.5
*Kpn*I	3	1.5, 17.0, 30.0
*Xba*I + *Xho*I	3	9.0, 15.0, 24.5
*Xba*I + *Kpn*I	4	1.5, 6.0, 17.0, 24.0

Conclusions:

(1) As λ DNA is linear, the number of restriction sites for each enzyme is *Xba*I 1, *Xho*I 1, *Kpn*I 2.

(2) The *Xba*I and *Xho*I sites can be mapped:



The only possibility is: *Xho*I *Xba*I 15.0 9.0 24.5

(3) All the *Kpn*I sites fall in the 24.5 kb *Xba*I fragment, as the 24.0 kb fragment is intact after *Xba*I–*Kpn*I double digestion. The order of the *Kpn*I fragments can be determined only by partial digestion.

Partial digestion

Enzyme	Fragment sizes (kb)
*Kpn*I – limiting conditions	1.5, 17.0, 18.5, 30.0, 31.5, 48.5

Conclusions:

(1) 48.5 kb fragment = uncut λ.
(2) 1.5, 17.0 and 30.0 kb fragments are products of complete digestion.
(3) 18.5 and 31.5 kb fragments are products of partial digestion.

The KpnI map must be: *Kpn*Is 30.0 1.5 17.0

Therefore the complete map is: *Xho*I *Xba*I *Kpn*Is 15.0 9.0 6.0 1.5 17.0

Figure 4.16

Restriction mapping. This example shows how the positions of the *Xba*I, *Xho*I and *Kpn*I sites on the λ DNA molecule can be determined.

migration rate to molecular mass has a logarithmic component (p. 59), which means that the difference in migration rates become increasingly small for larger molecules (Figure 4.17). In practice, molecules larger than about 50 kb cannot be resolved efficiently by standard gel electrophoresis.

This size limitation is not usually a problem when the restriction fragments being studied have been obtained by cutting the DNA with an enzyme with a tetranucleotide or hexanucleotide recognition sequence. Most, if not all, of the fragments produced in this way will be less than 30 kb in length and easily resolved by agarose gel electrophoresis. Difficulties might arise, however, if an enzyme with a longer recognition sequence is used, such as *Not*I, which cuts at an eight-nucleotide sequence (see Table 4.1). *Not*I would be expected, on average, to cut a DNA molecule once every

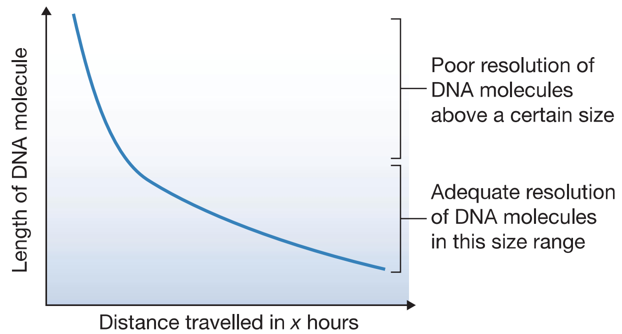
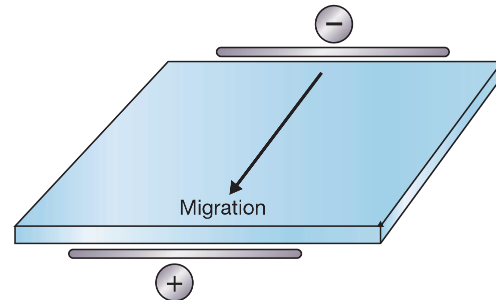
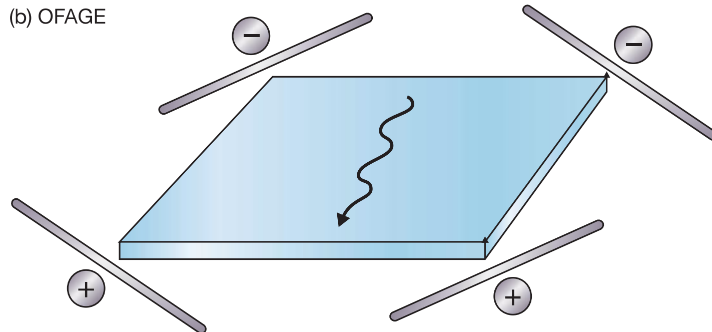


Figure 4.17
The influence of DNA size on migration rate during conventional gel electrophoresis.





