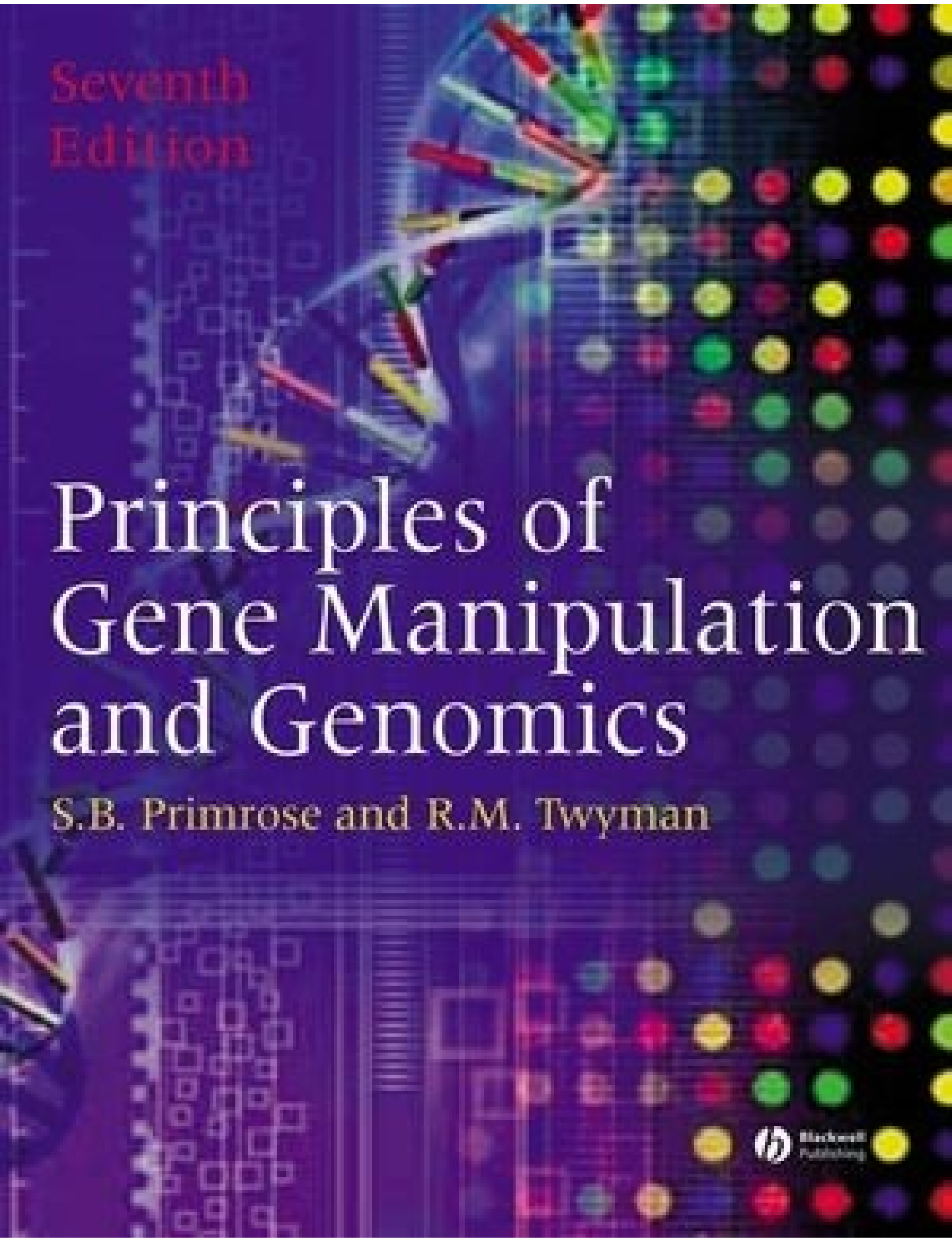

Seventh
Edition
Principles of
Gene Manipulation
and Genomics
S.B. Primrose and R.M. Twyman
Blackwell
Publishing

CHAPTER 3

Cutting and joining DNA molecules

Cutting DNA molecules

Before 1970 there was no method for cleaving DNA at discrete points. All the available methods for fragmenting DNA were non-specific. The available endonucleases had little site specificity and chemical methods produced very small fragments of DNA. The only method where any degree of control could be exercised was the use of mechanical shearing. The long, thin threads which constitute duplex DNA molecules are sufficiently rigid to be very easily broken by shear forces in solution. Intense sonication with ultrasound can reduce the length to about 300 nucleotide pairs. More controlled shearing can be achieved by high-speed stirring in a blender. Typically, high-molecular-weight DNA is sheared to a population of molecules with a mean size of about 8 kb by stirring at 1500 rev/min for 30 min (Wensink *et al.* 1974). Breakage occurs essentially at random with respect to DNA sequence. The termini consist of short, single-stranded regions which may have to be taken into account in subsequent joining procedures.

During the 1960s, phage biologists elucidated the biochemical basis of the phenomenon of host restriction and modification. The culmination of this work was the purification of the restriction endonuclease of *Escherichia coli* K12 by Meselson and Yuan (1968). Since this endonuclease cuts unmodified DNA into large discrete fragments, it was reasoned that it must recognize a target sequence. This in turn raised the prospect of controlled manipulation of DNA. Unfortunately, the K12 endonuclease turned out to be perverse in its properties. While the enzyme does bind to a defined recognition sequence, cleavage occurs at a "random" site several kilobases away (Yuan *et al.* 1980). The much sought-after breakthrough finally came in 1970 with the discovery in *Hemophilus influenzae* (Kelly & Smith 1970, Smith & Wilcox 1970) of an enzyme that behaves more simply. That is, the enzyme recognizes a particular target sequence in a duplex DNA molecule and breaks the polynucleotide chain within that sequence to give rise to discrete fragments of defined length and sequence.

The presence of restriction and modification systems is a double-edged sword. On the one hand, they provide a rich source of useful enzymes for DNA manipulation. On the other, these systems can significantly affect the recovery of recombinant DNA in cloning hosts. For this reason, some knowledge of restriction and modification is essential.

Understanding the biological basis of host-controlled restriction and modification of bacteriophage DNA led to the identification of restriction endonucleases

Restriction systems allow bacteria to monitor the origin of incoming DNA and to destroy it if it is recognized as foreign. Restriction endonucleases recognize specific sequences in the incoming DNA and cleave the DNA into fragments, either at specific sites or more randomly. When the incoming DNA is a bacteriophage genome, the effect is to reduce the efficiency of plating, i.e. to reduce the number of plaques formed in plating tests. The phenomena of restriction and modification were well illustrated and studied by the behavior of phage λ on two *E. coli* host strains.

If a stock preparation of phage λ, for example, is made by growth upon *E. coli* strain C and this stock is then titered upon *E. coli* C and *E. coli* K, the titers observed on these two strains will differ by several orders of magnitude, the titer on *E. coli* K being the lower. The phage are said to be *restricted* by the second host strain (*E. coli* K). When those phage that do result from the infection of *E. coli* K are now replated on *E. coli* K they are no longer restricted; but if they are first cycled through *E. coli* C they are once again restricted when plated upon *E. coli* K (Fig. 3.1). Thus the efficiency with which phage λ plates upon a particular host strain depends upon the strain on which it was last propagated. This non-heritable change conferred upon the phage by the second host strain

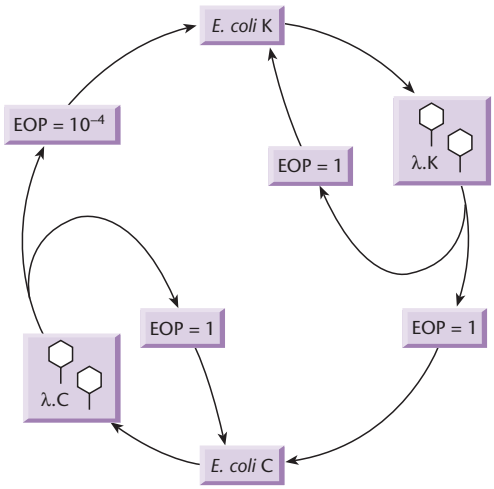


Fig. 3.1 Host-controlled restriction and modification of phage λ in *E. coli* strain K, analyzed by efficiency of plating (EOP). Phage propagated by growth on strains K or C (i.e. λ.K or λ.C) have EOPs on the two strains, as indicated by arrows.

(*E. coli* K) that allows it to be replated on that strain without further restriction is called *modification*.

The restricted phages adsorb to restrictive hosts and inject their DNA normally. When the phage are labeled with ^{32}P, it is apparent that their DNA is degraded soon after injection (Dussoix & Arber 1962) and the endonuclease that is primarily responsible for this degradation is called a *restriction endonuclease* or restriction enzyme (Lederberg & Meselson 1964). The restrictive host must, of course, protect its own DNA from the potentially lethal effects of the restriction endonuclease and so its DNA must be appropriately modified. Modification involves methylation of certain bases at a very limited number of sequences within DNA, which constitute the recognition sequences for the restriction endonuclease. This explains why phage that survive one cycle of growth upon the restrictive host can subsequently reinfect that host efficiently; their DNA has been replicated in the presence of the modifying methylase and so it, like the host DNA, becomes methylated and protected from the restriction system. Although phage infection has been chosen as our example to illustrate restriction and modification, these processes can occur whenever DNA is transferred from one bacterial strain to another.

Four different types of restriction and modification (R-M) system have been recognized but only one is widely used in gene manipulation

At least four different kinds of R-M system are known: type I, type II, type III, and type IIs. The essential differences between them are summarized in Table 3.1.

The type I systems were the first to be characterized and a typical example is that from *E. coli* K12. The active enzyme consists of two restriction subunits, two modification (methylation) subunits, and one recognition subunit. These subunits are the products of the *hsd*R, *hsd*M, and *hsd*S genes. The methylation and cutting reactions both require ATP and *S*-adenosylmethionine as cofactors. The recognition sequences are quite long with no recognizable features such as symmetry. The enzyme cuts unmodified DNA at some distance from the recognition sequence. However, because the methylation reaction is performed by the same enzyme which mediates cleavage, the target DNA may be modified before it is cut. These features mean that type I systems are of little value for gene manipulation (see also Box 3.1). However, their presence in *E. coli*

Table 3.1 Characteristics of the different types of endonucleases.

