From Genes to Genomes: Concepts and Applications of DNA Technology.
Jeremy W Dale and Malcom von Schantz
Copyright © 2002 John Wiley & Sons, Ltd.
ISBNs: 0-471-49782-7 (HB); 0-471-49783-5 (PB)

6 Vectors

In the previous chapter, we looked at the ways in which a potential insert can be ligated with a vector molecule, to enable that insert DNA to be replicated after insertion into a bacterial cell. We now need to consider further the nature of the vectors that can be used. Once again, we will initially be concentrating on the vectors that are used for cloning in bacteria (principally *E. coli*), and subsequently looking at the vectors needed for cloning in eukaryotes.

6.1 Plasmid Vectors

6.1.1 Properties of plasmid vectors

Plasmids are by far the most widely used, versatile, and easily manipulated vectors. They are the work-horses of the molecular biology laboratory. They are naturally occurring extrachromosomal DNA molecules, usually circular, double-stranded and supercoiled. Plasmids occur widely in nature, and are found in most bacterial species. They vary considerably in size, from a few thousand base pairs up to several hundred kilobases, although plasmids used as gene cloning vectors are usually small (typically 2–5 kb). Most of the commonly used ones are based on (or are closely related to) a naturally occurring *E. coli* plasmid called ColE1. Later in this chapter we will look at other types of vectors such as bacteriophages.

The most notorious property of plasmids lies in their ability to disseminate antibiotic resistance genes. They are responsible to a large extent for the spread of antibiotic resistance – although it should be noted that the plasmids used for gene cloning are nearly always unable to spread from one bacterium to another, and there are restrictions on experimental protocols to ensure that these experiments do not add new antibiotic resistance genes to clinically important pathogenic bacteria. Antibiotic resistance is not the limit of the ability of plasmids, nor the reason for their existence. Interest tends to focus on antibiotic resistance because of its importance in medical microbiology, and because of the ease with which resistance genes can be isolated and studied. However, many naturally occurring plasmids code for other properties, or even for none

at all, or at least none that we can discern. Plasmids exist because they can replicate within bacteria, and sometimes spread from one bacterium to another. That is all. They are a form of DNA parasite. Any advantage they confer on the host bacterium is a bonus that helps the plasmid to survive.

Plasmid replication

In genetic engineering, we make use of this ability of plasmids to be replicated, as it enables us to insert pieces of DNA which are then copied as part of the plasmid, and hence passed on to the progeny when the cell replicates. The most fundamental property of a plasmid, whether we are considering it as a cloning vector or as a natural phenomenon, is therefore the ability to replicate in the host bacterium. Most, or all, of the enzymes and other products needed for this replication are already present in the host cell; the amount of information that the plasmid has to supply may be only a few hundred base pairs. This region of the plasmid that is necessary for replication is generally referred to as the origin of replication, although literally the origin, or the site at which replication starts, is one specific base.

Plasmids that use the origin of replication from ColE1 or its relatives are multi-copy plasmids. Wild-type ColE1 is present at about 15 copies per cell, while most of the engineered vectors used today are present in numbers running into many hundreds of copies per cell. This is convenient in some ways as it makes it easier to purify large amounts of the plasmid, and if you want to express a cloned gene you also get a gene dosage effect. The presence of so many copies of the gene in the cell is reflected in higher levels of the product of that gene (see also Chapter 15). However, this can also be a disadvantage. Even without expression of the cloned gene, the large amount of plasmid DNA may make the cell grow more slowly. If the gene or its product is in any way harmful to the bacterium, it can sometimes be very difficult to isolate the required clone. For some specific purposes therefore it is desirable to use alternative vectors that exist at low copy number (or to use different vectors altogether that do not require continued viability of the cell, such as some types of bacteriophage vector – see below).

Some plasmids are able to replicate in a wide variety of bacterial species (broad host-range plasmids), but most of those that are used for gene cloning are rather more restricted in their host range. In one way this is useful: if there is any question about potential health hazards or environmental consequences associated with cloning a specific fragment of DNA, then using a narrow host-range plasmid makes it very unlikely that the gene will be transmitted to other organisms.

On the other hand, you may wish to carry out genetic manipulations in a bacterium other than *E. coli*, especially if your interest lies in studying the

behaviour of specific bacteria rather than simply using them to clone pieces of DNA. It will then usually be necessary to isolate or construct new vector plasmids, based on a replication origin that is functional in your chosen species. The host range of your new vector will probably also be limited, and it may well be unable to replicate in *E. coli*. This is a disadvantage, because you are likely to want to use *E. coli* as an intermediate host for the initial cloning and for studying the structure and behaviour of the gene that you have cloned. However, it is possible to insert two origins of replication into your plasmid, so that it will be replicated in *E. coli* using one origin, and in your chosen host using the alternative replication origin. Such a vector is known as a *shuttle plasmid*, because it can be transferred back and forth between the two species. We can also use shuttle vectors to transfer cloned genes between *E. coli* and a eukaryotic organism. We will be coming across various applications of shuttle vectors in subsequent chapters.

Therefore, the first essential characteristic of a plasmid cloning vector is the origin of replication, usually designated as *ori* in plasmid maps.

Cloning sites

The second characteristic that is necessary for a plasmid to be useful as a cloning vector is a *cloning site*. This is a unique restriction site, so that the enzyme concerned will cut the plasmid once only. If a circular molecule is broken at one position, it is converted into a linear molecule, and it is relatively simple to join the ends together to reform an intact circle. If an enzyme cuts more than once, the plasmid will be cut into two or more pieces, and joining them up again to make an intact plasmid will be much more complicated. A basic plasmid used as a cloning vector may contain only one or two such unique restriction sites, which must of course be located in a region of the plasmid that is not essential for replication or any other functions that we need.

With such a plasmid you are limited not only in your choice of restriction fragments that can be inserted and in the position of insertion, but also in the number of different fragments that can be inserted. This is because in most cases ligation of two restriction fragments, generated with the same enzyme, recreates the original restriction site. Thus when you insert say a *Bam*HI fragment into a site on the vector that has been cut with *Bam*HI, the resulting recombinant plasmid will have two *Bam*HI sites: one at each end of the inserted fragment (see Figure 6.1). This is the basis of a common test for the presence of such an insert. Digesting the supposed recombinant plasmid with, in this case, *Bam*HI will release a DNA fragment that should be the size of the insert you are trying to clone – see Chapter 8. However, since the recombinant plasmid now has two *Bam*HI sites, it would be difficult to clone further *Bam*HI fragments into it.

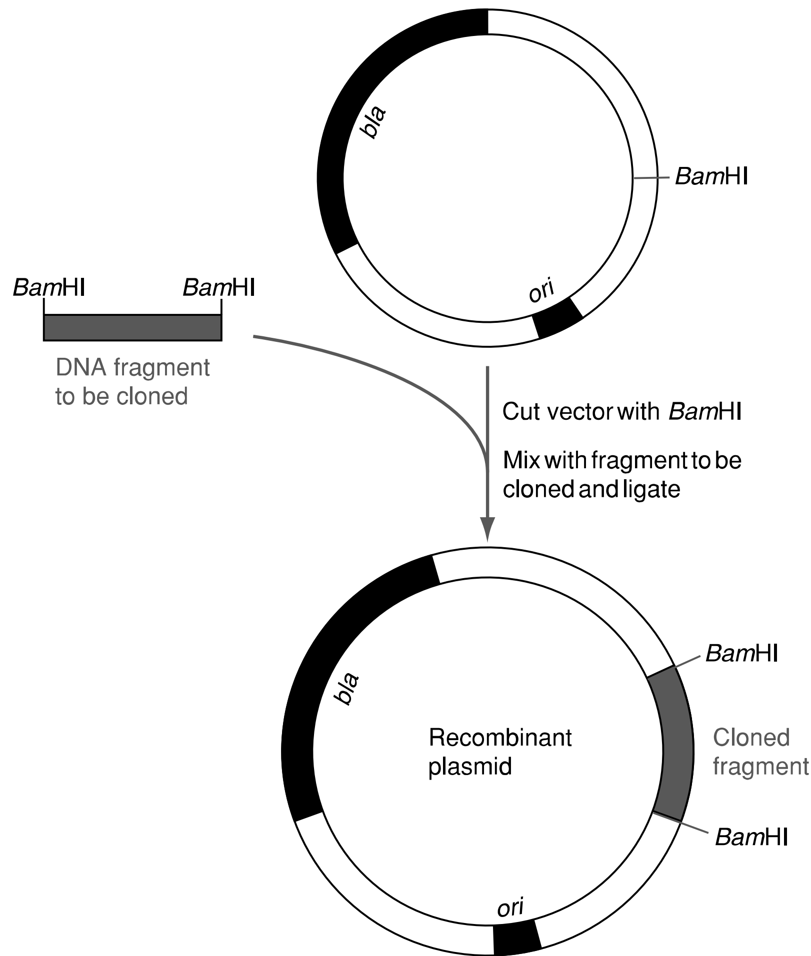
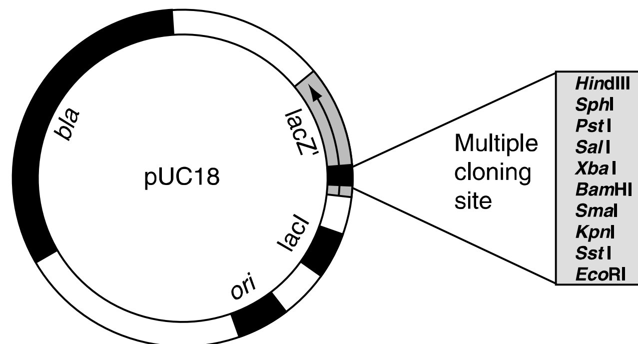


Figure 6.1 Cloning with a plasmid vector: *bla* = beta-lactamase (ampicillin resistance) selective marker; *ori* = origin of replication