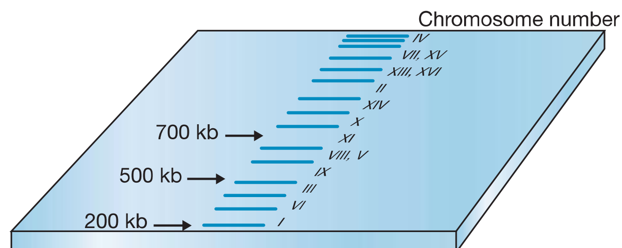


Figure 4.18
The difference between conventional gel electrophoresis and orthogonal field alternation gel electrophoresis (OFAGE).

4^8 = 65,536 bp. It is therefore unlikely that *Not*I fragments will be separated by standard gel electrophoresis.

The limitations of standard gel electrophoresis can be overcome if a more complex electric field is used. Several different systems have been designed, but the principle is best illustrated by **orthogonal field alternation gel electrophoresis (OFAGE)**. Instead of being applied directly along the length of the gel, as in the standard method (Figure 4.18a), the electric field now alternates between two pairs of electrodes, each

pair set at an angle of 45° to the length of the gel (Figure 4.18b). The result is a pulsed field, with the DNA molecules in the gel having continually to change direction in accordance with the pulses. As the two fields alternate in a regular fashion, the net movement of the DNA molecules in the gel is still from one end to the other, in more or less a straight line. However, with every change in field direction, each DNA molecule has to realign through 90° before its migration can continue. This is the key point, because a short molecule can realign faster than a long one, allowing the short molecule to progress toward the bottom of the gel more quickly. This added dimension increases the resolving power of the gel quite dramatically, so that molecules up to several thousand kilobases in length can be separated.

This size range includes not only restriction fragments but also the intact chromosomal molecules of many lower eukaryotes, including yeast, several important filamentous fungi, and protozoans such as the malaria parasite *Plasmodium falciparum*. OFAGE and related techniques such as **contour clamped homogeneous electric fields (CHEF)** and **field inversion gel electrophoresis (FIGE)** can therefore be used to prepare gels showing the separated chromosomes of these organisms (Figure 4.18c), enabling DNA from these individual chromosomes to be purified.

4.3 Ligation – joining DNA molecules together

The final step in construction of a recombinant DNA molecule is the joining together of the vector molecule and the DNA to be cloned (Figure 4.19). This process is referred to as ligation, and the enzyme that catalyzes the reaction is called DNA ligase.

4.3.1 The mode of action of DNA ligase

All living cells produce DNA ligases, but the enzyme used in genetic engineering is usually purified from *E. coli* bacteria that have been infected with T4 phage. Within the cell the enzyme carries out the very important function of repairing any discontinuities that may arise in one of the strands of a double-stranded molecule (see Figure 4.4a). A discontinuity is quite simply a position where a phosphodiester bond between adjacent nucleotides is missing (contrast this with a nick, where one or more nucleotides are absent). Although discontinuities may arise by chance breakage of the cell's DNA molecules, they are also a natural result of processes such as DNA replication and recombination. Ligases therefore play several vital roles in the cell.

In the test tube, purified DNA ligases, as well as repairing single-strand discontinuities, can also join together individual DNA molecules or the two ends of the same molecule. The chemical reaction involved in ligating two molecules is exactly the same as discontinuity repair, except that two phosphodiester bonds must be made, one for each strand (Figure 4.20a).

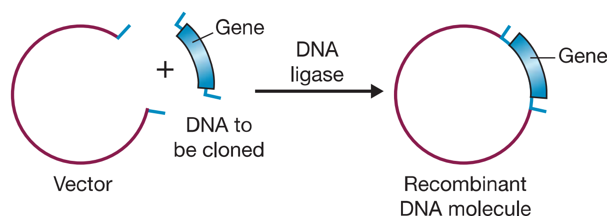


Figure 4.19
Ligation: the final step in construction of a recombinant DNA molecule.