System	Key features
Type I	One enzyme with different subunits for recognition, cleavage, and methylation. Recognizes and methylates a single sequence but cleaves DNA up to 1000 bp away
Type II	Two different enzymes which both recognize the same target sequence, which is symmetrical. The two enzymes either cleave or modify the recognition sequence
Type III	One enzyme with two different subunits, one for recognition and modification and one for cleavage. Recognizes and methylates same sequence but cleaves 24–26 bp away
Type IIs	Two different enzymes but recognition sequence is asymmetric. Cleavage occurs on one side of recognition sequence up to 20 bp away

Box 3.1 Restriction: from a phenomenon in bacterial genetics to a biological revolution

In the two related phenomena of host-controlled restriction and modification, a single cycle of phage growth in a particular host bacterium, alters the host range of the progeny virus. The virus may fail to plate efficiently on a second host; its host range is restricted. This modification of the virus differs fundamentally from mutation because it is imposed by the host cell on which the virus has been grown but it is not inherited; when the phage is grown in some other host, the modification may be lost. In the 1950s, restriction and modification were recognized as common phenomena, affecting many virulent and temperate (i.e. capable of forming lysogens) phages and involving various bacterial species (Luria 1953, Lederberg 1957).

Arber and Dussoix clarified the molecular basis of the phenomena. They showed that restriction of phage λ is associated with rapid breakdown of the phage DNA in the host bacterium. They also showed that modification results from an alteration of the phage DNA which renders the DNA insensitive to restriction. They deduced that a single modified strand in the DNA duplex is sufficient to prevent restriction (Arber & Dussoix 1962, Dussoix & Arber 1962). Subsequent experiments implicated methylation of the DNA in the modification process (Arber 1965).

Detailed genetic analysis, in the 1960s, of the bacterial genes (in *E. coli* K and *E. coli* B) responsible for restriction and modification supported the duality of the two phenomena. Mutants of the bacteria could be isolated that were both restriction-deficient and modification-deficient (R^-M^-), or that were R^-M^+. The failure to recover R^+M^- mutants was correctly ascribed to the suicidal failure to confer protective modification upon the host's own DNA.

The biochemistry of restriction advanced with the isolation of the restriction endonuclease from *E. coli* K (Meselson & Yuan 1968). It was evident that the restriction endonucleases from *E. coli* K and *E. coli* B were important examples of proteins that recognize specific structures in DNA, but the properties of these type I enzymes as they are now known, were complex. Although the recognition sites in the phage could be mapped genetically (Murray *et al.* 1973a), determined efforts to define the DNA sequences cleaved were unsuccessful (Eskin & Linn 1972, Murray *et al.* 1973b).

The breakthrough came with Hamilton Smith's discovery of a restriction endonuclease from *Hemophilus influenzae* strain Rd (Smith & Wilcox 1970) and the elucidation of the nucleotide sequence at its cleavage sites in phage T7 DNA (Kelly & Smith 1970). This enzyme is now known as *Hin*dII. The choice of T7 DNA as the substrate for cleavage was a good one, because the bacterium also contains another type II restriction enzyme, *Hin*dIII, in abundance. Fortunately, *Hin*dIII does not cleave T7 DNA, and so any contaminating *Hin*dIII in the *Hin*dII preparation could not be problematical (Old *et al*. 1975). Shortly after the discovery of *Hin*dII, several other type II restriction endonucleases were isolated and characterized. *Eco*RI was foremost among these (Hedgepeth *et al*. 1972). They were rapidly exploited in the first recombinant DNA experiments.

By the mid-1960s, restriction and modification had been recognized as important and interesting phenomena within the field of bacterial genetics (see, for example, Hayes 1968), but who could have foreseen the astonishing impact of restriction enzymes upon biology?

strains can affect recovery of recombinants (see p. 24). Type III enzymes have symmetrical recognition sequences but otherwise resemble type I systems and are of little value.

Most of the useful R-M systems are of type II. They have a number of advantages over type I and III systems. First, restriction and modification are mediated by separate enzymes so it is possible to cleave DNA in the absence of modification. Secondly, the restriction activities do not require cofactors such as ATP or *S*-adenosylmethionine, making them easier to use. Most important of all, type II enzymes recognize a defined, usually symmetrical, sequence *and cut within it*. Many of them also make a staggered break in the DNA and the usefulness of this will become apparent. Although type IIs systems have similar cofactors and macromolecular structure to those of type II systems, the fact that restriction occurs at a distance from the recognition site limits their usefulness.

The classification of R-M systems into types I to III is convenient but may require modification as new discoveries are made. For example, the *Eco*571 system comprises a single polypeptide which has both restriction and modification activities (Petrusyte *et al.* 1988). Other restriction systems are known which fall outside the above classification. Examples include the *mcr* and *mrr* systems (see p. 43) and homing endonucleases. The latter are double-stranded deoxyribonucleases (DNases) derived from introns or inteins (Belfort & Roberts 1997). They have large, asymmetric recognition sequences and, unlike standard restriction endonucleases, tolerate some sequence degeneracy within their recognition sequence.

The naming of restriction endonucleases provides information about their source

The discovery of a large number of restriction and modification systems called for a uniform system of nomenclature. A suitable system was proposed by Smith and Nathans (1973) and a simplified version of this is in use today. The key features are:

- The species name of the host organism is identified by the first letter of the genus name and the first two letters of the specific epithet to generate a three-letter abbreviation. This abbreviation is always written in italics.
- Where a particular strain has been the source then this is identified.
- When a particular host strain has several different R-M systems, these are identified by roman numerals.

Some examples are given in Table 3.2.

Homing endonucleases are named in a similar fashion except that intron-encoded endonucleases are given the prefix "I-" (e.g. I-*Ceu*I) and intein endonucleases have the prefix "PI-" (e.g. Pl-*Psp*I). Where it is necessary to distinguish between the restriction and methylating activities, they are given the prefixes "R" and "M", respectively, e.g. R.*Sma*I and M.*Sma*I.

Restriction enzymes cut DNA at sites of rotational symmetry and different enzymes recognize different sequences

Most, but not all, type II restriction endonucleases recognize and cleave DNA within particular sequences of four to eight nucleotides which have a twofold axis of *rotational symmetry*. Such sequences are often referred to as *palindromes* because of their similarity to words that read the same backwards as forwards. For example, the restriction and modification enzymes R.*Eco*RI and M.*Eco*RI recognize the sequence:

5′-G A A ⋮ T T C-3′
3′-C T T ⋮ A A G-5′
Axis of symmetry

Table 3.2 Examples of restriction endonuclease nomenclature.

Enzyme	Enzyme source	Recognition sequence
*Sma*I	*Serratia marcescens*, 1st enzyme	CCCGGG
*Hae*III	*Hemophilus aegyptius*, 3rd enzyme	GGCC
*Hin*dII	*Hemophilus influenzae*, strain d, 2nd enzyme	GTPyPuAC
*Hin*dIII	*Hemophilus influenzae*, strain d, 3rd enzyme	AAGCTT
*Bam*HI	*Bacillus amyloliquefaciens*, strain H, 1st enzyme	GGATCC

The position at which the restricting enzyme cuts is usually shown by the symbol "/" and the nucleotides methylated by the modification enzyme are usually marked with an asterisk. For *Eco*RI these would be represented thus:

5′-G/AA*T T C-3′
3′-C TT A*A/G-5′

For convenience it is usual practice to simplify the description of recognition sequences by showing only one strand of DNA, that which runs in the 5′ to 3′ direction. Thus the *Eco*RI recognition sequence would be shown as G/AATTC.

From the information shown above we can see that *Eco*RI makes single-stranded breaks four bases apart in the opposite strands of its target sequence so generating fragments with protruding 5′ termini:

5′-G 5′-AATTC-3′
3′-CTTAA-5′ G-5′

These DNA fragments can associate by hydrogen bonding between overlapping 5′ termini, or the fragments can circularize by intramolecular reaction (Fig. 3.2). For this reason the fragments are said to have *sticky* or *cohesive* ends. In principle, DNA fragments from diverse sources can be joined by means of the cohesive ends and, as we shall see later, the nicks in the molecules can be sealed to form an intact *artificially recombinant* DNA molecule.

Not all type II enzymes cleave their target sites like *Eco*RI. Some, such as *Pst*I (CTGCA/G), produce fragments bearing 3′ overhangs, while others, such as *Sma*I (CCC/GGG), produce *blunt* or *flush* ends.

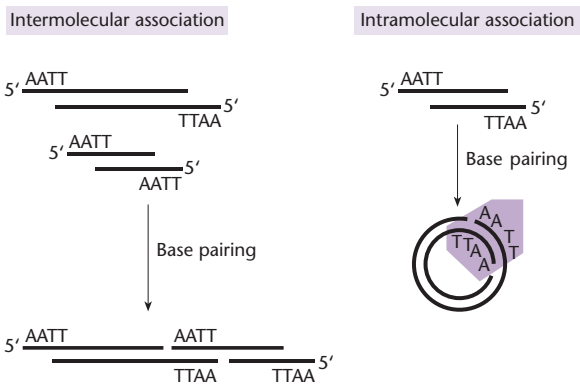


Fig. 3.2 Cohesive ends of DNA fragments produced by digestion with *Eco*RI.

Table 3.3 Some restriction endonucleases and their recognition sites.

Enzyme	Recognition sequence
4-base cutters	
*Mbo*I, *Dpn*I, *Sau*3AI	/GATC
*Msp*I, *Hpa*II	C/CGG
*Alu*I	AG/CT
*Hae*III	GG/CC
*Taq*I	ACGT/
6-base cutters	
*Bgl*II	A/GATCT
*Cla*I	AT/CGAT
*Pvu*II	CAG/CTG
*Pvu*I	CGAT/CG
*Kpn*I	GGTAC/C
8-base cutters	
*Not*I	GC/GGCCGC
*Sbf*I	CCTGCA/GG

To date, over 10,000 microbes from around the world have been screened for restriction enzymes. From these, over 3000 enzymes have been found representing approximately 200 different sequence specificities. Some representative examples are shown in Table 3.3. For a comprehensive database of information on restriction endonucleases and their associated methylases, including cleavage sites, commercial availability, and literature references, the reader should consult the website maintained by New England Biolabs (www.rebase.neb.com).

Occasionally enzymes with novel DNA sequence specificities are found but most prove to have the same specificity as enzymes already known. Restriction enzymes with the same sequence specificity and cut site are known as *isoschizomers*. Enzymes that recognize the same sequence but cleave at different points, for example *Sma*I (CCC/GGG) and *Xma*I (C/CCGGG), are sometimes known as *neoschizomers*.

Under extreme conditions, such as elevated pH or low ionic strength, restriction endonucleases are capable of cleaving sequences which are similar but not identical to their defined recognition sequence. This altered specificity is known as *star* activity. The most common types of altered activity are acceptance of base substitutions and truncation of the number of bases in the recognition sequence. For example, *Eco*RI* (*Eco*RI star activity) cleaves the sequence N/AATTN, where N is any base, whereas *Eco*RI cleaves the sequence GAATTC.