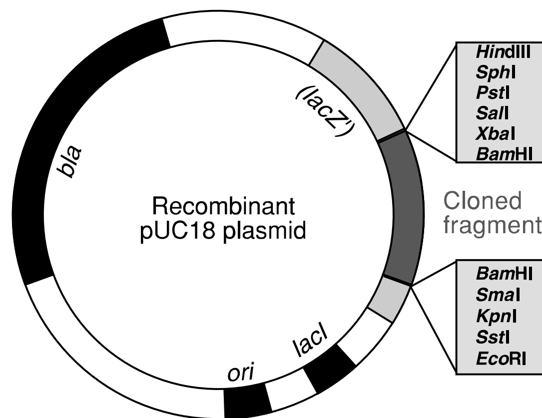
The problem here is that in many cases we *do* want to insert several fragments into the same plasmid. We may want to combine the expression signals of one gene with the coding region of another, or we may want to insert additional markers that can be used to identify the presence of the plasmid. Or we may, as described above, want to insert the replication origin from another plasmid so as to create a shuttle vector, and still leave sites available for further inserts. The best way around this problem is to create a *multiple cloning site* (MCS), i.e. a short DNA region that contains recognition sites for a number of different enzymes. This is done by synthesizing a short piece of DNA with the required restriction sites, and inserting that into the plasmid in the usual way. Figure 6.2 shows the structure of pUC18, one of a family of similar plasmids that are commonly used as cloning vectors, and you will see

that pUC18 contains such a multiple cloning site. Insertion of a fragment into the *Bam*HI site, as in Figure 6.3, will still leave a selection of other sites available for further inserts.



bla = beta-lactamase (ampicillin resistance); selective marker
ori = origin of replication
lacZ' = beta-galactosidase (partial gene)
lacI = repressor of *lac* promoter

Figure 6.2 Structure of the plasmid cloning vector pUC18



The *lacZ* gene has been disrupted by insertion of a DNA fragment, resulting in white colonies on X-gal plates

bla = beta-lactamase (ampicillin resistance); selective marker
ori = origin of replication
lacZ' = beta-galactosidase (partial gene)
lacI = repressor of *lac* promoter

Figure 6.3 Use of the plasmid cloning vector pUC18

Selectable markers

So we have a plasmid with a replication origin and one or more restriction sites. One further feature is essential for a functionally useful vector, and that is a selectable marker. The need for this arises from the inefficiency both of ligation and of bacterial transformation. Even with the high efficiency systems that are now available for *E. coli*, the best yield available, using native plasmid DNA, implies that only about 1 per cent of the bacterial cells actually take up the DNA. In practice the yields are likely to be lower than this – and if you are using a host other than *E. coli*, many orders of magnitude lower. Therefore in order to be able to recover the transformed clones, it is necessary to be able to prevent the non-transformed cells (i.e. those cells that have not taken up the plasmid) from growing. The presence of an antibiotic resistance gene on the plasmid vector means that you can simply plate out the transformation mix on an agar plate containing the relevant antibiotic, and only the transformants will be able to grow. In Figure 6.2, you will see that pUC18 carries a β-lactamase gene (*bla*), coding for an enzyme that hydrolyses β-lactam (penicillin-like) antibiotics such as ampicillin (and hence often referred to as Amp^R, for ampicillin resistance).

Insertional inactivation

In Figure 6.3 you will see a further feature of pUC18. The multiple cloning site is located near to the 5′ end of a β-galactosidase gene (*lacZ*). The synthetic oligonucleotide that creates the multiple cloning site was designed so that it did not affect the reading frame of the *lacZ* gene; it merely results in the production of β-galactosidase with some additional amino acids near to the amino terminus of the protein. This does not affect the function of the enzyme; it is still able to hydrolyse lactose. More accurately, we should say that pUC18 carries a part of the *lacZ* gene; we use *E. coli* strains that carry the remainder of the gene. The product of the host gene is unable to hydrolyse lactose by itself, and so the host strain without the plasmid is Lac^-, i.e. it does not ferment lactose. When pUC18 is inserted into the host, the plasmid-encoded polypeptide will associate with the host product to form a functional enzyme. We say that pUC18 is capable of *complementing* the host defect in *lacZ*.

We can easily detect the activity of the β-galactosidase by plating the organism onto agar containing the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (universally known as *X-gal*), together with the inducing agent isopropyl thiogalactoside (*IPTG*). The X-gal substrate is colourless but the action of β-galactosidase releases the dye moiety, resulting in a deep blue colour. Colonies carrying pUC18 are therefore blue

when grown on this medium. However, if we are successful in inserting a DNA fragment at the cloning site, the gene will (normally) be disrupted and the resulting *E. coli* colonies are referred to as 'white'. The advantage of this *insertional inactivation* is that we can tell not only that the cells have been transformed with the plasmid (since they are able to grow in the presence of ampicillin), but also that the plasmid is a recombinant, and not merely the original pUC18 self-ligated. An insertional inactivation marker such as this is not an essential feature of a cloning vector but it does provide a useful way of monitoring the success of the ligation strategy and overcoming some of the problems referred to in the previous chapter.

One word of caution: if the insert is relatively small, and if it happens to consist of a multiple of three bases, transcription and translation of the *lacZ* gene may still occur and the enzyme may still have enough activity (despite the addition of still more amino acids at the N terminus) to produce a detectable blue colour. Conversely, a white colony is not a guarantee of cloning success as the deletion of even a single base at the cloning site, or the insertion of undesirable junk fragments (in other than multiples of three bases), will put the *lacZ* gene in the wrong reading frame and thus inactivate it.

The advantage of plasmid vectors, compared with the other vectors described subsequently in this chapter, is that they are small and easy to manipulate; also they are conceptually simple and universal. You can make and use plasmid vectors for a wide range of organisms without a detailed knowledge of the molecular biology of the host or the vector. On the other hand, the basic plasmid vectors that we have been considering so far are limited in their cloning capacity, i.e. the size of the insert they can accommodate. Later in this chapter, we will look at other vectors that will accommodate larger inserts, but first we need to consider the ways in which we can introduce the recombinant plasmids into host bacterial cells.

6.1.2 Transformation

Bacterial transformation was discovered in 1928 with the demonstration by Fred Griffith that cultures of the pneumococcus (*Streptococcus pneumoniae*) that had lost virulence could have their pathogenicity restored by addition of an extract of a killed virulent strain. It was the identification, many years later (by Avery, MacLeod and McCarty), that the 'transforming principle' is DNA that resolved the question of the chemical nature of the genetic material.

This experiment rests on the natural ability of the pneumococcus to take up 'naked' DNA from its surroundings. This ability is known as *competence*. Competence develops naturally in some bacterial species, but although the range of species that exhibit natural competence is much wider than was thought for many years to be the case, it is still too limited in scope (or too

inefficient or too selective) to be of much use for genetic engineering. In particular, *E. coli* does not seem to exhibit natural competence. It was therefore necessary to develop alternative ways of introducing plasmid DNA into bacterial cells. Although these methods are radically different, they are still referred to as transformation, which is defined as the uptake of naked DNA, to distinguish it from other methods of horizontal gene transfer, namely *conjugation* (direct transfer by cell to cell contact) and *transduction* (which is mediated by bacteriophage infection).

The breakthrough came with the demonstration that competence in *E. coli* cells could be induced by washing them with ice-cold calcium chloride, followed by adding the plasmid DNA and subjecting the mixture to a brief, mild heat shock (e.g. 2 min at 42°C). It is then necessary to dilute the cells in growth medium and incubate for a while (30–60 min) to allow the bacteria to recover and to express the resistance marker introduced on the plasmid, before plating them onto a selective medium containing the appropriate antibiotic. Although this simple basic process represented a major, and essential, step forward, it was very inefficient with yields of perhaps 10^4 transformants per μg of pure, supercoiled plasmid DNA (and less with ligation mixtures or non-supercoiled DNA). Gradually, over the years, improvements have been made in the transformation process, both by modifying the preparation of competent cells (e.g. using salts other than calcium chloride) and also by selecting *E. coli* strains with mutations that make them easier to transform. With the best of these systems it is now possible to obtain transformation frequencies in excess of 10^9 transformants per μg of plasmid DNA. Where such high yields are required, it is cost-effective to purchase pre-prepared competent cells of a strain with high transformation efficiency. Note that although transformation frequencies are generally quoted in these terms (number of transformants per μg of DNA), there is not a linear relationship between the number of transformants and the amount of DNA used. Transformation works best with low levels of DNA, and the efficiency with which bacterial cells take up DNA falls off as the concentration of DNA is increased. If you increase the amount of DNA too much, you might even decrease not just the efficiency but the actual numbers of transformants.

Reference back to the discussion of ligation in Chapter 5 will disclose a quandary here. Ligation works best with high concentrations of DNA. The following step, transformation, works best with small amounts of DNA. The resolution is clear, although unpalatable: use only a small proportion of your ligation mix in the transformation step. If you really need very large numbers of transformants, scaling up the transformation step does not work very well – it is usually much better to carry out several separate small-scale transformations.

Transformation based on induced competence and heat shock can be used for bacterial species other than *E. coli*, but you immediately lose all the

advantages that have been gained by optimization of transformation conditions for selected strains of *E. coli*. At best, therefore, transformation is likely to be very inefficient – and in most cases simply using an *E. coli* procedure will not work at all. Therefore laboratories that are interested in manipulating other bacterial species have had to develop alternative methods of transformation.

Electroporation is the most versatile transformation procedure. Bacterial cells, washed with water to remove electrolytes from the growth medium, are mixed with DNA and subjected to a brief pulse of high-voltage electricity. This appears to induce temporary holes in the cell envelope through which the DNA can enter. The cells are then diluted into a recovery medium before plating on a selective medium in the same way as above. Although it is comparatively easy to obtain *some* transformants with a wide range of bacteria (or other cells), there are many parameters that need to be adjusted to obtain optimum performance, including the conditions under which the cells are grown, the temperature of the suspension, and the duration and voltage of the electric pulse.