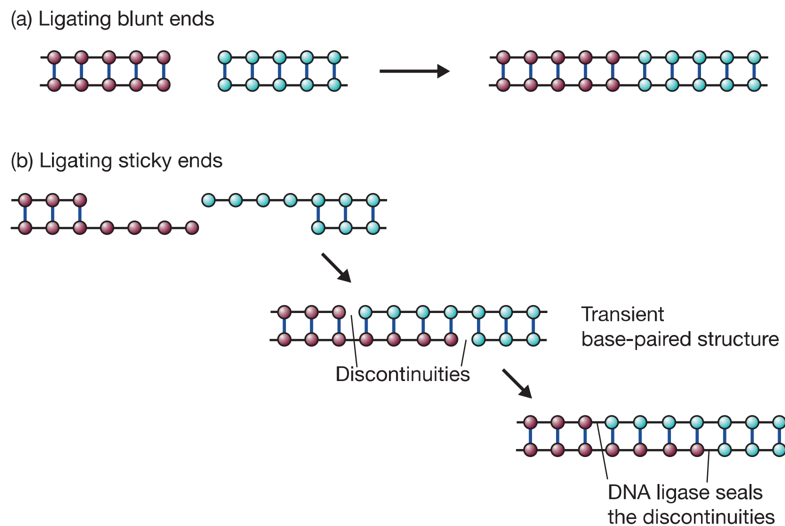


Figure 4.20

The different joining reactions catalysed by DNA ligase: (a) ligation of blunt-ended molecules; (b) ligation of sticky-ended molecules.

4.3.2 Sticky ends increase the efficiency of ligation

The ligation reaction in Figure 4.20a shows two blunt-ended fragments being joined together. Although this reaction can be carried out in the test tube, it is not very efficient. This is because the ligase is unable to "catch hold" of the molecule to be ligated, and has to wait for chance associations to bring the ends together. If possible, blunt end ligation should be performed at high DNA concentrations, to increase the chances of the ends of the molecules coming together in the correct way.

In contrast, ligation of complementary sticky ends is much more efficient. This is because compatible sticky ends can base pair with one another by hydrogen bonding (Figure 4.20b), forming a relatively stable structure for the enzyme to work on. If the phosphodiester bonds are not synthesized fairly quickly then the sticky ends fall apart again. These transient, base-paired structures do, however, increase the efficiency of ligation by increasing the length of time the ends are in contact with one another.

4.3.3 Putting sticky ends onto a blunt-ended molecule

For the reasons detailed in the preceding section, compatible sticky ends are desirable on the DNA molecules to be ligated together in a gene cloning experiment. Often these sticky ends can be provided by digesting both the vector and the DNA to be cloned with the same restriction endonuclease, or with different enzymes that produce the same sticky end, but it is not always possible to do this. A common situation is where the vector molecule has sticky ends, but the DNA fragments to be cloned are blunt-ended. Under these circumstances one of three methods can be used to put the correct sticky ends onto the DNA fragments.

Linkers

The first of these methods involves the use of linkers. These are short pieces of double-stranded DNA, of known nucleotide sequence, that are synthesized in the test tube.