The G+C content of a DNA molecule affects its susceptibility to different restriction endonucleases

The number and size of the fragments generated by a restriction enzyme depend on the frequency of occurrence of the target site in the DNA to be cut. Assuming a DNA molecule with a 50% G+C content and a random distribution of the four bases, a four-base recognition site occurs every 4^4 (256) bp. Similarly, a six-base recognition site occurs every 4^6 (4096) bp and an eight-base recognition sequence every 4^8 (65,536) bp. In practice, there is not a random distribution of the four bases and many organisms can be AT- or GC-rich, e.g. the nuclear genome of mammals is 40% G+C and the dinucleotide CG is fivefold less common than statistically expected. Similarly, CCG and CGG are the rarest trinucleotides in most A+T-rich bacterial genomes and CTAG is the rarest tetranucleotide in G+C-rich bacterial genomes. Thus different restriction endonucleases with six-base recognition sites can produce average fragment sizes significantly different from the expected 4096 bp (Table 3.4).

Certain restriction endonucleases show preferential cleavage of some sites in the same DNA molecule. For example, phage λ DNA has five sites for *Eco*RI but the different sites are cleaved non-randomly (Thomas & Davis 1975). The site nearest the right terminus is cleaved 10 times faster than the sites in the middle of the molecule. There are four sites for *Sac*II in λ DNA but the three sites in the center of the molecule are cleaved 50 times faster than the remaining site. There is a group of three restriction enzymes which show an even more dramatic site preference. These are *Nar*I, *Nae*I, and *Sac*II and they require simultaneous interaction with two copies of their recognition sequence before they will cleave DNA (Kruger 1988, Conrad & Topal, 1989). Thus *Nar*I will rapidly cleave two of the four recognition sites on plasmid pBR322 DNA but will seldom cleave the remaining two sites.

Simple DNA manipulations can convert a site for one restriction enzyme into a site for another enzyme

In order to join two fragments of DNA together, it is not essential that they are produced by the same restriction endonuclease. Many different restriction endonucleases produce compatible cohesive ends. For example, *Age*I (A/CCGGT) and *Ava*I (C/CCGGG) produce molecules with identical 5′ overhangs and so can be ligated together. There are many other examples of compatible cohesive ends. What is more, if the cohesive ends were produced by six-base cutters, the ligation products are often recleavable by four-base cutters. Thus, in the example cited above, the hybrid site ACCGGG can be cleaved by *Hpa*II (C/CGG), *Nci*I (CC/GGG), or *ScrF*I (CC/NGG).

New restriction sites can be generated by filling in the overhangs generated by restriction endonucleases and ligating the products together. Figure 3.3 shows that after filling in the cohesive ends produced by *Eco*RI, ligation produces restriction sites recognized by four other enzymes. Many other examples of creating new target sites by filling and ligation are known.

There are also many examples of combinations of blunt-end restriction endonucleases that produce recleavable ligation products. For example, when molecules generated by cleavage with *Alu*I (AG/CT) are joined to ones produced by *Eco*RV (GAT/ATC), some of the ligation sites will have the sequence GATCT and others will have the sequence AGATC. Both can be cleaved by *Mbo*I (GATC).

A methyltransferase, M.*Sss*I, that methylates the dinucleotide CpG (Nur *et al.* 1985) has been isolated from *Spiroplasma*. This enzyme can be used to modify

Table 3.4 Average fragment size (bp) produced by different enzymes with DNA from different sources.

Enzyme	Target	*Arabidopsis*	Nematode	*Drosophila*	*E. coli*	Human
*Apa*I	GGGCCC	25,000	40,000	6,000	15,000	2,000
*Avr*II	CCTAGG	15,000	20,000	20,000	150,000	8,000
*Bam*HI	GGATCC	6,000	9,000	4,000	5,000	5,000
*Dra*I	TTTAAA	2,000	1,000	1,000	2,000	2,000
*Spe*I	ACTAGT	8,000	8,000	9,000	60,000	10,000

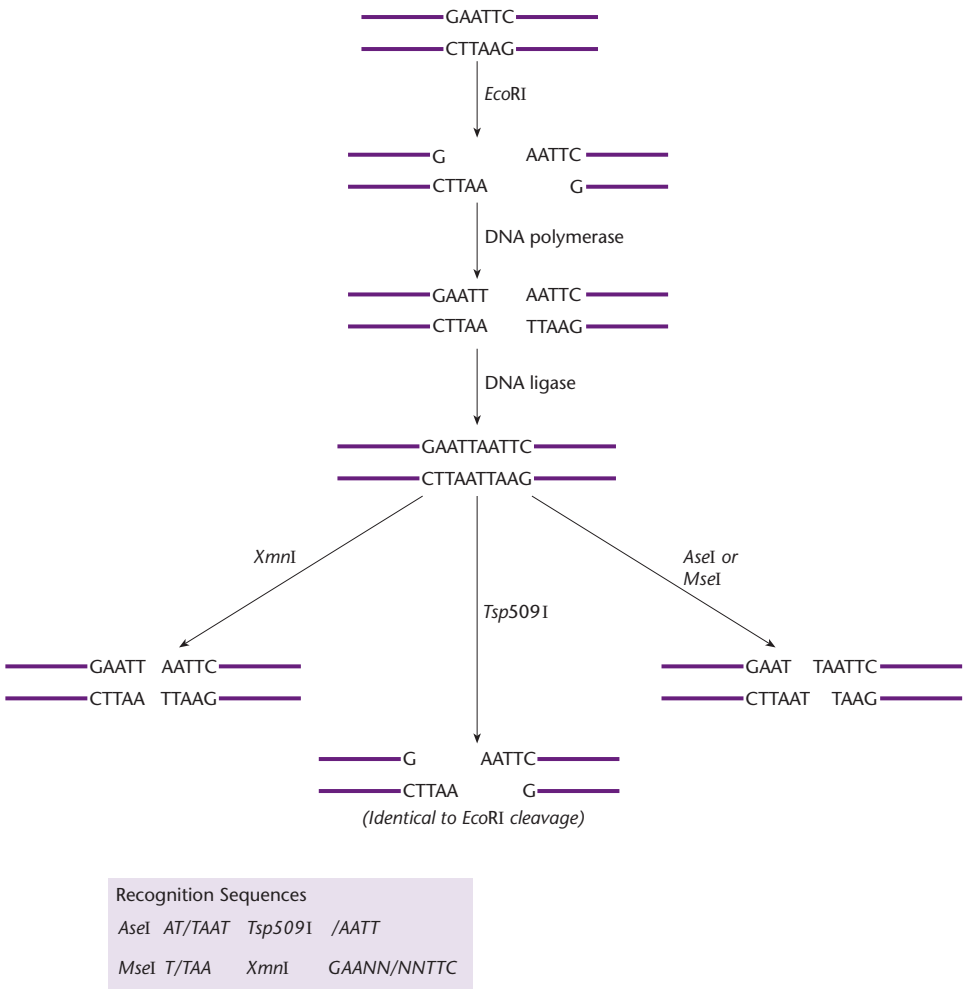


Fig. 3.3 The generation of three new restriction sites after filling in the overhangs produced by endonuclease *Eco*RI and ligating the products together. Note that there are two target sites, 4 bp apart, in the reconstituted molecule for endonuclease *Tsp*509I.

in vitro restriction endonuclease target sites which contain the CG sequence. Some of the target sequences modified in this way will be resistant to endonuclease cleavage, while others will remain sensitive. For example, if the sequence CCGG is modified with *Sss*I, it will be resistant to *Hpa*II but sensitive to *Msp*I. Since 90% of the methyl groups in the genomic DNA of many animals, including vertebrates and echinoderms, occur as 5-methylcytosine in the sequence CG, M.*Sss* can be used to imprint DNA from other sources with a vertebrate pattern.

Methylation can reduce the susceptibility of DNA to cleavage by restriction endonucleases and the efficiency of DNA transformation

Most laboratory strains of *E. coli* contain three site-specific DNA methylases. The methylase encoded by the *dam* gene transfers a methyl group from *S*-adenosylmethionine to the N^6 position of the adenine residue in the sequence GATC. The methylase encoded by the *dcm* gene (the Dcm methylase, previously called the Mec methylase) modifies the internal cytosine residues in the sequences CCAGG and CCTGG at the C^5 position (Marinus *et al.* 1983). In DNA in which the GC content is 50%, the sites for these two methylases occur, on average, every 256–512 bp. The third methylase is the enzyme M.*Eco*KI but the sites for this enzyme are much rarer and occur about once every 8 kb.

These enzymes are of interest for two reasons. First, some or all of the sites for a restriction endonuclease may be resistant to cleavage when isolated from strains expressing the Dcm or Dam methylases. This occurs when a particular base in the recognition site of a restriction endonuclease is methylated. The relevant base may be methylated by one of the *E. coli* methylases if the methylase recognition site overlaps the endonuclease recognition site. For example, DNA isolated from Dam^+ *E. coli* is completely

resistant to cleavage by *Mbo*I, but not *Sau*3AI, both of which recognize the sequence GATC. Similarly, DNA from a Dcm$^+$ strain will be cleaved by *Bst*NI but not by *Eco*RII, even though both recognize the sequence CCATGG. It is worth noting that most cloning strains of *E. coli* are Dam$^+$ Dcm$^+$ but double mutants are available (Marinus *et al.* 1983).

The second reason these methylases are of interest is that the modification state of plasmid DNA can affect the frequency of transformation in special situations. Transformation efficiency will be reduced when Dam-modified plasmid DNA is introduced into Dam$^-$ *E. coli* or Dam- or Dcm-modified DNA is introduced into other species (Russell & Zinder 1987). When DNA is to be moved from *E. coli* to another species it is best to use a strain lacking the Dam and Dcm methylases.

As will be seen later, it is difficult to clone DNA that contains short, direct repetitive sequences stably. Deletion of the repeat units occurs quickly, even when the host strain is deficient in recombination. However, the deletion mechanism appears to involve Dam methylation, for it does not occur in *dam* mutants (Troester *et al.* 2000).

It is important to eliminate restriction systems in *E. coli* strains used as hosts for recombinant DNA

If foreign DNA is introduced into an *E. coli* host it may be attacked by restriction systems active in the host cell. An important feature of these systems is that the fate of the incoming DNA in the restrictive host depends not only on the sequence of the DNA but also upon its history: the DNA sequence may or may not be restricted, depending upon its source immediately prior to transforming the *E. coli* host strain. As we have seen, post-replication modifications of the DNA, usually in the form of methylation of particular adenine or cytosine residues in the target sequence, protect against cognate restriction systems but not, in general, against different restriction systems.