Since the added DNA seems to simply diffuse through the holes created (briefly) by the electric pulse, the effect is not specific for DNA; other substances, notably RNA or proteins, can also be introduced into bacterial cells by electroporation. Nor is it directionally specific. Material within the cell can diffuse out as well, and the procedure has been used for isolating plasmid DNA from bacterial cells. It follows from this that, since the plasmid that comes out of one cell can enter another one, electroporation can be used to transfer plasmids from one strain to another, simply by applying it to a mixture of the two strains.

Other methods that are used more commonly with animal and plant cells, including microinjection, biolistics, and protoplast transformation are considered in Chapter 17.

6.2 Vectors Based on the Lambda Bacteriophage

6.2.1 *Lambda biology*

Plasmid vectors are at their best when cloning relatively small fragments of DNA. Although there is probably no fixed limit to the size of a DNA fragment that can be inserted into a plasmid, the recombinant plasmid may become rather less stable with larger DNA inserts, the efficiency of transformation is reduced, and the plasmid will give a much smaller yield when grown and purified in *E coli*. Vectors based on bacteriophage lambda allow efficient cloning of larger fragments, which is important in constructing gene libraries. The larger the inserts, the fewer clones you have to screen to find the one you

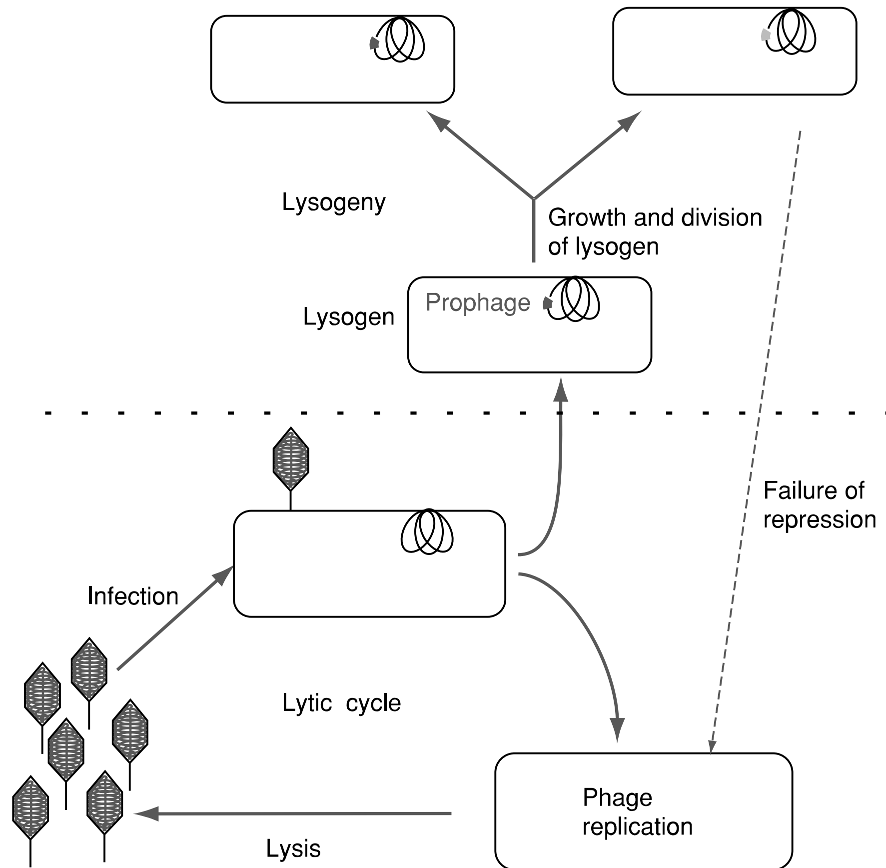


Figure 6.4 Lytic cycle and lysogeny

want (see Chapters 7 and 8). Lambda vectors also have advantages in gene library construction, as it is much easier to screen large libraries when using bacteriophage vectors; the results with bacteriophage plaques are much cleaner than those obtained with bacterial colonies.

In order to understand the nature and use of lambda cloning vectors, some knowledge of the basic biology of bacteriophage lambda is necessary. While we hope you are familiar with this, a recap of the salient features (summarized in Figure 6.4) will be useful.

Lysogeny

Lambda is a temperate bacteriophage, i.e. on infection of *E. coli* it may enter a productive lytic cycle, resulting in lysis of the cell and liberation of a number of phage particles, or it may enter a more or less stable relationship with the host known as *lysogeny*. In the lysogenic state, expression of almost all of the phage genes is switched off by the action of a phage-encoded repressor protein, the product of the *cI* gene. The expression of this gene during the establishment of

lysogeny requires two other genes, *cII* and *cIII*. The proportion of infected cells going down each route is influenced by environmental conditions, as well as by the genetic composition of the phage and the host. Some phage mutants will only produce lytic infection, and these give rise to clear plaques, while the wild-type phage produces turbid plaques due to the presence of lysogens which are resistant to further attack by lambda phage (known as *superinfection immunity*). On the other hand, some bacterial host strains carrying a mutation known as *hfl* (high frequency of lysogenization) produce a much higher proportion of lysogens when infected with wild-type lambda – which can be useful if we want a more stably altered host strain, for example if we are studying the expression of genes carried by the phage. Generally, when using lambda vectors we are more interested in the recombinant phage carrying the cloned genes, and the lytic cycle is the more relevant one in such cases.

Although the lysogenic state is relatively stable, that stability is not absolute. A culture of a bacterial lysogen will normally contain phage particles in the supernatant, due to a low level of spontaneous failure of the repression mechanism. This rate of breakdown of repression can be increased by treating the culture with agents that damage the DNA, such as UV irradiation; the DNA damage induces the production of repair enzymes which amongst other things destroy the *cI* repressor protein, allowing initiation of the lytic cycle. Some widely used lambda vectors carry a mutation in the *cI* gene which makes the protein more temperature-sensitive (*cI857* mutation). A lysogen carrying such a mutant phage can be grown as a lysogen at a reduced temperature and the lytic cycle will be induced by raising the temperature, due to inactivation of the repressor protein.

In the lysogenic state, lambda is normally integrated into the bacterial chromosome, by site-specific recombination at a specific position, and is therefore replicated as part of the bacterial DNA. Induction of the lysogen requires excision from the chromosome. However, this integration, although common amongst temperate phages, is not an essential feature of lysogeny. Lambda can continue to replicate in an extrachromosomal, plasmid-like state; with some bacteriophages (including P1, which we will encounter later in this chapter) this is the normal mode of replication in lysogeny.

Particles of wild-type bacteriophage lambda have a double-stranded linear DNA genome of 48 514 base pairs, in which the 12 bases at each end are unpaired but complementary (see Figure 6.5). These ends are therefore 'sticky' or 'cohesive', much like the ends of many restriction fragments – but the longer length of these sticky ends makes the pairing much more stable, even at 37°C. The ends can be separated by heating lambda DNA, and if it is then cooled rapidly you will get linear monomeric lambda DNA. At low temperatures, the ends of the molecule will move slowly, and therefore the re-annealing of the sticky ends will take a long time. Eventually, however, it will resume a circular (although not covalently joined) structure. When lambda infects a bacterial cell

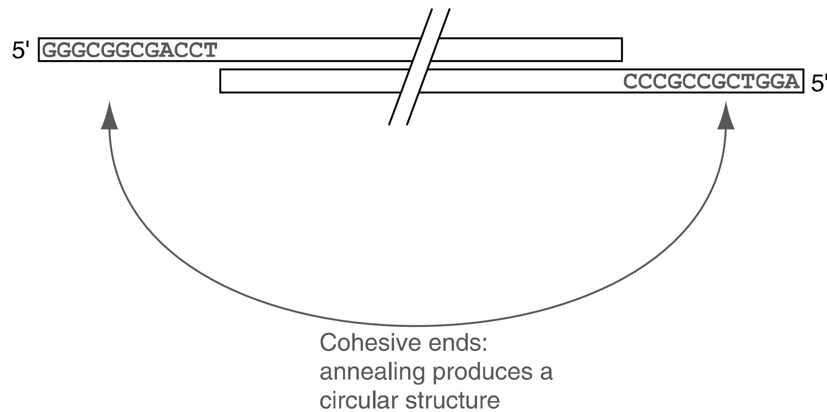


Figure 6.5 Cohesive ends of lambda DNA

and injects its DNA into the cell, it will therefore form a circular structure, with the nicks being repaired *in vivo* by bacterial DNA ligase. At about this time, a complex series of events occur that affect subsequent gene expression, determining whether the phage enters the lytic cycle or establishes lysogeny. We do not need to consider the details of the *lytic–lysogenic decision*, except to emphasize that it is essentially irreversible so that once started on one or the other route, the phage is committed to that process. However, we do need to consider the events in the lytic cycle.

Lytic cycle

In the lytic cycle, this circular DNA structure is initially replicated, in a plasmid-like manner (*theta* replication), to produce more circular DNA. Eventually, however, replication switches to an alternative mode (*rolling-circle replication*) which generates a long linear DNA molecule containing a large number of copies of the lambda genome joined end to end in a continuous structure (see Figure 6.6). While all this is going on, the genes carried by the phage are being expressed to produce the components of the phage particle. These proteins are assembled first of all into two separate structures: the head (as an empty precursor structure into which the DNA will be inserted), and the tail (which will be joined to the head after the DNA has been packaged).