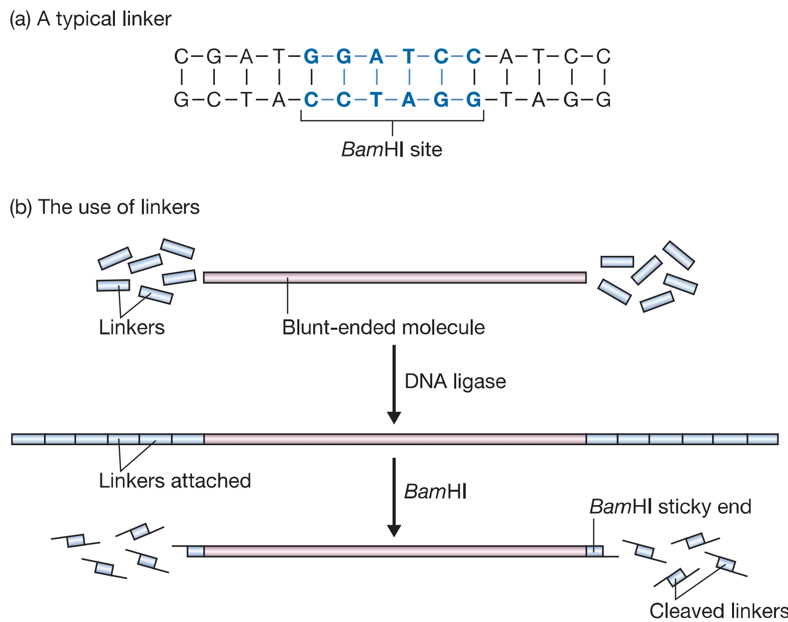


Figure 4.21

Linkers and their use: (a) the structure of a typical linker; (b) the attachment of linkers to a blunt-ended molecule.

A typical linker is shown in Figure 4.21a. It is blunt-ended, but contains a restriction site, *Bam*HI in the example shown. DNA ligase can attach linkers to the ends of larger blunt-ended DNA molecules. Although a blunt end ligation, this particular reaction can be performed very efficiently because synthetic oligonucleotides, such as linkers, can be made in very large amounts and added into the ligation mixture at a high concentration.

More than one linker will attach to each end of the DNA molecule, producing the chain structure shown in Figure 4.21b. However, digestion with *Bam*HI cleaves the chains at the recognition sequences, producing a large number of cleaved linkers and the original DNA fragment, now carrying *Bam*HI sticky ends. This modified fragment is ready for ligation into a cloning vector restricted with *Bam*HI.

Adaptors

There is one potential drawback with the use of linkers. Consider what would happen if the blunt-ended molecule shown in Figure 4.21b contained one or more *Bam*HI recognition sequences. If this was the case, the restriction step needed to cleave the linkers and produce the sticky ends would also cleave the blunt-ended molecule (Figure 4.22). The resulting fragments will have the correct sticky ends, but that is no consolation if the gene contained in the blunt-ended fragment has now been broken into pieces.

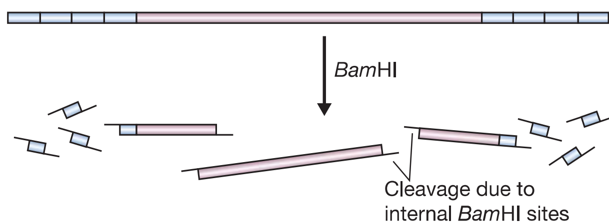
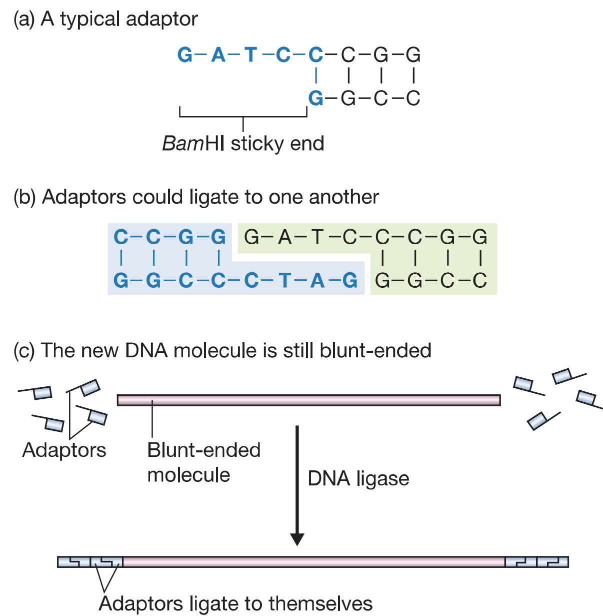


Figure 4.22

A possible problem with the use of linkers. Compare this situation with the desired result of *Bam*HI restriction, as shown in Figure 4.21(b).

Figure 4.23

Adaptors and the potential problem with their use. (a) A typical adaptor. (b) Two adaptors could ligate to one another to produce a molecule similar to a linker, so that (c) after ligation of adaptors a blunt-ended molecule is still blunt-ended and the restriction step is still needed.