Because restriction provides a natural defense against invasion by foreign DNA, it is usual to employ a K restriction-deficient *E. coli* K12 strain as a host in transformation with newly created recombinant molecules. Thus where, for example, mammalian DNA has been ligated into a plasmid vector, transformation of the *Eco*K restriction-deficient host eliminates the possibility that the incoming sequence will be restricted, even if the mammalian sequence

Fig. 3.4 The immigration control region of *E. coli* strain K12.

contains an unmodified *Eco*K target site. If the host happens to be *Eco*K restriction-deficient but *Eco*K modification-*proficient*, propagation on the host will confer modification methylation and hence allow subsequent propagation of the recombinant in *Eco*K restriction-proficient strains, if desired.

Whereas the *Eco*KI restriction system, encoded by the *hsd*RMS genes, cleaves DNA that is not protected by methylation at the target site, the McrA, McrBC, and Mrr endonucleases cut DNA that is methylated at specific positions. All three endonucleases restrict DNA modified by CpC methylase (M.*Sss*I) and the Mrr endonuclease will attack DNA with methyladenine in specific sequences. The significance of these restriction enzymes is that DNA from many bacteria, and from all plants and higher animals, is extensively methylated and its recovery in cloning experiments will be greatly reduced if the restriction activity is not eliminated. There is no problem with DNA from *Saccharomyces cerevisiae* or *Drosophila melanogaster* since there is little methylation of their DNA.

All the restriction systems in *E. coli* are clustered together in an "immigration control region" about 14 kb in length (Fig. 3.4). Some strains carry mutations in one of the genes. For example, strains DH1 and DH5 have a mutation in the *hsd*R gene and so are defective for the *Eco*KI endonuclease but still mediate the *Eco*KI modification of DNA. Strain DP50 has a mutation in the *hsd*S gene and so lacks both the *Eco*KI restriction and modification activities. Other strains, such as *E. coli* C and the widely used cloning strain HB101, have a deletion of the entire *mcrC–mrr* region and hence lack all restriction activities.

The success of a cloning experiment is critically dependent on the quality of any restriction enzymes that are used

Restriction enzymes are available from many different commercial sources. In choosing a source of enzyme, it is important to consider the quality of the enzyme supplied. High-quality enzymes are purified extensively to remove contaminating exonucleases and endonucleases and tests for the absence of such contaminants form part of routine quality control

(QC) on the finished product (see below). The absence of exonucleases is particularly important. If they are present, they can nibble away the overhangs of cohesive ends, thereby eliminating or reducing the production of subsequent recombinants. Contaminating phosphatases can remove the terminal phosphate residues, thereby preventing ligation. Even where subsequent ligation is achieved, the resulting product may contain small deletions. The message is clear: cheap restriction enzymes are in reality poor value for money!

A typical QC procedure is as follows. DNA fragments are produced by an excessive overdigestion of substrate DNA with each restriction endonuclease. These fragments are then ligated and recut with the same restriction endonuclease. Ligation can occur only if the 3′ and 5′ termini are left intact, and only those molecules with a perfectly restored recognition site can be recleaved. A normal banding pattern after cleavage indicates that both the 3′ and 5′ termini are intact and the enzyme preparation is free of detectable exonucleases and phosphatases (Fig. 3.5).

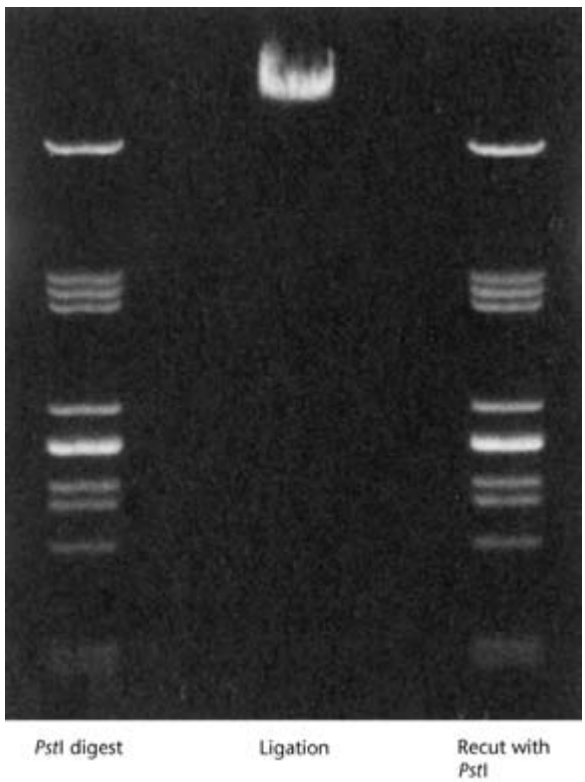


Fig. 3.5 Quality control of the enzyme *Pst*I. DNA was overdigested with the endonuclease and the fragments were ligated together and then recut. Note that the two digests give an identical banding pattern upon agarose gel electrophoresis.

An additional QC test is the *blue/white screening* assay. In this, the starting material is a plasmid carrying the *E. coli lac*Z′ gene in which there is a single recognition site for the enzyme under test. The plasmid is overdigested with the restriction enzyme, religated and transformed into a lacZ$^-$ strain of *E. coli.* The transformants are plated on media containing the β-galactosidase substrate Xgal. If the *lac*Z gene remains intact after digestion and ligation, it will give rise to a blue colony. If any degradation of the cut ends occurred, then a white colony will be produced (Box 3.2).

Joining DNA molecules

Having described the methods available for cutting DNA molecules, we must consider the ways in which DNA fragments can be joined to create artificially recombinant molecules. There are currently three methods for joining DNA fragments *in vitro*. The first of these capitalizes on the ability of DNA ligase to join covalently the annealed cohesive ends produced by certain restriction enzymes. The second depends upon the ability of DNA ligase from phage T4-infected *E. coli* to catalyze the formation of phosphodiester bonds between blunt-ended fragments. The third utilizes the enzyme terminal deoxynucleotidyltransferase to synthesize homopolymeric 3′ single-stranded tails at the ends of fragments. We can now look at these three methods a little more deeply.

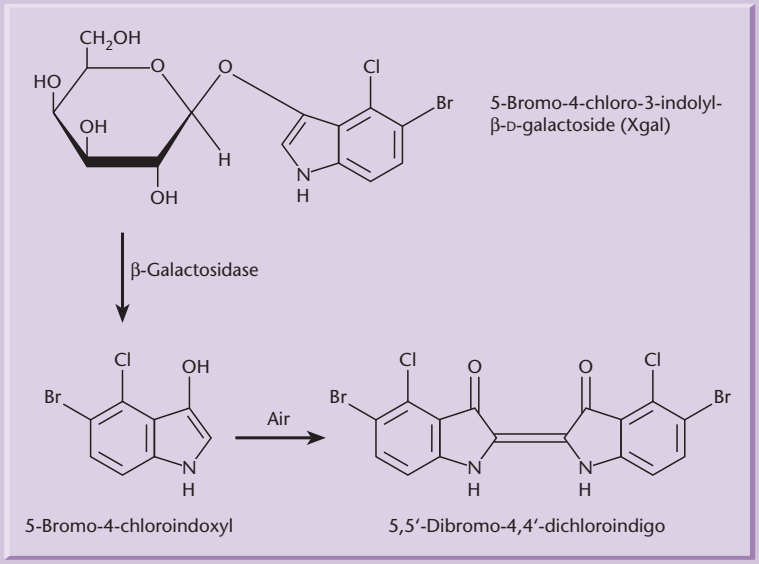
The enzyme DNA ligase is the key to joining DNA molecules *in vitro*

E. coli and phage T4 encode an enzyme, DNA ligase, which seals single-stranded nicks between adjacent nucleotides in a duplex DNA chain (Olivera *et al.* 1968, Gumport & Lehman 1971). Although the reactions catalyzed by the enzymes of *E. coli* and T4-infected *E. coli* are very similar, they differ in their cofactor requirements. The T4 enzyme requires ATP, while the *E. coli* enzyme requires NAD^+. In each case the cofactor is split and forms an enzyme–AMP complex. The complex binds to the nick, which must expose a 5′ phosphate and 3′ OH group, and makes a covalent bond in the phosphodiester chain, as shown in Fig. 3.6.

When termini created by a restriction endonuclease that creates cohesive ends associate, the joint has nicks a few base pairs apart in opposite strands. DNA ligase can then repair these nicks to form an intact duplex. This reaction, performed *in vitro* with purified

Box 3.2 α-Complementation of β-galactosidase and the use of Xgal

The activity of the enzyme β-galactosidase is easily monitored by including in the growth medium the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Xgal). This compound is colorless but on cleavage releases a blue indolyl derivative. On solid medium, colonies that are expressing active β-galactosidase are blue in color while those without the activity are white in color. This is often referred to as blue/white screening. Since Xgal is not an inducer of β-galactosidase, the non-substrate (*gratuitous*) inducer isopropyl-β-D-thiogalactoside (IPTG) is also added to the medium.



The phenomenon of α-complementation of β-galactosidase is widely used in molecular genetics. The starting-point for α-complementation is the M15 mutant of *E. coli*. This has a deletion of residues 11–41 in the *lac*Z gene and shows no β-galactosidase activity. Enzyme activity can be restored to the mutant enzyme *in vitro* by adding a cyanogen bromide peptide derived from amino acid residues 3–92 (Langley *et al.* 1975, Langley & Zabin 1976). Complementation can also be shown *in vivo*. If a plasmid carrying the N-terminal fragment of the *lac*Z gene encompassing the missing region is introduced into the M15 mutant, then β-galactosidase is produced, as demonstrated by the production of a blue color on medium containing Xgal. In practice, the plasmid usually carries the *lac*I gene and the first 146 codons of the *lac*Z gene, because in the early days of genetic engineering this was a convenient fragment of DNA to manipulate.

Since wild-type β-galactosidase has 1021 amino acids, it is encoded by a gene 3.1 kb in length. While a gene of this length is easily manipulated *in vitro*, there are practical disadvantages to using the whole gene. As will be seen later, it is preferable to keep cloning vectors and their inserts as small as possible. The phenomenon of α-complementation allows genetic engineers to take advantage of the *lac* system without having to have the entire Z gene on the vector.

DNA ligase, is fundamental to many gene-manipulation procedures, such as that shown in Fig. 3.7.