The packaging process involves enzymes recognizing specific sites on the multiple-length DNA molecule generated by rolling-circle replication, and making asymmetric cuts in the DNA at these positions. These staggered breaks in the DNA give rise to the cohesive ends seen in the mature phage DNA; these sites are known as *cohesive end sites* (*cos* sites). Accompanying these cleavages, the region of DNA between two *cos* sites – representing a unit length of the lambda genome – is wound tightly into the phage head. This process is known

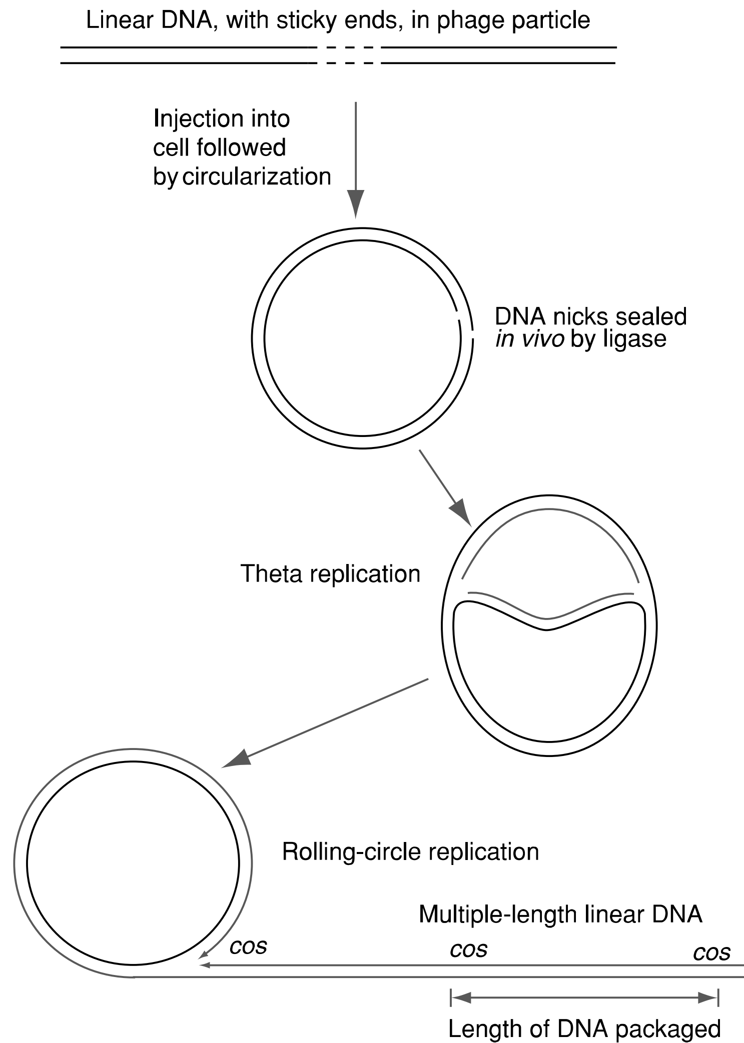


Figure 6.6 Replication of bacteriophage lambda DNA

as *packaging*. Following successful packaging of the DNA into the phage head, the tail is added to produce the mature phage particle, which is eventually released when the cell lyses.

Lysis of the bacterial cell is accomplished largely through the action of a phage-encoded protein, the product of gene *S*. Mutations in this gene can cause a delay or failure of lysis – which can be advantageous in increasing the yield of bacteriophages, as the replication of the phage will therefore continue for a longer time instead of being interrupted by lysis of the host cell. Many lambda vectors have such a mutation.

One of the most important features of this process from our point of view is that the length of DNA that will be packaged into the phage head is determined by the distance between two *cos* sites. If we insert a piece of DNA into our lambda vector, we will increase that distance, and so the amount of DNA to be

packaged will be bigger. However, the head is a fixed size, and can only accommodate a certain amount of DNA. It can take somewhat more than is present in wild-type lambda (up to about 51 kb altogether, which is about 5 per cent more than wild-type). As one of the reasons for using lambda is to be able to clone large pieces of DNA, this would be a serious limitation. The way round it is to delete some of the DNA that is normally present. This is possible, because the lambda genome contains a number of genes that are not absolutely necessary – especially if we only need lytic growth, when we can delete any genes that are solely required for the establishment of lysogeny. However, we cannot delete too much. The stability of the phage head requires a certain amount of DNA so, even though there are more genes that are not required, we cannot delete all that DNA. To produce viable phage, there has to be a minimum of 37 kb of DNA (about 75 per cent of wild-type) between the two *cos* sites that are cleaved.

The existence of these *packaging limits* is a very important feature of the design and application of lambda vectors, and also of cosmids which we will discuss later.

6.2.2 In vitro *packaging*

Naked bacteriophage DNA can be introduced into a host bacterial cell by transformation (often referred to as *transfection* when talking about phage DNA), in much the same way as we described for a plasmid. The big difference is that in this case instead of plating on a selective agar and counting bacterial colonies, we would mix the transfection mix with a culture of a phage-sensitive indicator bacterium in molten soft agar and look for *plaques* (zones of clearing due to lysis of the bacteria) when overlaid onto an agar plate. Note that in this case we do not need an antibiotic resistance gene as a selective marker.

However, the large size of most bacteriophage DNA molecules, including that of lambda, makes transfection an inefficient process compared with plasmid transformation, and not suitable for the generation of gene libraries which is the principal application of lambda vectors. However, there is a more efficient alternative. Some mutant lambda phages, in an appropriate bacterial host strain, will produce empty phage heads (as they lack a protein needed for packaging the DNA), while others are defective in the production of the head, but contain the proteins needed for packaging. The two extracts are thus complementary to one another. Use of the mixture allows productive packaging of added DNA, which occurs very effectively *in vitro* (including the addition of the tails). The resulting phage particles can then be assayed by addition of a sensitive bacterial culture and plating as an overlay, as above. Since *in vitro* packaging of lambda DNA is much more effective than transfection, it is the method that is almost always used.

One feature of this system that is markedly different from working with plasmid vectors is that the packaging reaction is most efficient with multiple-length DNA. The enzyme involved in packaging the DNA normally cuts the DNA at two different *cos* sites on a multiple-length molecule; monomeric circular molecules with a single *cos* site are packaged very poorly. So whereas with plasmid vectors the ideal ligation product is a monomeric circular plasmid consisting of one copy of the vector plus insert, for lambda vectors it is advantageous to adjust the ligation conditions so that we *do* get multiple end-to-end ligation of lambda molecules together with the insert fragments. The stickiness of the ends of the linear lambda DNA means this happens very readily.

6.2.3 Insertion vectors

The simplest form of lambda vectors is that known as an *insertion vector*. These are similar in concept to a plasmid vector, in that they contain a single cloning site into which DNA can be inserted. However, wild-type lambda DNA contains many sites for most of the commonly used restriction enzymes; you cannot just cut it with say *Hin*dIII and ligate it with your insert DNA. *Hin*dIII has seven sites in normal lambda DNA, and so will cut it into eight pieces. (Note the difference between a circular DNA molecule such as a plasmid, and a linear molecule like lambda: cut a circular DNA molecule once and you still have one fragment; cut linear DNA once and you have two fragments.) It would be almost impossible to join all these fragments (and your insert) together in the right order. To circumvent this, all lambda vectors have been genetically manipulated to remove unwanted restriction sites. In some cases this has been done by deleting regions of DNA carrying these sites (and hence also increasing the cloning capacity); another strategy is to select mutants with alterations in their sequence that result in the loss of the unwanted restriction sites.

In Figure 6.7, we see one example of a lambda vector, known as lambda gt10. In this vector, there is only a single site at which *Eco*RI will cut the DNA. The deletions and other manipulations that this phage has undergone have removed some of the unwanted sites, and have also reduced the overall size of the phage DNA to 43.3 kb (which is still large enough to produce viable phage particles, and still contains the genes that are needed for viability), and hence allows the insertion of foreign DNA up to a maximum of 7.6 kb.

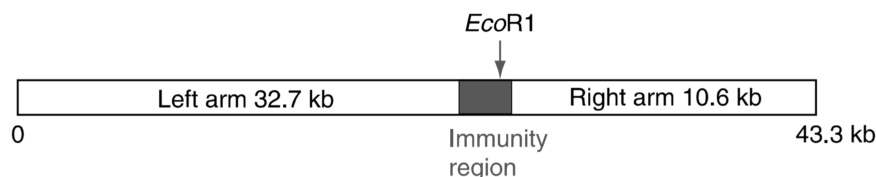


Figure 6.7 Lambda insertion vector gt10

Cutting this vector with *Eco*RI will produce two DNA fragments, referred to as the left and right arms. Although it therefore appears that the insert would have to be ligated to two different pieces of DNA, in practice this does not complicate the ligation as much as might be imagined. One end of each of the two fragments is derived from the cohesive ends of the lambda DNA, and will therefore anneal quite stably at 37°C – so although not covalently joined they can be considered as a single DNA fragment.

Lambda gt10 also provides us with another example of how insertional inactivation can be used to distinguish the parental vector (which may form by religation of the arms without an insert) from the recombinants. The *Eco*RI site is found within the repressor (*c*I) gene, so the recombinant phage, which carry an insert in this position, are unable to make functional repressor. As a consequence, they will be unable to establish lysogeny and will give rise to clear plaques, whereas the parental gt10 phage will give rise to turbid plaques. The distinction can be made even more marked, by using a host strain carrying the *hfl* (high frequency of lysogenization) mutation. In such a strain, any parental phage will establish lysogeny extremely efficiently, rather than entering the lytic cycle, with the result that few, if any, plaques will be obtained. This does not affect the recombinants, which are unable to make functional repressor, and therefore do not establish lysogeny. So you can achieve substantial enrichment of recombinant phage over the religated vector without having to resort to dephosphorylation with alkaline phosphatase.

Another example of a lambda insertion vector, lambda gt11, is used in a rather different way. Since it allows expression of the cloned fragment, it is considered later in this chapter, together with other expression vectors.

The packaging limits for lambda DNA are between 37 kb and 51 kb, as described above. In other words, we cannot make an insertion vector smaller than 37 kb, or we will be unable to grow it to produce the DNA that we need. Also, we cannot insert a DNA fragment so big that it would make the product larger than 51 kb; the recombinant DNA would be unable to be packaged into the phage heads. It follows from this that the maximum cloning capacity for an insertion vector is (51 minus 37) kb = 14 kb. This is larger than we would normally clone comfortably in a plasmid vector, but still smaller than we would like for some purposes. In order to increase the available cloning capacity, we have to turn to a different type of lambda vector, known as a *replacement vector*.