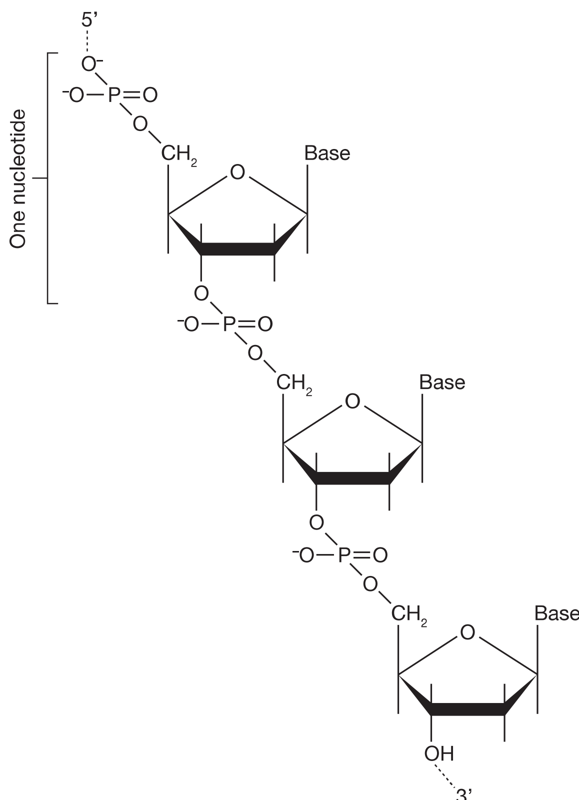


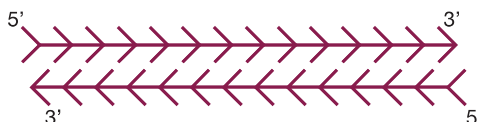
The second method of attaching sticky ends to a blunt-ended molecule is designed to avoid this problem. **Adaptors**, like linkers, are short synthetic oligonucleotides. But unlike linkers, an adaptor is synthesized so that it already has one sticky end (Figure 4.23a). The idea is of course to ligate the blunt end of the adaptor to the blunt ends of the DNA fragment, to produce a new molecule with sticky ends. This may appear to be a simple method but in practice a new problem arises. The sticky ends of individual adaptor molecules could base pair with each other to form dimers (Figure 4.23b), so that the new DNA molecule is still blunt-ended (Figure 4.23c). The sticky ends could be recreated by digestion with a restriction endonuclease, but that would defeat the purpose of using adaptors in the first place.

The answer to the problem lies in the precise chemical structure of the ends of the adaptor molecule. Normally the two ends of a polynucleotide strand are chemically distinct, a fact that is clear from a careful examination of the polymeric structure of DNA (Figure 4.24a). One end, referred to as the 5′ terminus, carries a phosphate group (5′-P); the other, the 3′ terminus, has a hydroxyl group (3′-OH). In the double helix the two strands are antiparallel (Figure 4.24b), so each end of a double-stranded molecule consists of one 5′-P terminus and one 3′-OH terminus. Ligation takes place between the 5′-P and 3′-OH ends (Figure 4.24c).

Adaptor molecules are synthesized so that the blunt end is the same as "natural" DNA, but the sticky end is different. The 3′-OH terminus of the sticky end is the same as usual, but the 5′-P terminus is modified: it lacks the phosphate group, and is in fact a 5′-OH terminus (Figure 4.25a). DNA ligase is unable to form a phosphodiester bridge between 5′-OH and 3′-OH ends. The result is that, although base pairing is always occurring between the sticky ends of adaptor molecules, the association is never stabilized by ligation (Figure 4.25b).

Adaptors can therefore be ligated to a blunt-ended DNA molecule but not to themselves. After the adaptors have been attached, the abnormal 5′-OH terminus is converted to the natural 5′-P form by treatment with the enzyme polynucleotide kinase (p. 50), producing a sticky-ended fragment that can be inserted into an appropriate vector.





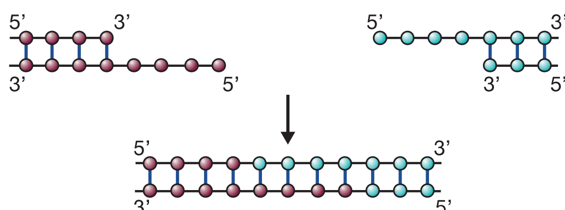


Figure 4.24

The distinction between the 5′ and 3′ termini of a polynucleotide.