The optimum temperature for ligation of nicked DNA is 37°C, but at this temperature the hydrogen-bonded join between the sticky ends is unstable. *Eco*RI-generated termini associate through only four AT base pairs and these are not sufficient to resist thermal disruption at such a high temperature. The optimum

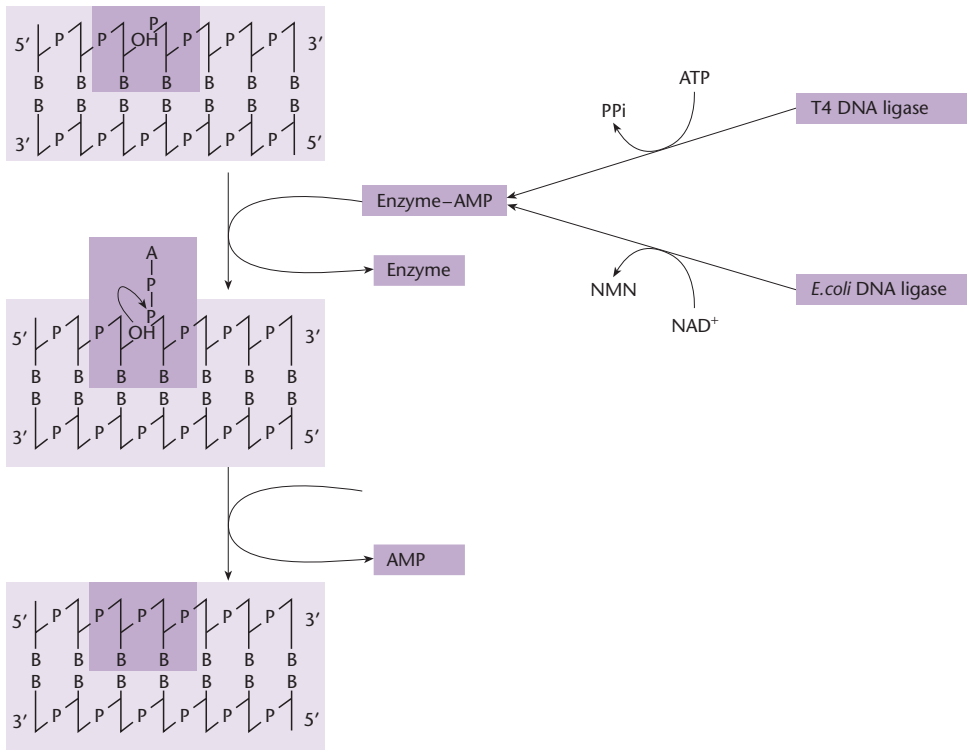


Fig. 3.6 Action of DNA ligase. An enzyme–AMP complex binds to a nick bearing 3′ OH and 5′ P groups. The AMP reacts with the phosphate group. Attack by the 3′ OH group on this moiety generates a new phosphodiester bond, which seals the nick.

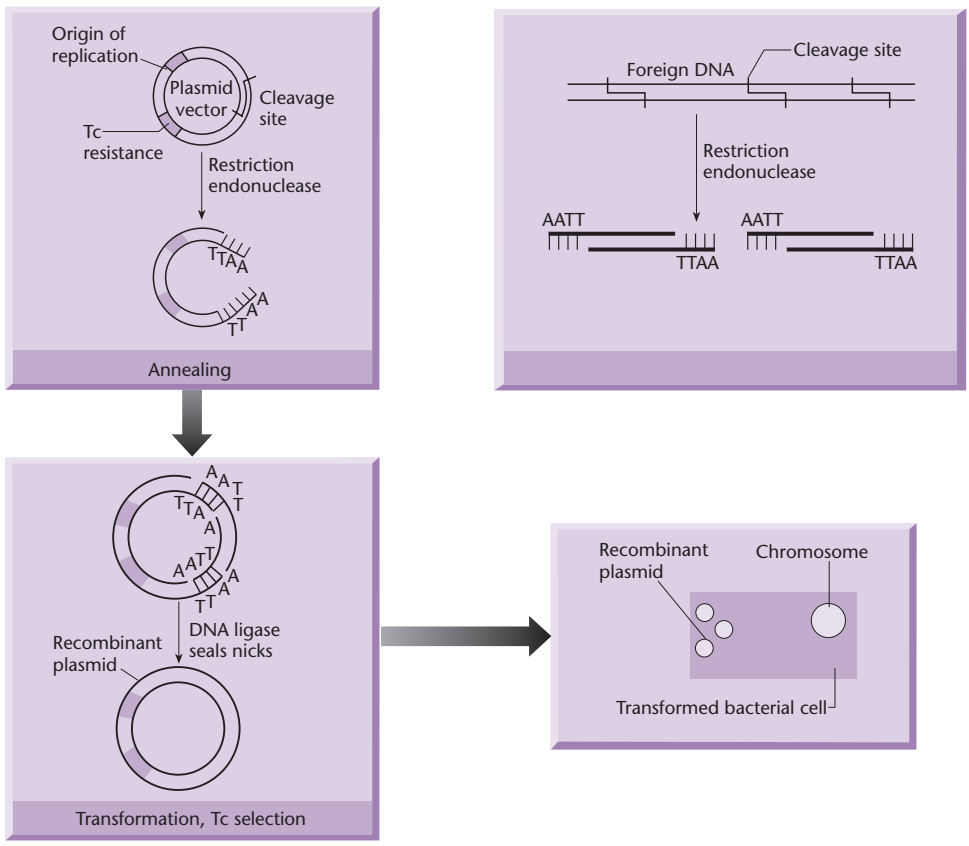


Fig. 3.7 Use of DNA ligase to create a covalent DNA recombinant joined through association of termini generated by *Eco*RI.

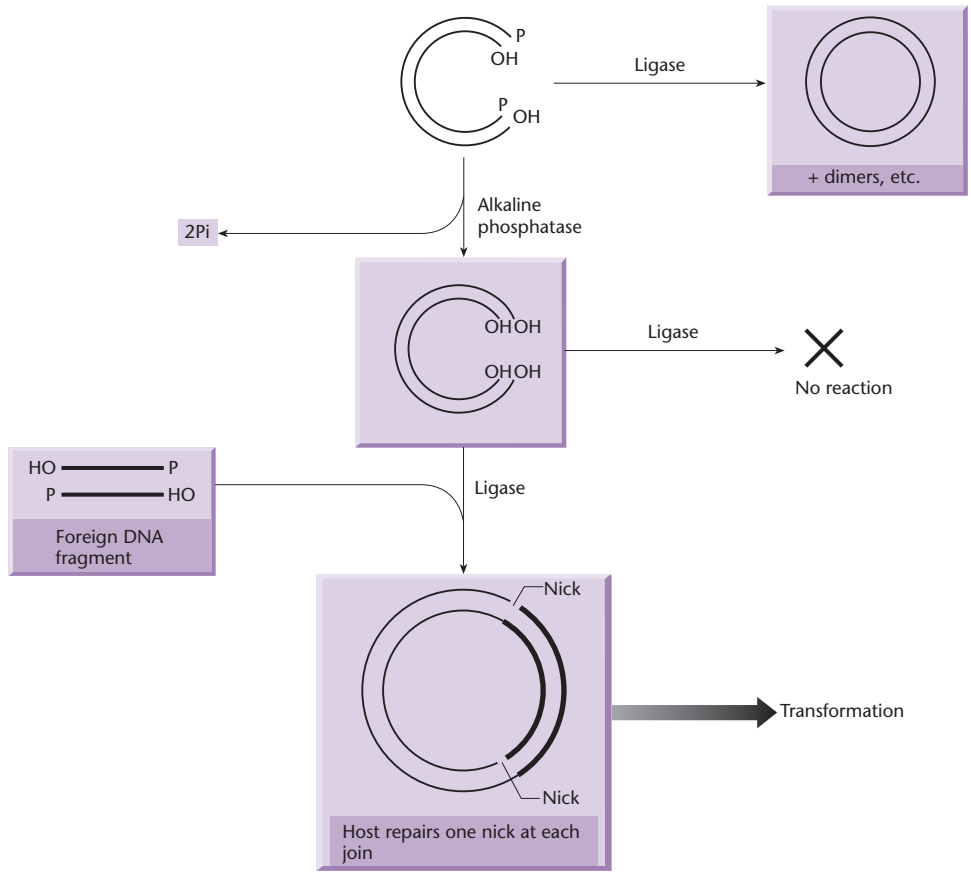


Fig. 3.8 Application of alkaline phosphatase treatment to prevent recircularization of vector plasmid without insertion of foreign DNA.

temperature for ligating the cohesive termini is therefore a compromise between the rate of enzyme action and association of the termini, and has been found experimentally to be in the range 4–15°C (Dugaiczyk *et al.* 1975, Ferretti & Sgaramella 1981).

The ligation reaction can be performed so as to favor the formation of recombinants. First, the population of recombinants can be increased by performing the reaction at a high DNA concentration; in dilute solutions *circularization* of linear fragments is relatively favored because of the reduced frequency of intermolecular reactions. Secondly, by treating linearized plasmid vector DNA with alkaline phosphatase to remove 5′-terminal phosphate groups, both recircularization and plasmid dimer formation are prevented (Fig. 3.8). In this case, circularization of the vector can occur only by insertion of non-phosphatase-treated foreign DNA which provides one 5′-terminal phosphate at each join. One nick at each join remains unligated, but, after transformation of host bacteria, cellular repair mechanisms reconstitute the intact duplex.

Joining DNA fragments with cohesive ends by DNA ligase is a relatively efficient process which has been used extensively to create artificial recombinants. A modification of this procedure depends upon the ability of T4 DNA ligase to join blunt-ended DNA molecules (Sgaramella 1972). The *E. coli* DNA ligase will not catalyze blunt ligation except under special reaction conditions of macromolecular crowding (Zimmerman & Pheiffer 1983). Blunt ligation is most usefully applied to joining blunt-ended fragments via *linker* molecules; in an early example of this, Scheller *et al.* (1977) synthesized self-complementary decameric oligonucleotides, which contain sites for one or more restriction endonucleases. One such molecule is shown in Fig. 3.9. The molecule can be ligated to both ends of the foreign DNA to be cloned, and then treated with restriction endonuclease to produce a sticky-ended fragment, which can be incorporated into a vector molecule that has been cut with the same restriction endonuclease. Insertion by means of the linker creates restriction-enzyme target sites at each end of the foreign DNA and so enables the foreign DNA to be excised and recovered after cloning and amplification in the host bacterium.