6.2.4 Replacement vectors

The packaging limits that restrict the cloning capacity of insertion vectors are imposed by the physical requirements of the phage head rather than by the nature of the genes needed. There are more genes that are not essential for lytic

growth and could be deleted, except that it would make the phage DNA too small to produce viable progeny. That provides a clue to an alternative design of lambda cloning vectors. Instead of merely inserting extra DNA, arrange the vector so that a piece of DNA can be removed and replaced by your insert – hence the term *replacement vector*.

Figure 6.8 shows an example of a lambda replacement vector, EMBL4. Instead of being cut just once by the restriction enzyme of choice (in this case *Bam*HI), there are two sites where the DNA will be cleaved. The vector DNA will therefore be cut into three fragments: the left and right arms (which will anneal by virtue of their cohesive ends) and a third fragment which is not needed (except to maintain the size of the DNA) and can be discarded. Since the only purpose of this fragment is to help to fill up the phage head it is known as a *stuffer fragment*.

In use, therefore, this vector would be cut with *Bam*HI and the fragments separated, e.g. by gel electrophoresis. The stuffer fragment would be thrown away, and the arms mixed with the restriction fragments to be cloned. These could be generated by *Bam*HI digestion, or by cleavage of the target with another enzyme that produces compatible ends, e.g. *Sau*3A (see Chapter 5). Ligation of the mixture would produce recombinant phage DNA that would be packaged into phage heads by *in vitro* packaging. The cloning capacity of the vector is thus considerably increased; in this case the size of the arms combined comes to 29 kb and thus you can clone fragments up to (51 minus 29) kb = 22 kb.

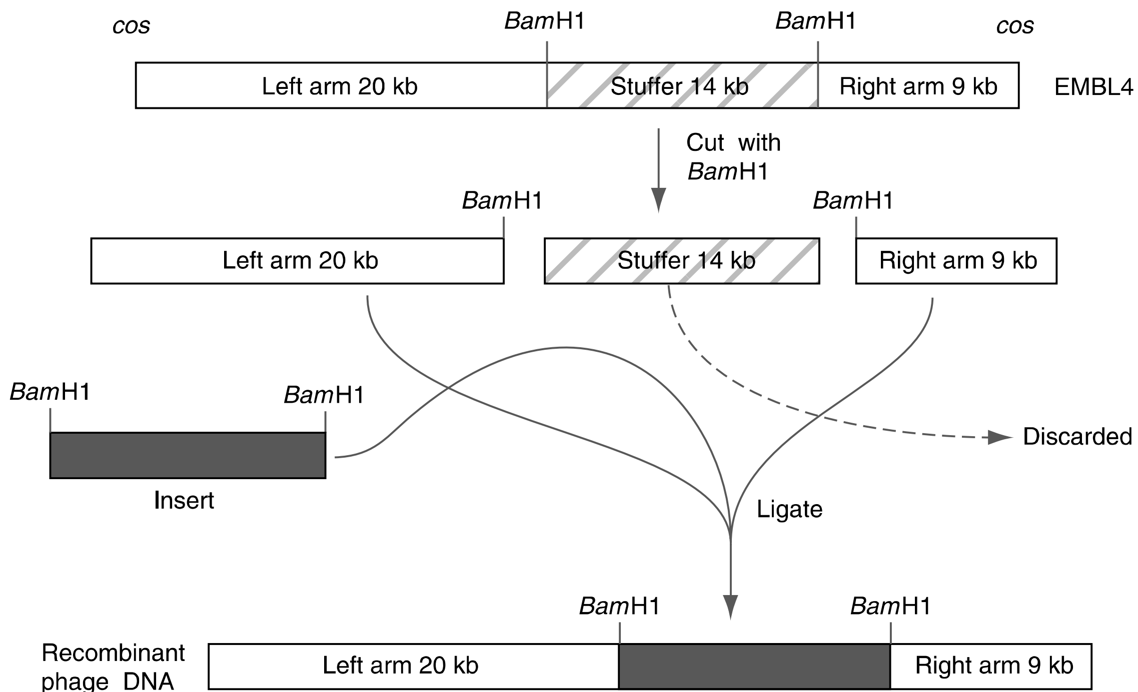


Figure 6.8 EMBL4

There is a further advantage to such a vector. The combined size of the arms is only 29 kb, which is less than the minimum required for packaging. Any pairs of arms that are ligated without an insert will therefore be too small to produce viable phage particles. Such viable particles will only be produced if ligation results in an insert of at least (37–29) kb = 8 kb. The vector thus provides a positive selection for recombinants as opposed to parental phage, and furthermore for recombinants that contain an insert of at least 8 kb. The gene library will therefore be free of non-recombinant phage.

In Chapter 5 we discussed strategies, with plasmid vectors, for ensuring that we obtained recombinant progeny rather than parental vector molecules, including the use of alkaline phosphatase treatment of the vector to prevent recircularization. This is not necessary with replacement vectors. Since we do not have to treat the vector with phosphatase, we have another possibility – dephosphorylation of the *insert*. In the production of a gene library, the insertion of more than one fragment into the same vector molecule is a problem that can give rise to anomalies in characterizing the insert in relation to the genome it came from. Phosphatase treatment of the insert will prevent insert–insert ligation, and hence will ensure that all of the recombinants carry only a single insert fragment.

There is yet another useful feature that can be built into a replacement vector. Since the stuffer fragment is not necessary for phage production (apart from filling up the phage head), it does not have to be lambda DNA. It can be anything we want, so we could for example put in a fragment carrying a β-galactosidase gene. Then any plaques formed by phage that still carry the stuffer fragment would be blue (on a medium containing X-gal). Of course ideally there should not be any; but there may be some phage DNA molecules that have not been cut completely, or there may be some stuffer DNA contaminating the preparation of the vector arms, which could then be ligated back into the vector. Any plaques containing the stuffer will be blue – so you have an immediate check that everything has gone according to plan, or not.

So we see that lambda vectors provide a highly versatile and efficient system for primary cloning of unknown fragments, especially in the construction of genomic and cDNA libraries (see Chapter 7). They extend the cloning capacity over that readily obtainable with plasmid vectors, and can easily generate the very large numbers of recombinants that are required for a gene library. However, some people do not like working with lambda systems, mainly because it requires a different set of techniques for growing, assaying and maintaining phage preparations. There is nothing really difficult about it; it is just unfamiliar. The only real disadvantage to lambda cloning systems is the size of the vector DNA. With an insertion vector, your recombinant may contain 5 kb of insert and 45 kb of vector. This makes it more difficult to analyse or manipulate your insert than would be the case with a plasmid

recombinant – especially as the lambda vector will contain a substantial number of recognition sites for different restriction enzymes. The normal procedure therefore, having identified the recombinant clone of interest, would be to reclone the insert (or part of it) into a plasmid vector for further analysis and manipulation.

Although lambda phages are the most widely used phage vectors, there are other phages, or vectors based on them, that are used for specific purposes, including P1 and M13. These are discussed further in subsequent sections in this chapter; but first we need to look at a special class of vector that combines some of the features of lambda and plasmid vectors, and enables the cloning of even larger pieces of DNA. These are the *cosmids*.

6.3 Cosmids

The lambda packaging reaction has two fundamental requirements: the presence of a *cos* site, and the physical size of the DNA. Cosmids exploit this to provide cloning vectors with a capacity larger than can be achieved with lambda replacement vectors.

Basically, a cosmid is simply a plasmid which contains a *cos* site. As with all plasmid vectors, it has an origin of replication, a selectable marker (usually an antibiotic resistance gene), and a cloning site. Digestion, ligation with the potential insert fragments, and subsequent purification of recombinant clones, are carried out more or less as for a normal plasmid vector. However, instead of transforming bacterial cells with the ligation mix, as you would normally with a plasmid, you subject the ligation mixture to *in vitro* packaging as described above for lambda vectors. Since the cosmid carries a *cos* site, it can be a substrate for *in vitro* packaging, but only if it is big enough. The vector itself is quite small – in the example shown in Figure 6.9, it is 5.4 kb, which is much too small for successful packaging. The packaging reaction will only be successful if you have inserted a DNA fragment between about 32 kb and 45 kb

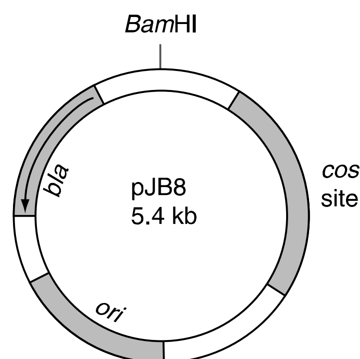


Figure 6.9 Structure of a cosmid

in size. So you not only have an increased cloning capacity, but also a positive selection for an insert, and for an insert that is quite large.

There is one difference in the ligation procedure between cosmids and plasmids. The *in vitro* packaging reaction requires two separate *cos* sites and therefore works best with multimeric structures resembling the normal lambda DNA substrate. The efficiency of cosmid cloning can therefore be enhanced by adjusting the ligation conditions to favour multimer formation, or by using more complex cosmid designs and/or more sophisticated digestion and ligation strategies that ensure that only multimeric ligation products are made. Note, however, that although the ligation results in multimeric structures, the *in vitro* packaging reaction will ensure that each recombinant clone contains only a monomeric cosmid.

Of course the products of the packaging reaction, although they are phage particles, will not give rise to more phages after infection of a host bacterium. They do not carry any of the genes that are needed for production of more phage particles, nor for lysis of the cell. So you will not get phage plaques. However, the cosmid will replicate as a plasmid, and hence will give rise to more cosmid-containing cells, and these can be selected as colonies on agar containing an antibiotic (in this case ampicillin).