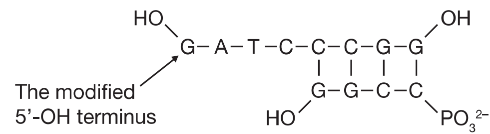
Producing sticky ends by homopolymer tailing

The technique of **homopolymer tailing** offers a radically different approach to the production of sticky ends on a blunt-ended DNA molecule. A homopolymer is simply a polymer in which all the subunits are the same. A DNA strand made up entirely of, say, deoxyguanosine is an example of a homopolymer, and is referred to as polydeoxyguanosine or poly(dG).

Tailing involves using the enzyme terminal deoxynucleotidyl transferase (p. 50) to add a series of nucleotides onto the 3′-OH termini of a double-stranded DNA molecule. If this reaction is carried out in the presence of just one deoxyribonucleotide, a homopolymer tail is produced (Figure 4.26a). Of course, to be able to ligate together

Figure 4.25

The use of adaptors: (a) the actual structure of an adaptor, showing the modified 5′-OH terminus; (b) conversion of blunt ends to sticky ends through the attachment of adaptors.



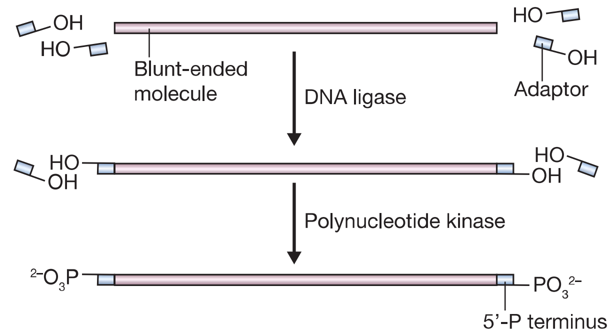
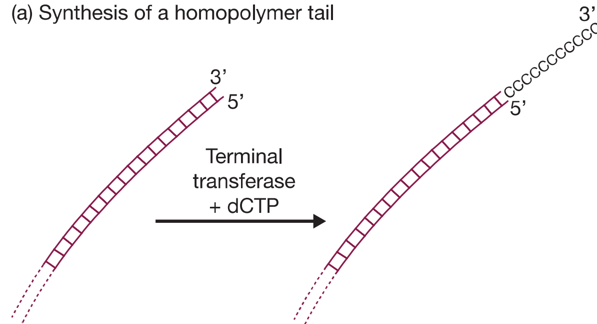
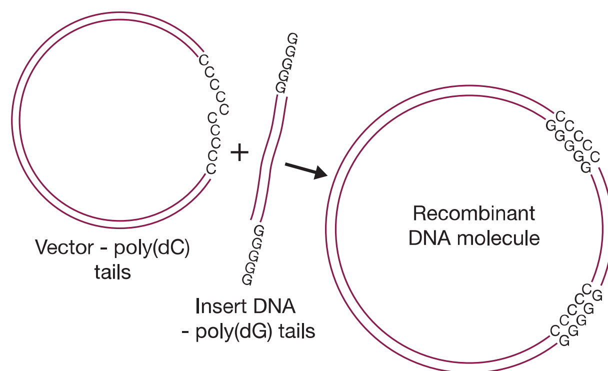
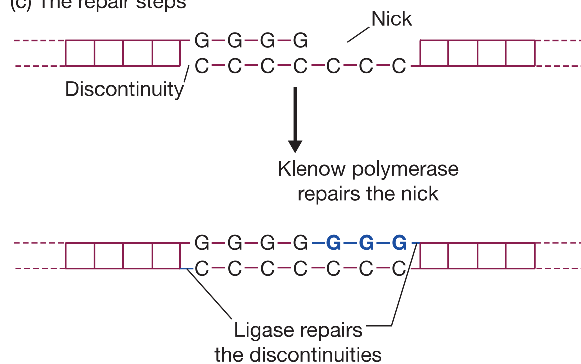


Figure 4.26

Homopolymer tailing: (a) synthesis of a homopolymer tail; (b) construction of a recombinant DNA molecule from a tailed vector plus tailed insert DNA; (c) repair of the recombinant DNA molecule.







two tailed molecules, the homopolymers must be complementary. Frequently polydeoxycytosine (poly(dC)) tails are attached to the vector and poly(dG) to the DNA to be cloned. Base pairing between the two occurs when the DNA molecules are mixed (Figure 4.26b).