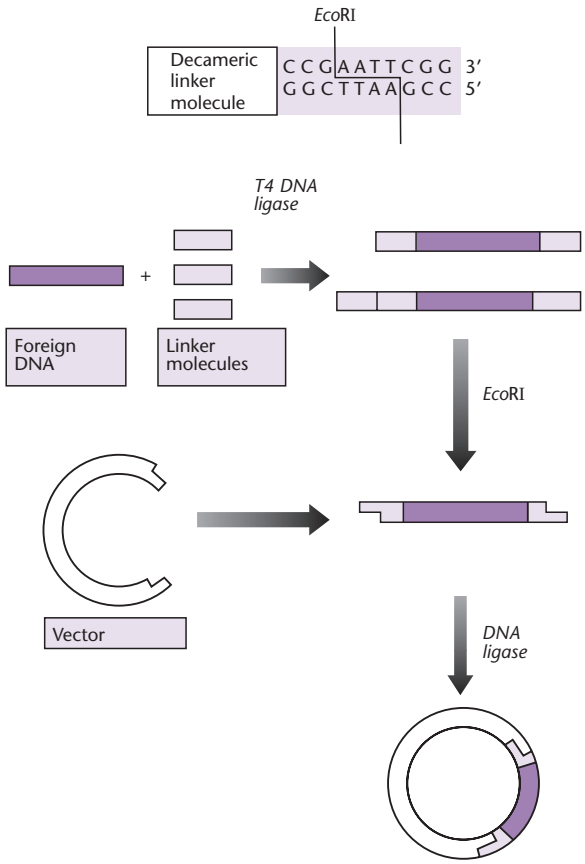


Fig. 3.9 A decameric linker molecule containing an *Eco*RI target site is joined by T4 DNA ligase to both ends of flush-ended foreign DNA. Cohesive ends are then generated by *Eco*RI. This DNA can then be incorporated into a vector that has been treated with the same restriction endonuclease.

Adaptors and linkers are short double-stranded DNA molecules that permit different cleavage sites to be interconnected

It may be the case that the restriction enzyme used to generate the cohesive ends in the linker will also cut the foreign DNA at internal sites. In this situation, the foreign DNA will be cloned as two or more sub-fragments. One solution to this problem is to choose another restriction enzyme, but there may not be a suitable choice if the foreign DNA is large and has sites for several restriction enzymes. Another solution is to methylate internal restriction sites with the appropriate modification methylase. An example of this is described in Fig. 6.2 (p. 99). Alternatively, a general solution to the problem is provided by chemically synthesized adaptor molecules which have a *preformed* cohesive end (Wu *et al.* 1978). Consider a blunt-ended foreign DNA containing an internal *Bam*HI site (Fig. 3.10), which is to be cloned in a *Bam*HI-cut vector. The *Bam*HI adaptor molecule has one blunt end bearing a 5′ phosphate group and a *Bam*HI cohesive end which is not phosphorylated. The adaptor can be ligated to the foreign DNA ends. The foreign DNA plus added adaptors is then phosphorylated at the 5′ termini and ligated into the *Bam*HI site of the vector. If the foreign DNA were to be recovered from the recombinant with *Bam*HI, it would be obtained in two fragments. However, the adaptor is designed to contain two other restriction sites (*Sma*I, *Hpa*II), which may enable the foreign DNA to be recovered intact.

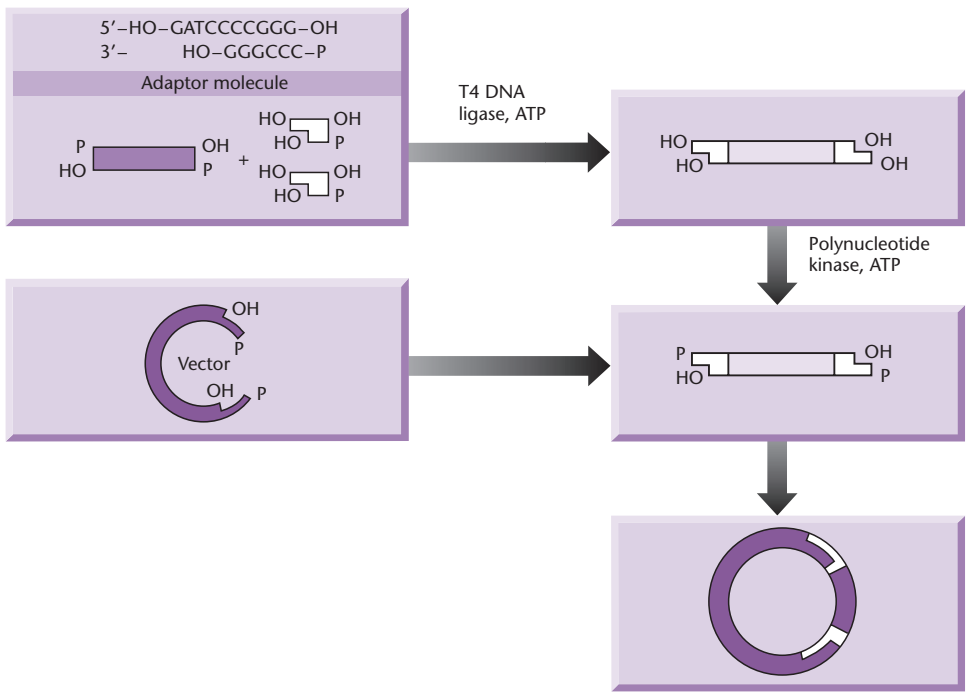


Fig. 3.10 Use of a *Bam*HI adaptor molecule. A synthetic adaptor molecule is ligated to the foreign DNA. The adaptor is used in the 5′-hydroxyl form to prevent self-polymerization. The foreign DNA plus ligated adaptors is phosphorylated at the 5′-termini and ligated into the vector previously cut with *Bam*HI.

Note that the only difference between an adaptor and a linker is that the former has cohesive ends and the latter has blunt ends. A wide range of adaptors is available commercially.

Homopolymer tailing is a general method for joining DNA molecules that has special uses

A general method for joining DNA molecules makes use of the annealing of complementary homopolymer sequences. Thus, by adding oligo(dA) sequences to the 3′ ends of one population of DNA molecules and oligo(dT) blocks to the 3′ ends of another population, the two types of molecule can anneal to form mixed dimeric circles (Fig. 3.11).

An enzyme purified from calf thymus, terminal deoxynucleotidyltransferase, provides the means by which the homopolymeric extensions can be synthesized, for if presented with a single deoxynucleotide triphosphate it will repeatedly add nucleotides to the 3′ OH termini of a population of DNA molecules (Chang & Bollum 1971). DNA with exposed 3′ OH groups, such as arise from pretreatment with phage λ exonuclease or restriction with an enzyme such as *Pst*I, is a very good substrate for the transferase. However, conditions have been found in which the enzyme will extend even the shielded 3′ OH of 5′ cohesive termini generated by *Eco*RI (Roychoudhury *et al.* 1976, Humphries *et al.* 1978).

The terminal transferase reactions have been characterized in detail with regard to their use in gene manipulation (Deng & Wu 1981, Michelson & Orkin 1982). Typically, 10–40 homopolymeric residues are added to each end.

One of the earliest examples of the construction of recombinant molecules, the insertion of a piece of λ DNA into SV40 viral DNA, made use of homopolymer tailing (Jackson *et al.* 1972). In their experiments, the single-stranded gaps which remained in the two strands at each join were repaired *in vitro* with DNA polymerase and DNA ligase so as to produce covalently closed circular molecules. The recombinants were then transfected into susceptible mammalian cells (see Chapter 12). Subsequently, the homopolymer method, using either dA.dT or dG.dC homopolymers was used extensively to construct recombinant plasmids for cloning in *E. coli*. In recent years, homopolymer tailing has been largely replaced as a result of the availability of a much wider range of restriction endonucleases and other DNA-modifying enzymes. However, it is still important for cDNA cloning (see p. 102 *et seq.*).

Special methods are often required if DNA produced by PCR amplification is to be cloned

Many of the strategies for cloning DNA fragments do not work well with PCR products. The reason for this is that the polymerases used in the PCR have a terminal transferase activity. For example, the *Taq*

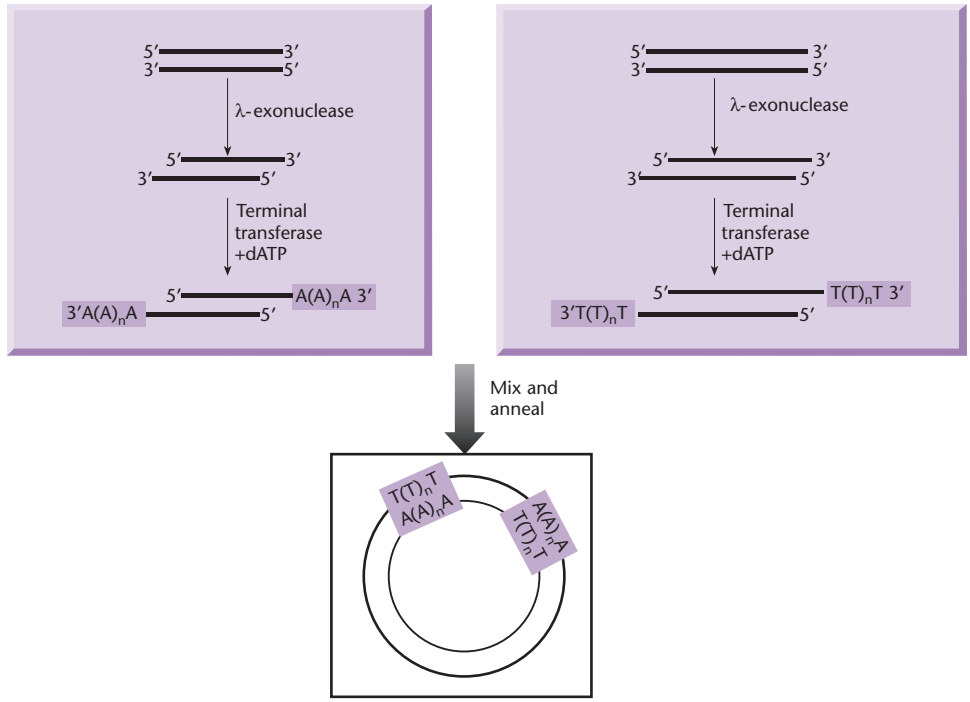


Fig. 3.11 Use of calf-thymus terminal deoxynucleotidyltransferase to add complementary homopolymer tails to two DNA molecules.