For very large genomes, such as mammalian ones, the cloning capacity of cosmids is still rather too small for convenience, and alternative vectors with even greater capacity are now available (see below). However, for smaller and medium-sized genomes, cosmids can be extremely valuable. A complete bacterial genome can be covered by a selection of only a few hundred cosmids, which can be useful in genome mapping and sequencing projects.

Cosmids have some advantages over phage lambda, particularly in that they can be propagated and purified by conventional plasmid-oriented techniques, without having to become familiar with phage technology. Furthermore, with cosmids the vector is small compared with the insert, in contrast to lambda where more than half of the DNA of a recombinant phage is derived from the vector. Subcloning of fragments of your insert, to obtain fragments carrying just the gene you need, is therefore rather easier with cosmids than with lambda vectors. On the other hand, a lambda library can be easily stored as a pool of phage particles in a single tube.

6.4 M13 Vectors

Like lambda, M13 is a bacteriophage that infects *E. coli*. That is about as far as the resemblance goes, either biologically or in their use as cloning vectors.

M13 is a 'sex-specific' bacteriophage, or more accurately F-specific. It attaches to the tips of the pili that are produced on the surface of bacteria that carry an F-type plasmid, and is therefore unable to infect bacteria that do

not carry such a plasmid. The very long, thin filamentous phage particles contain a circular, single-stranded DNA molecule of about 6 kb. After this DNA enters the cell, it is converted to a double-stranded molecule (the *replicative form, RF)* by synthesis of the complementary strand. This molecule is replicated by producing a circular single-stranded copy of one strand of the RF. This single-stranded DNA is again converted to a double-stranded form (see Figure 6.10). The production of the single-stranded intermediate requires a specific signal on the DNA, and is therefore completely strand-specific, i.e. it is always the same strand that appears in this form. This separation of the synthesis of the two strands is not unique to M13 but is found in some other bacteriophages and also some classes of plasmids. However, most of the phage DNA within the cell is double-stranded circular DNA (RF), and can be isolated by conventional plasmid purification methods.

Continued replication of the phage DNA leads to a build up of these plasmid-like DNA molecules within the cell. At the same time, expression of phage genes occurs, and the product of one of these genes binds to the single-stranded

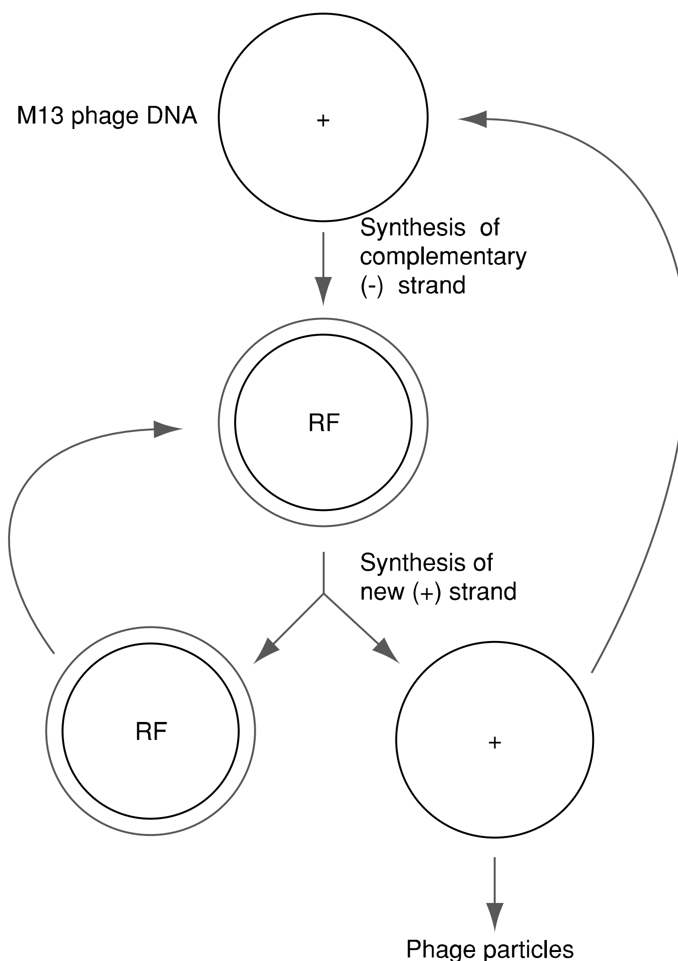


Figure 6.10 Replication of single-strand bacteriophages

product, initiating production of phage particles. This occurs by extrusion of the DNA through the cell membrane, during which process it becomes coated with phage proteins. The length of the filamentous phage particle is determined by the length of the DNA molecule, unlike lambda where the size of the particle is determined by the structure of the proteins of which it is composed. Hence, there are no absolute packaging limits for M13, although the phage does become increasingly fragile if large DNA fragments are inserted.

A curious, and significant, feature of M13 is that infection does not lead to bacterial lysis. Phage particles continue to be produced, and the cell remains viable, although it grows more slowly. Infection does result in the appearance of 'plaques' in a bacterial lawn, but these are zones of reduced growth rather than zones of lysis. As a consequence of the continuing viability of the host cell, very high titres of phage can be produced.

The main advantage of M13 is that it provides a very convenient way of obtaining single-stranded versions of a gene, which would be difficult to do in any other way. Single-stranded DNA obtained from M13 clones has been widely used for DNA sequencing (see Chapter 10). Although, nowadays, double-stranded DNA templates are commonly used for sequencing, single-stranded vectors are still preferred by some laboratories. Another application where single-stranded DNA can be advantageous (although not essential) is site-directed mutagenesis (Chapter 15). M13 vectors also have another role, not connected with the production of single-stranded DNA: this is the technique known as *phage display* (see Chapter 14).

Vectors based on M13 have been engineered to contain multiple cloning sites, and these vectors usually include a beta-galactosidase gene to distinguish recombinants from parental vector, as was described for pUC18 earlier in this chapter. DNA fragments can be cloned into such sites, using the plasmid-like replicative form; after transformation, the progeny are detected by a plaque assay, with 'white' plaques indicating a recombinant clone and blue plaques (on a medium containing X-gal) suggesting non-recombinant phage. As discussed earlier, in relation to pUC18, some blue plaques may occur through insertion of small fragments if they do not shift the reading frame of the *lac*Z gene.

If you then isolate phage particles from the supernatant of an infected culture, these will contain your gene in single-stranded form. Note that because of the specificity of the replication process, it is always the same strand that is found in all phage particles.

6.5 Expression Vectors

The above discussion has assumed that all you want to do is to clone a piece of DNA. It does not consider the possibility that you might want to obtain

expression of the gene encoded by that DNA. If you take a DNA fragment from another organism and clone it in *E. coli*, there are many reasons why it may not be expressed. At the simplest level, these relate to the signals necessary for initiating transcription (a promoter) and translation (a ribosome-binding site and start codon). The basic way of encouraging (although not ensuring) expression of the cloned gene is to incorporate these signals into the vector, adjacent to the cloning site. This is then known as an *expression vector*.

Expression vectors are of two main types (see Figures 6.11 and 6.12). If the vector just carries a promoter, and relies on the translation signals present in the cloned DNA, it is referred to as a *transcriptional fusion* vector. On the other hand, if the vector supplies the translational signals as well (so you are inserting the cloned fragment into the coding region of a vector gene), then you have a *translational fusion*. Note that in this case the insert must be in frame with the start codon. The plasmid vector pUC18 that we looked at earlier is actually a translational fusion vector, although not often used as such. A better example is the lambda vector gt11 (see Figure 6.13). This is an insertional vector, 43.7 kb in length (making the maximum cloning capacity about 7 kb). It has been engineered to contain a β-galactosidase gene, and has a single *Eco*RI restriction site within that gene. This confers two properties on the vector. Firstly, insertion of DNA at the *Eco*RI site will inactivate the β-galactosidase gene, so that recombinants will give 'white' plaques on a medium containing X-gal. Secondly, the insert, if in the correct orientation and in frame, will give rise to a