In practice, the poly(dG) and poly(dC) tails are not usually exactly the same length, and the base-paired recombinant molecules that result have nicks as well as discontinuities (Figure 4.26c). Repair is therefore a two-step process, using Klenow polymerase to fill in the nicks followed by DNA ligase to synthesize the final phosphodiester bonds. This repair reaction does not always have to be performed in the test tube. If the complementary homopolymer tails are longer than about 20 nucleotides, then quite stable base-paired associations are formed. A recombinant DNA molecule, held together by base pairing although not completely ligated, is often stable enough to be introduced into the host cell in the next stage of the cloning experiment (see Figure 1.1). Once inside the host, the cell's own DNA polymerase and DNA ligase repair the recombinant DNA molecule, completing the construction begun in the test tube.

4.3.4 Blunt end ligation with a DNA topoisomerase

A more sophisticated, but easier and generally more efficient way of carrying out blunt end ligation, is to use a special type of enzyme called a **DNA topoisomerase**. In the cell, DNA topoisomerases are involved in processes that require turns of the double helix to be removed or added to a double-stranded DNA molecule. Turns are removed during DNA replication in order to unwind the helix and enable each polynucleotide to be replicated, and are added to newly synthesized circular molecules to introduce supercoiling. DNA topoisomerases are able to separate the two strands of a DNA molecule without actually rotating the double helix. They achieve this feat by causing transient single- or double-stranded breakages in the DNA backbone (Figure 4.27). DNA topoisomerases therefore have both nuclease and ligase activities.

To carry out blunt end ligation with a topoisomerase, a special type of cloning vector is needed. This is a plasmid that has been linearized by the nuclease activity of the DNA topoisomerase enzyme from vaccinia virus. The vaccinia topoisomerase cuts DNA at the sequence CCCTT, which is present just once in the plasmid. After cutting the

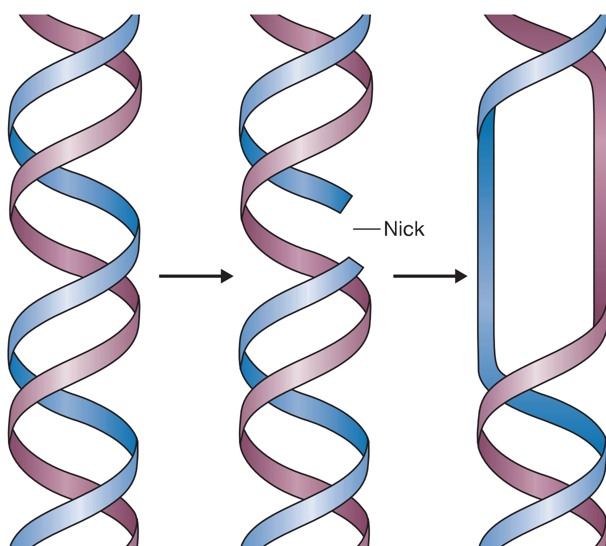


Figure 4.27

The mode of action of a Type 1 DNA topoisomerase, which removes or adds turns to a double helix by making a transient break in one of the strands.