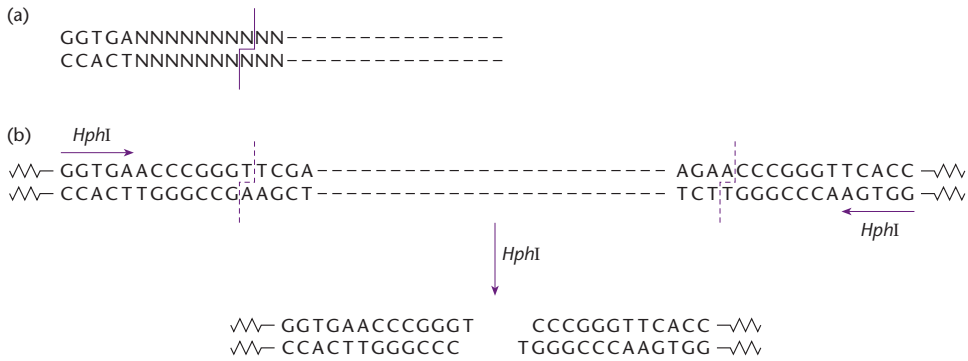


Fig. 3.12 Cleavage of a vector DNA molecule to generate single thymidylate overhangs. (a) The recognition sequence and cleavage point for the restriction endonuclease *Hph*I. (b) Sequences in the vector DNA which result in desired overhangs after cleavage with *Hph*I.

polymerase adds a single 3′ A overhang to each end of the PCR product. Thus PCR products cannot be blunt-end-ligated unless the ends are first *polished* (blunted). A DNA polymerase like Klenow can be used to fill in the ends. Alternatively, *Pfu* DNA polymerase can be used to remove extended bases with its 3′ to 5′ exonuclease activity. However, even when the PCR fragments are polished, blunt-end-ligating them into a vector still may be very inefficient. One solution to this problem is to use T/A cloning (Mead *et al.* 1991). In this method, the PCR fragment is ligated to a vector DNA molecule with a single 3′ deoxythymidylate extension (Fig. 3.12).

A PCR primer may be designed which, in addition to the sequence required for hybridization with the input DNA, includes an extra sequence at its 5′ end. The extra sequence does not participate in the first hybridization step – only the 3′ portion of the primer hybridizes – but it subsequently becomes incorporated into the amplified DNA fragment (Fig. 3.13). Because the extra sequence can be chosen at the will of the experimenter, great flexibility is available here.

A common application of this principle is the incorporation of restriction sites at each end of the amplified product. Figure 3.13 illustrates the addition of a *Hin*dIII site and an *Eco*RI site to the ends of an amplified DNA fragment. In order to ensure that the restriction sites are good substrates for the restriction endonucleases, four nucleotides are placed between the hexanucleotide restriction sites and the extreme ends of the DNA. The incorporation of these restriction sites provides one method for cloning amplified DNA fragments (see below).

DNA molecules can be joined without DNA ligase

In all the cutting and joining reactions described above, two separate protein components were required: a site-specific endonuclease and a DNA ligase. Shuman (1994) has described a novel approach to the synthesis of recombinant molecules in which a single enzyme, vaccinia DNA topoisomerase, both cleaves and rejoins DNA molecules. Placement of the CCCTT cleavage motif for vaccinia topoisomerase near the end of a duplex DNA permits efficient generation of a stable, highly recombinogenic protein DNA adduct that can only religate to acceptor DNAs that contain complementary single-strand extensions. Linear DNAs containing CCCTT cleavage sites at both ends can be activated by topoisomerase and inserted into a plasmid vector.

Heyman *et al.* (1999) have used the properties of vaccinia topoisomerase to develop a ligase-free technology for the covalent joining of DNA fragments to plasmid vectors. Whereas joining molecules with DNA ligase requires an overnight incubation, topoisomerase-mediated ligation occurs in 5 min. The method is particularly suited to cloning PCR fragments. A linearized vector with single 3′ T extensions is activated with the topoisomerase. On addition of the PCR product with 3′ A overhangs, ligation is very rapid. In addition, the high substrate specificity of the enzyme means that there is a low rate of formation of vectors without inserts.

Amplified DNA can be cloned using *in vitro* recombination

Conservative site-specific recombinases (see Box 3.3) are topoisomerases that catalyze rearrangements of DNA at specific sequences that are considerably longer than the cleavage sequence favored by vaccinia topoisomerase. These sites can be incorporated into PCR-amplified fragments by including them at the 5′ end of both of the amplification primers in exactly the same way as shown in Fig. 3.13 for restriction sites. All that is required to clone such

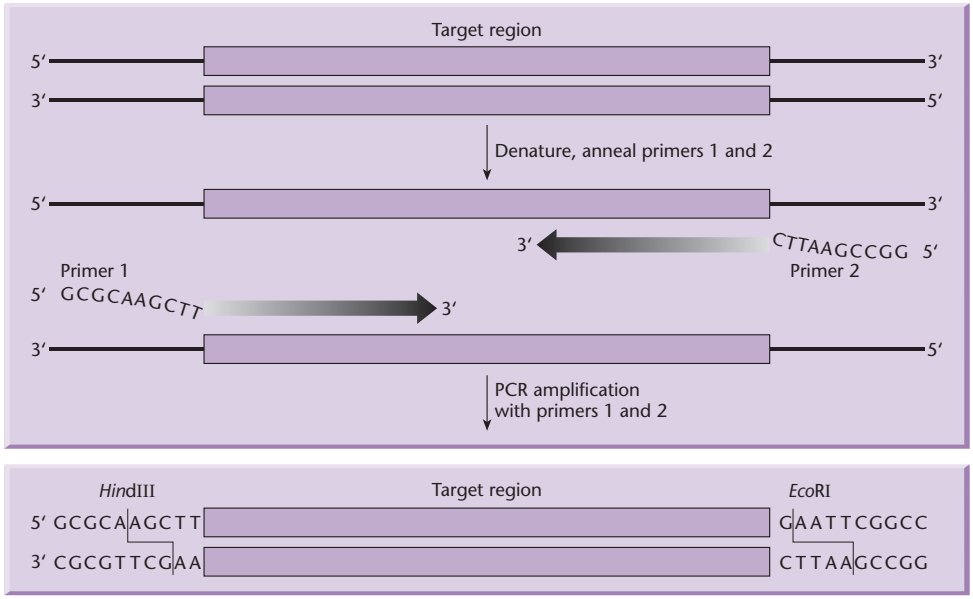


Fig. 3.13 Incorporation of extra sequence at the 5′ end of a primer. Two primers have sequences designed to hybridize at the ends of the target region. Primer 1 has an extra sequence near its 5′ end which forms a *Hin*dIII site (AAGCTT), and primer 2 has an extra sequence near its 5′ end which forms an *Eco*RI (GAATTC) site. Each primer has an additional 5′-terminal sequence of four nucleotides so that the hexanucleotide restriction sites are placed within the extreme ends of the amplified DNA, and so present good substrates for endonuclease cleavage.

Box 3.3 Site-specific recombinases

Site-specific recombinases catalyze a reciprocal double-stranded DNA exchange between two DNA segments provided both segments carry very specific sequences. The target DNA sequences can be on the same or different DNA molecules. If they are on different molecules, and one of the molecules is circular, then an insertion event will occur. The best example of this is the chromosomal insertion of the genome of bacteriophage λ during the process of lysogeny (Fig. B3.1). However, if the two DNA molecules are linear then segments of DNA will be exchanged.

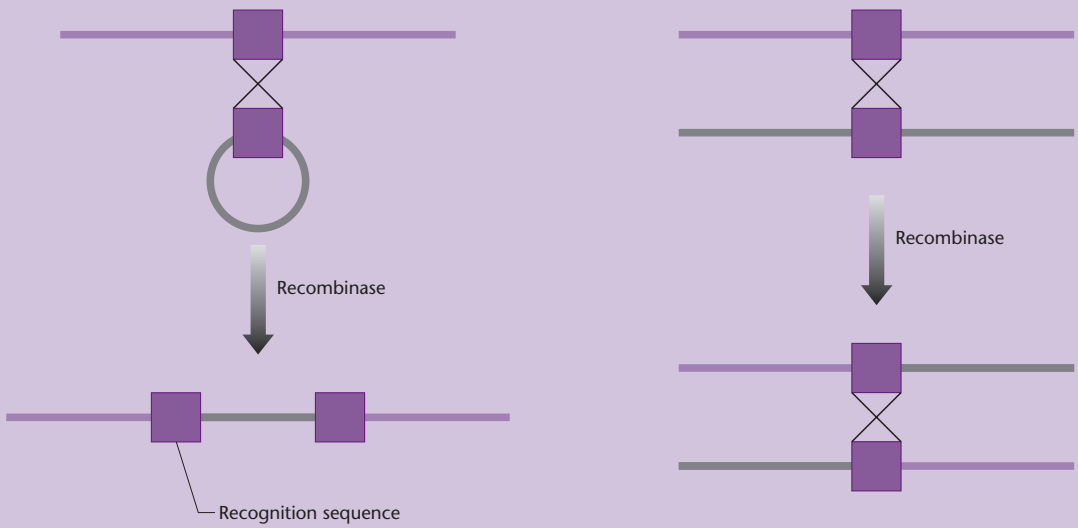


Fig. B3.1 The outcomes of recombinase action when the target sites are on different DNA molecules.

continued

Box 3.3 *continued*

If the recognition sites for the recombinase are on the same DNA molecule then the outcome depends on their alignment. If the two sites are aligned in the same direction then the intervening DNA is deleted. If the two sites are in opposite orientation then inversion of the intervening DNA occurs (Fig. B3.2).

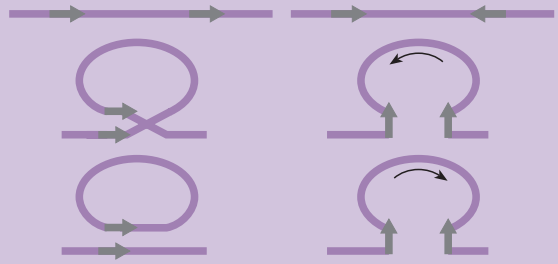


Fig. B3.2 The outcomes of recombinase action when the target sequences are on the same DNA molecule.

The site-specific recombinases catalyze DNA exchange by two different mechanisms: the Int–Flp and resolvase–invertase mechanisms. Only the Int–Flp recombinases are used in gene manipulation and the best-known ones are phage λ Int, Flp from the yeast *Saccharomyces cerevisiae* and the Cre protein from bacteriophage P1. With some recombinases, like Cre and Flp, the sites where recombination takes place (Fig. B3.3) are identical and the enzyme can function without any accessory proteins.

With others, such as Int, the sites of recombination exhibit homology but are not identical and additional proteins are required for activity (Fig. B3.4). All of them share a conserved carboxyterminal region that includes the active site tyrosine residue that forms a covalent protein–DNA intermediate during the recombination step. Since recombination does not involve the gain or loss of nucleotides it is said to be conservative.