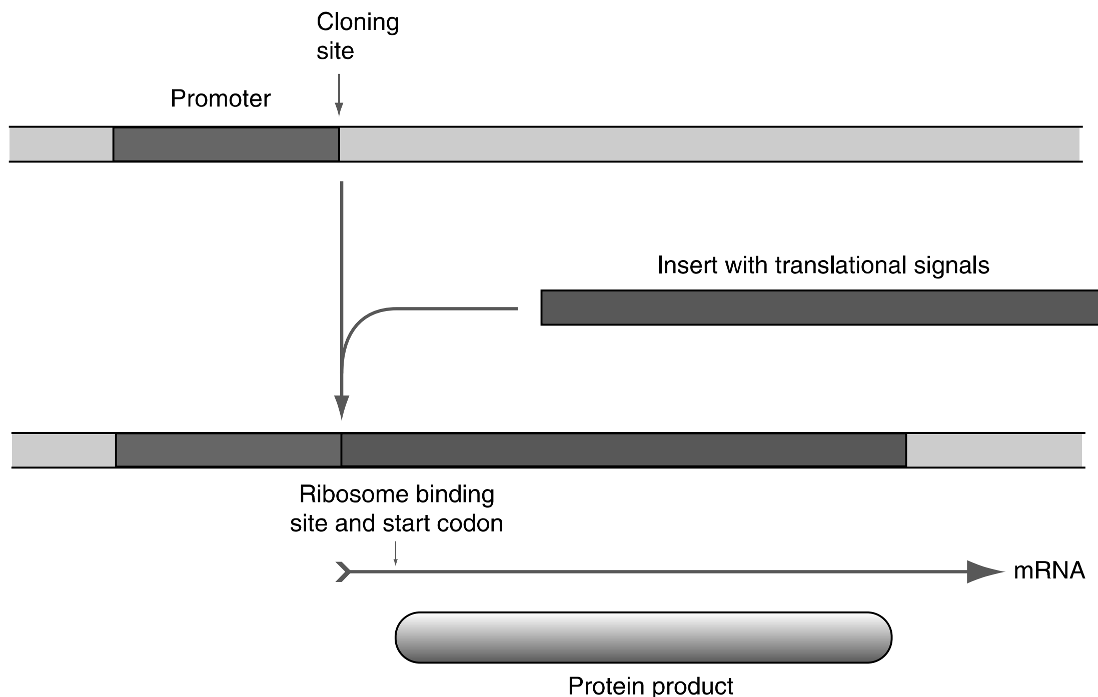


Figure 6.11 Expression vectors: transcriptional fusions

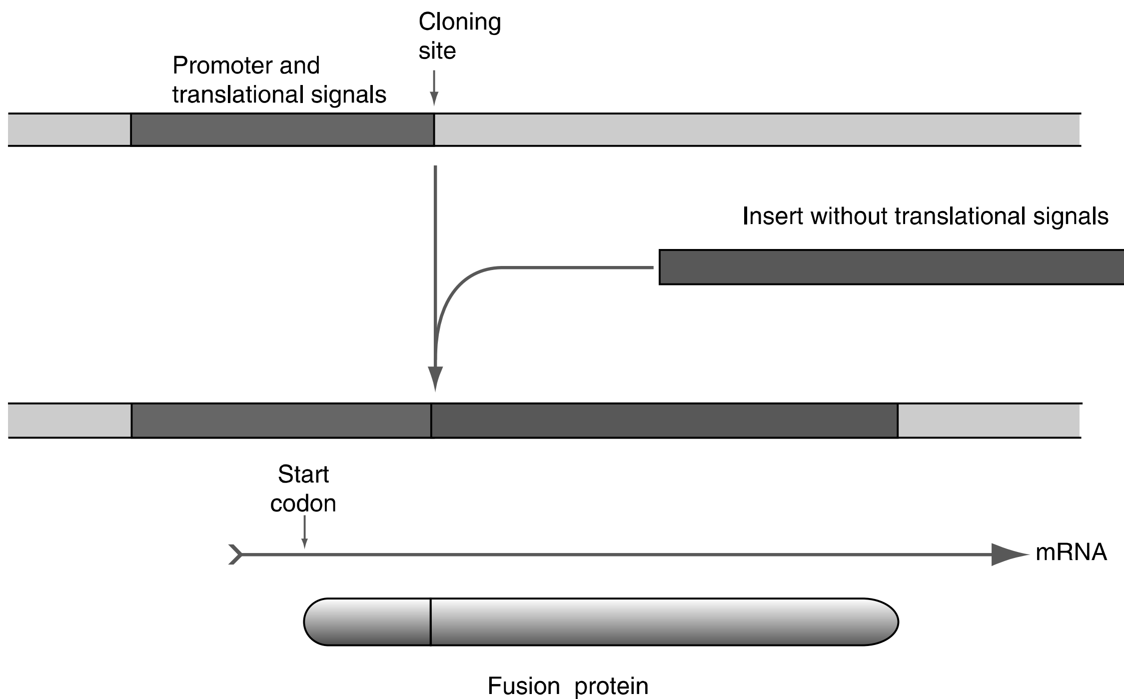


Figure 6.12 Expression vectors: translational fusions

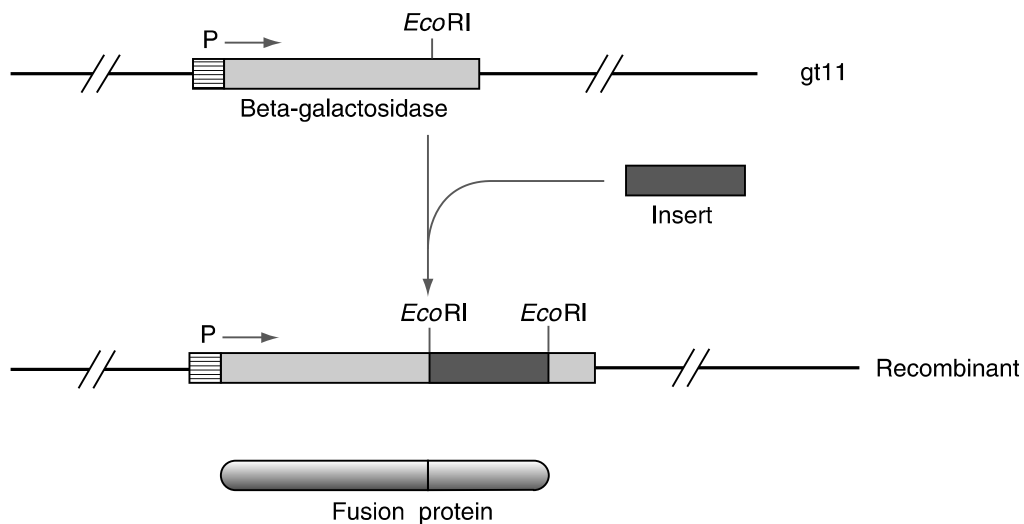


Figure 6.13 Use of lambda gt11 for generation of fusion proteins

fusion protein containing the product encoded by the insert fused to the β-galactosidase protein. This fusion protein is unlikely to have the biological functions associated with your cloned gene – it contains too much extraneous material – but that is not the point. It *is* reasonably likely to react with some antibodies to the natural product, which makes it a useful way of detecting the clone of interest, as we will see in Chapter 8.

To make full use of an expression vector you need to be able to choose whether to have expression on or off. The regulation of most inducible bacterial promoters, such as that of the *lac* operon, is rather 'leaky', i.e. there is still some expression even in the uninduced or repressed state. Firmer control can be achieved by the use of promoters from bacteriophages, notably one from the bacteriophage T7. In T7, this promoter controls the expression of the 'late' genes, i.e. the genes that are only switched on at a late stage of infection. This promoter is not recognized by *E. coli* RNA polymerase, but requires the T7 RNA polymerase, a product of genes expressed earlier in the infection cycle. So if we clone our DNA fragment downstream from a T7 promoter, using an 'ordinary' *E. coli* host (lacking a T7 polymerase gene) we will get no expression at all. This can be useful, as the product might be deleterious to the cell. Once we are satisfied that we have made the right construct, we can isolate the plasmid and put it into another *E. coli* strain that has been engineered to contain a T7 polymerase gene, and hence will allow transcription of the cloned gene. If the expression of the T7 polymerase gene is itself regulated, for example by putting it under the control of a *lac* promoter, then we can turn the expression of the T7 polymerase up (by adding IPTG) or down, so we still have control over the level of expression. Further devices can be included to inhibit the low level of T7 polymerase arising from the leakiness of the uninduced *lac* promoter.

The pGEM® series of vectors (see Figure 6.14) provide an example. In this case, there is a multiple cloning site adjacent to the T7 promoter, so any DNA inserted will be under the control of the T7 promoter. This is a transcriptional fusion vector, and it is often more useful for generating substantial amounts of an RNA copy of your cloned fragment, which can then be used as a probe for hybridization. The vector actually has a second specific promoter, derived from another bacteriophage (SP6), at the other side of the multiple cloning site, so if you provide an SP6 polymerase you will get an RNA copy of the other strand, the antisense strand. The usefulness of this will be apparent when we consider applications of antisense RNA, such as the RNase protection assay (Chapter 13).

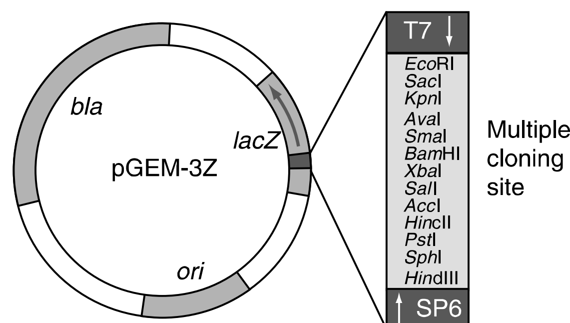


Figure 6.14 Structure of the expression vector pGEM-3Z: *bla* = beta-lactamase (ampicillin resistance); *ori* = origin of replication

There are also a variety of translational fusion vectors with T7 promoters, which are designed for obtaining high, controllable, levels of protein expression. We will return to the concept of expression vectors, and other factors that have to be considered for the optimization of protein production, in Chapter 15.

6.6 Vectors for Cloning and Expression in Eukaryotic Cells

Most primary cloning (i.e. the initial isolation of a gene or other DNA fragment from a target organism) is carried out using bacterial hosts (usually *E. coli*), because of the ease of manipulation and the range of powerful techniques that have been developed. Eukaryotic hosts are more commonly used for studying the behaviour of genes that have already been cloned (but in an environment more closely related to their original source), for analysing their effect on the host cell and modifying it, or for obtaining a product which is not made in its natural state in a bacterial host. There is therefore more emphasis with eukaryotic vectors in obtaining gene expression rather than making gene libraries or primary cloning (with the notable exception of YAC vectors, see below). There is a bewildering variety of vectors available for cloning in different eukaryotic hosts, and a full review of them is way beyond the scope of this book. In this chapter, we want to introduce some of the main concepts, many of which are similar in principle to those of bacterial cloning vectors although there are significant differences; we will consider further the use of eukaryotic hosts for product formation in Chapter 15, and the genetic modification of animal and plant cells (or whole animals and plants) in Chapter 17.

6.6.1 Yeasts

Microbiologically, ‘yeasts’ are single-celled fungi, as opposed to filamentous fungi, but the term is quite imprecise. Not all ‘yeasts’ are related taxonomically, and indeed some filamentous fungi can also grow in a unicellular form that is referred to as a yeast form. Although in common usage the term ‘yeast’ would be taken to mean the brewer’s/baker’s yeast *Saccharomyces cerevisiae*, even molecular biologists are starting to have to recognize the existence of other yeasts (especially members of the genus *Pichia*, which we will encounter again in Chapter 15). However, for the moment we will limit ourselves to *S. cerevisiae*.