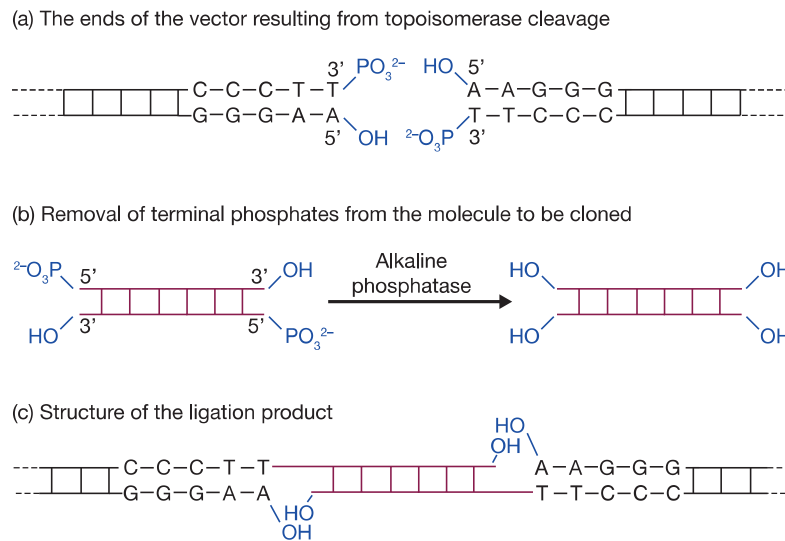


Figure 4.28

Blunt end ligation with a DNA topoisomerase. (a) Cleavage of the vector with the topoisomerase leaves blunt ends with 5′-OH and 3′-P termini. (b) The molecule to be cloned must therefore be treated with alkaline phosphatase to convert its 5′-P ends into 5′-OH termini. (c) The topoisomerase ligates the 3′-P and 5′-OH ends, creating a double-stranded molecule with two discontinuities, which are repaired by cellular enzymes after introduction into the host bacteria.

plasmid, topoisomerase enzymes remain covalently bound to the resulting blunt ends. The reaction can be stopped at this point, enabling the vector to be stored until it is needed.

Cleavage by the topoisomerase results in 5′-OH and 3′-P termini (Figure 4.28a). If the blunt-ended molecules to be cloned have been produced from a larger molecule by cutting with a restriction enzyme, then they will have 5′-P and 3′-OH ends. Before mixing these molecules with the vector, their terminal phosphates must be removed to give 5′-OH ends that can ligate to the 3′-P termini of the vector. The molecules are therefore treated with alkaline phosphatase (Figure 4.28b).

Adding the phosphatased molecules to the vector reactivates the bound topoisomerases, which proceed to the ligation phase of their reaction. Ligation occurs between the 3′-P ends of the vectors and the 5′-OH ends of the phosphatased molecules. The blunt-ended molecules therefore become inserted into the vectors. Only one strand is ligated at each junction point (Figure 4.28c), but this is not a problem because the discontinuities will be repaired by cellular enzymes after the recombinant molecules have been introduced into the host bacteria.

Further reading

FURTHER READING

Deng, G. & Wu, R. (1981) An improved procedure for utilizing terminal transferase to add homopolymers to the 3′ termini of DNA. *Nucleic Acids Research*, 9, 4173–4188.

Helling, R.B., Goodman, H.M. & Boyer, H.W. (1974) Analysis of endonuclease R·*Eco*RI fragments of DNA from lambdoid bacteriophages and other viruses by agarose-gel electrophoresis. *Journal of Virology*, 14, 1235–1244.

Heyman, J.A., Cornthwaite, J., Foncerrada, L. et al. (1999) Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Research*, 9, 383–392. [A description of ligation using topoisomerase.]

Jacobsen, H., Klenow, H. & Overgaard-Hansen, K. (1974) The N-terminal amino acid sequences of DNA polymerase I from *Escherichia coli* and of the large and small fragments obtained by a limited proteolysis. *European Journal of Biochemistry*, 45, 623–627. [Production of the Klenow fragment of DNA polymerase I.]

Lehnman, I.R. (1974) DNA ligase: structure, mechanism, and function. *Science*, 186, 790–797.

REBASE: http://rebase.neb.com/rebase/ [A comprehensive list of all the known restriction endonucleases and their recognition sequences.]

Rothstein, R.J., Lau, L.F., Bahl, C.P., Narang, N.A. & Wu, R. (1979) Synthetic adaptors for cloning DNA. *Methods in Enzymology*, 68, 98–109.

Schwartz, D.C. & Cantor, C.R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 37, 67–75.

Smith, H.O. & Wilcox, K.W. (1970) A restriction enzyme from *Haemophilus influenzae*. *Journal of Molecular Biology*, 51, 379–391. [One of the first full descriptions of a restriction endonuclease.]

Zipper, H., Brunner, H., Bernhagen, J. & Vitzthum, F. (2004) Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Research*, 32, e103. [Details of one the DNA dyes now used as an alternative to ethidium bromide for staining agarose gels.]