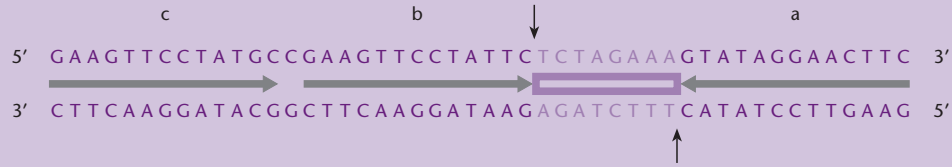


Fig. B3.3 The recognition site (*FRT*) for the Flp recombinases. The site consists of three 13 bp symmetry elements (shown by arrows), one of which (a) is in a different orientation to the other two (b and c). The "a" and "b" elements are separated by an 8-bp asymmetric sequence across which recombination takes place. The "c" element is dispensable. The Cre recognition site (*lox*P) is similar to *FRT* in consisting of two 13-bp inverted repeats separated by an asymmetric 8-bp region (Fig. 15.7).

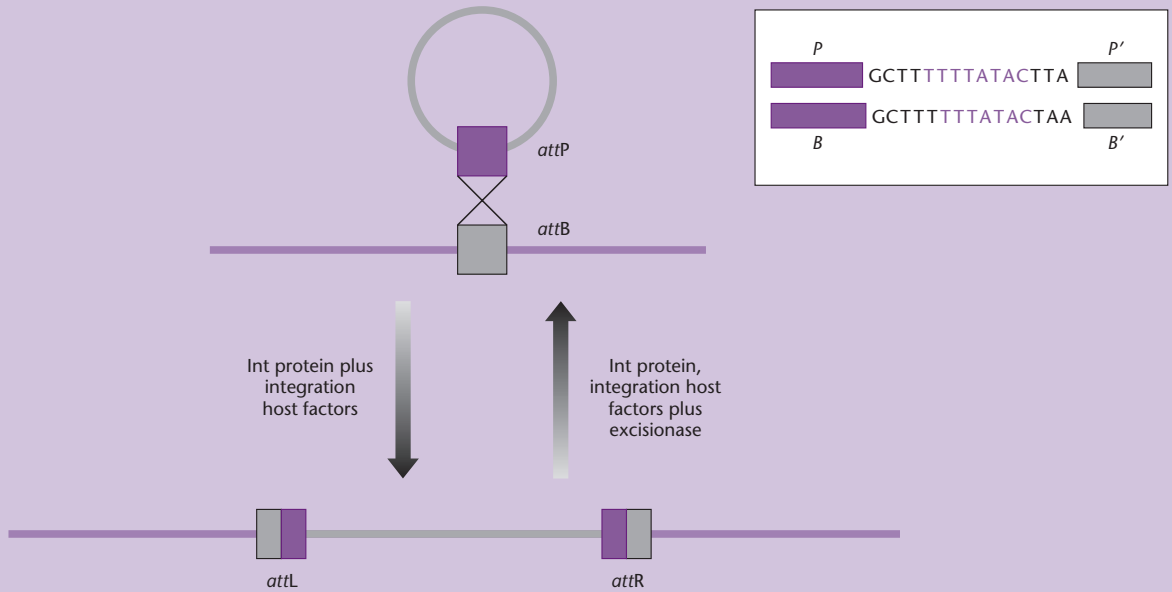


Fig. B3.4 The integration of DNA into the chromosome of *E. coli*. The *att*P and *att*B sites are mostly dissimilar and have only the central 15 bases in common (inset). Recombination occurs between the bases shown in red.

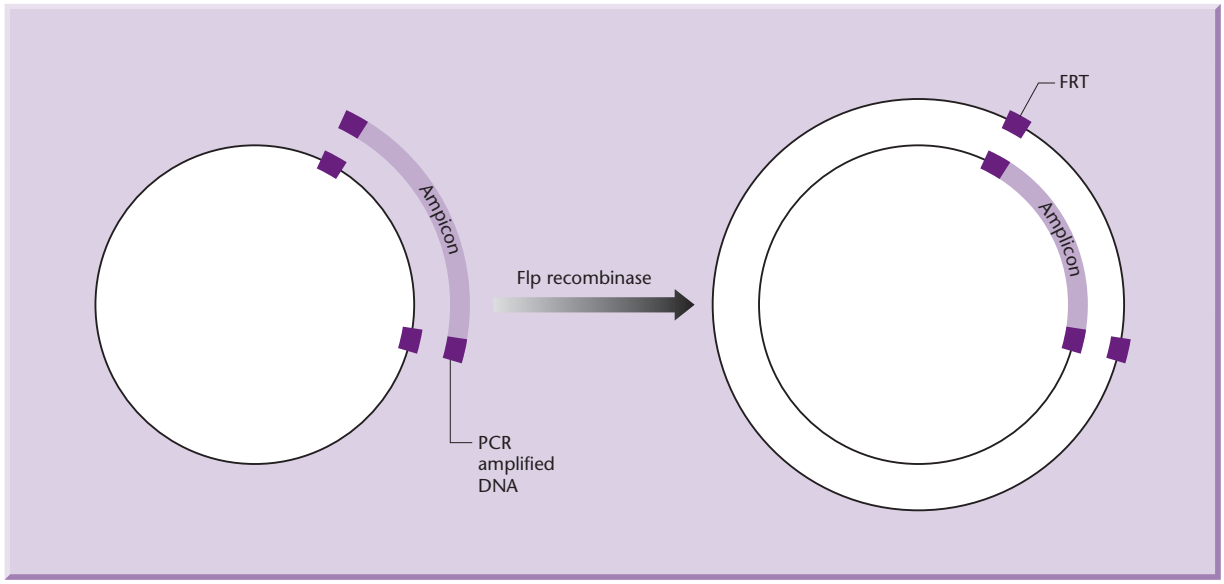


Fig. 3.14 Cloning PCR-amplified DNA using a site-specific recombinase. The DNA to be cloned is amplified using primers containing the recognition site for a site-specific recombinase (in this case, *FRT*, the site for the Flp recombinase). The amplified DNA is mixed with a vector containing two *FRT* sites and the Flp recombinase. The recombinase mediates the replacement of the vector section between the two *FRT* sites with the amplified DNA.

amplified DNA is to mix it with a vector containing two copies of the homologous recognition site (Fig. 3.14). Two variations of this method have been developed. In the original version, developed by Hartley *et al.* (2000), the λ integrase (Int) is used to catalyze the recombination event. This results in a recombinant molecule in which the insert is flanked by different sites than were present on the original amplified DNA. In the variation described by Sadowski (2003), the Flp recombinase is used and this generates recombinants in which the flanking sites remain unchanged.

There are two major applications of cloning with recombinases: recloning and recombineering. Once a PCR amplicon has been cloned it may be necessary to reclone it in another vector and the presence of recombinase sites makes this recloning easy (Fig. 3.15). A series of specialist vectors (the Gateway vectors) have been designed for this purpose and these are described on p. 94. In certain experiments it is necessary to clone very large fragments of DNA (>50 kb). It is difficult to manipulate *in vitro* vectors containing such large inserts and usually it is preferable to do the manipulations *in vivo*. This is known as recombineering and it is facilitated by the use of the Cre and Flp recombinases in conjunction with specialist vectors known as BACs and PACs (see p. 76). Other applications of recombinases are described in the review of Schweizer (2003).

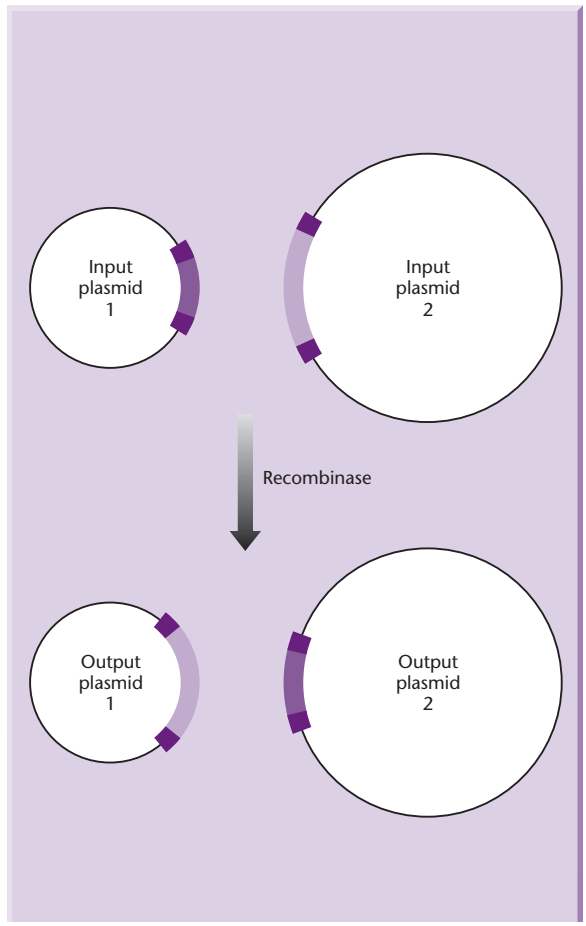


Fig. 3.15 Using recombinases to move cloned DNA from one vector to another.

Suggested reading

Loenen W.A. (2003) Tracking *Eco*KI and DNA fifty years on: a golden era full of surprises. *Nucleic Acids Research* **31**, 7059–69.

An excellent review of host-controlled restriction and modification which provides an historical perspective on the development of restriction enzymes.

Buryanov Y. & Shevchuk T. (2005) The use of prokaryotic DNA methylation transferases as experimental and analytical tools in modern biology. *Analytical Biochemistry* **338**, 1–11.

Pingoud A., Fuxreiter M., Pingoud V. & Wende W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell and Molecular Life Sciences* **62**, 685–707.

Two reviews that provide additional detail to the material covered in this chapter.

Copeland N.G., Jenkins N.A. & Court D.L. (2001) Recombineering: a powerful new tool for mouse functional genomics. *Nature Reviews Genetics* **2**, 769–79.

Muyrers J.P.P., Zhang Y. & Stewart F. (2001) Techniques: recombinogenic engineering – new options for cloning and manipulating DNA. *Trends in Biochemical Sciences* **26**, 325–31.

Two excellent reviews on the topic of recombineering.

Useful website

http://www.neb.com/nebecomm/tech_reference/

New England Biolabs are the premier supplier of restriction enzymes and other enzymes involved in *in vitro* gene manipulation. This website details everything that you need to know before using any of these enzymes.