The vectors that will be most familiar, after reading about bacterial cloning vectors, are the yeast episomal plasmids (YEp). These are based (Figure 6.15) on a naturally occurring yeast plasmid known as the 2 μm plasmid, and they are therefore able to replicate independently in yeast, at a high copy number

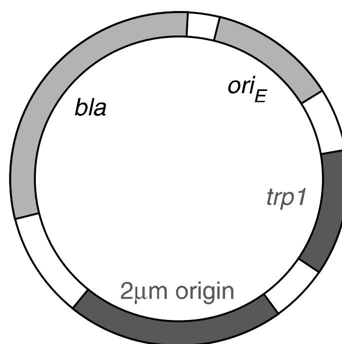


Figure 6.15 Structure of a yeast episomal vector: *bla* = beta-lactamase (ampicillin resistance); *ori* = origin of replication in *E. coli*; *trp1* = selectable marker in *S. cerevisiae* auxotrophs; 2 µm origin = origin of replication in *S. cerevisiae*

(25–100 copies per cell). As is usually the case with vectors for eukaryotic cells, these plasmids also have an *E. coli* origin of replication, enabling them to be grown and manipulated in an *E. coli* host (i.e. they are *shuttle vectors*). There is one point of detail in which they differ from bacterial cloning vectors, and that is the nature of the selectable marker. For bacterial vectors, we can exploit the large number of antibacterial antibiotics (and the correspondingly large number of antibiotic resistance genes) to enable us to select our transformants. There are fewer antibiotics available to which yeasts are sensitive (although some fungicides can be used), and therefore selection more commonly makes use of complementation of auxotrophic mutations in the host strain. For example, a host strain of *S. cerevisiae* with a mutation in the *trp1* gene will be unable to grow on a medium lacking tryptophan. If the vector plasmid carries a functional *trp1* gene, then transformants can be selected on a tryptophan-deficient medium. Other commonly used selectable markers include *ura3* (uracil), *leu2* (leucine), and *his3* (histidine). These vectors would usually also carry an antibiotic resistance marker for selection in *E. coli*.

Vectors that replicate as plasmids in *S. cerevisiae* are often rather unstable, in that they tend to be lost from the culture as plasmid-free daughter cells accumulate. This is due to erratic partitioning during mitosis. Newer versions of YEp vectors, taking advantage of a better understanding of the biology of the 2 µm plasmid, are more stable.

Autonomously replicating plasmids can also be constructed by inserting a specific sequence from a yeast chromosome; this sequence is known, unsurprisingly, as an *autonomously replicating sequence*, or *ars*. These plasmids are very unstable, but some constructs that also include a centromere are more stable (Figure 6.16). In contrast to the YEp vectors, these yeast centromere plasmids (YCp), are normally maintained at a low copy number (1–2 copies per cell), which can be advantageous if your product is in any way harmful to the cell, or if you want to study its regulation.

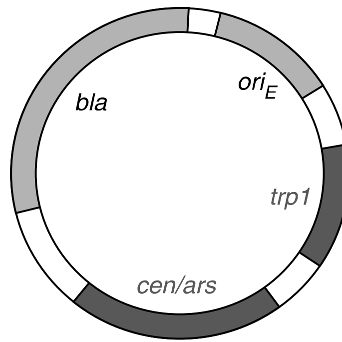


Figure 6.16 Structure of a yeast centromere vector: *bla* = beta-lactamase (ampicillin resistance); *ori* = origin of replication in *E. coli; trp1* = selectable marker in *S. cerevisiae* auxotrophs; *cen/ars* = centromere and autonomously replicating sequence, providing an origin of replication in *S. cerevisiae*

Since one of the purposes of using yeast as a host is to express the cloned gene, these vectors are commonly designed as expression vectors. The principles involved are similar to those described above for bacterial expression vectors, except that of course the expression signals involved are those applicable to *S. cerevisiae* rather than *E. coli*. If required, this can include signals for secretion, or for targeting to the nucleus or other cellular compartments. Note that *S. cerevisiae*, although eukaryotic, has very few introns, and is not the host of choice if you want to ensure correct excision of introns.

The vectors described are maintained in yeast as circular DNA molecules, much like a bacterial plasmid. Two other classes of vectors deserve a mention. First, there are the yeast integrating plasmids (YIp). These do not replicate independently but integrate into the chromosome by recombination (Figure 6.17). The frequency of transformation is very low, and it is difficult to recover the recombinant vector after transformation. The main advantage is that the transformants are much more stable than those obtained with the autonomously replicating plasmids.

Second, there are the yeast artificial chromosomes (YACs), which carry telomeres that enable their maintenance in *S. cerevisiae* as linear structures resembling a chromosome. The use of these vectors, for cloning very large pieces of DNA, is quite distinct from the uses of the vectors described above. YACs are considered further, along with other vectors used for the same purpose, in a later section of this chapter.

6.6.2 *Mammalian cells*

In bacteria, the cloning vectors used replicate separately from the chromosome, as plasmids or bacteriophages. As we have seen above, the same is true of many

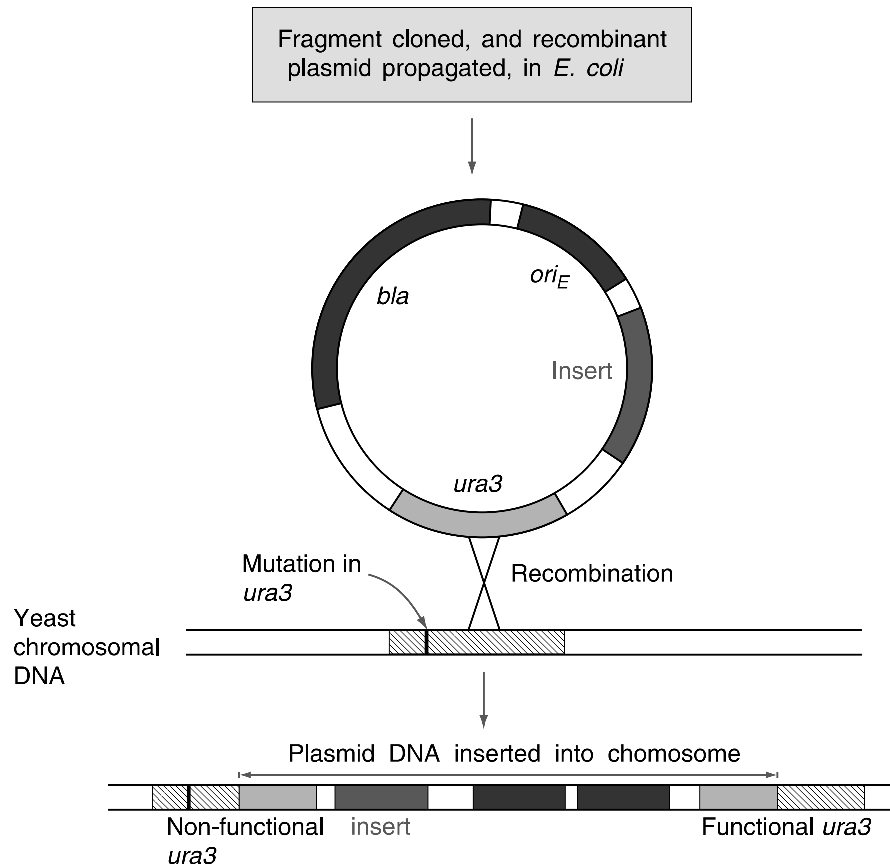


Figure 6.17 Structure and use of a yeast integrative plasmid (YIp)

types of vectors used in yeast. The situation with cloning in mammalian cells is somewhat different in that independent, plasmid-like, replication is often not sustained. Some vectors are capable of plasmid-like replication, especially those carrying the origin of replication from the virus SV40 (simian virus 40), which replicate episomally in some mammalian cells (such as COS cells). More stable clones are obtained by inserting the DNA into the chromosome, which happens readily in mammalian cells. In either case, the cloning vector enables you to organize your cloned gene in relation to a set of expression signals, many of which are derived from viruses such as SV40 or cytomegalovirus (CMV). The general features of such a vector are exemplified in Figure 6.18. A gene inserted at the multiple cloning site (MCS) enables high level constitutive expression from the CMV promoter, while the presence of a polyadenylation signal increases mRNA stability. The SV40 origin allows episomal replication in COS cells, and the neomycin phosphotransferase gene permits selection for resistance to the antibiotic G-418 (Geneticin®). Note that this is a shuttle vector, carrying an *E. coli* origin of replication and an ampicillin resistance gene (β-lactamase), so the construction can be carried out in *E. coli* before

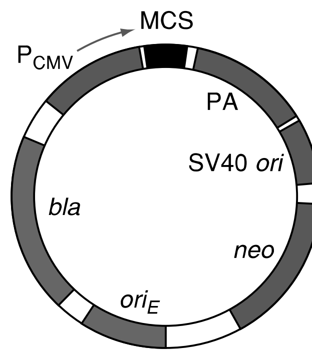


Figure 6.18 Structure of a basic episomal vector for gene expression in mammalian cells: P_{CMV} = CMV promoter – high level constitutive expression in mammalian cells; MCS = multiple cloning site; PA = polyadenylation signal; SV40 *ori* = origin of replication, episomal replication in *cos* cells; *neo* = neomycin phosphotransferase, resistance to G418 for selection in mammalian cells; ori_E = *E. coli* replication origin; *bla* = beta-lactamase (ampicillin resistance) for selection in *E. coli*

transferring the recombinant plasmid to a mammalian cell line. There is a wide variety of commercially available expression vectors for mammalian cells, with more sophisticated features than that shown. We will return to the topic of expression of cloned genes in mammalian cells in Chapter 15.

There are other types of vector available, based on various viruses, which can be used to transmit your cloned gene from one cell to another. Of these, the *retroviral* vectors deserve a special mention, and in order to understand these we need a brief account of retroviral biology. Retroviruses have an RNA genome. When a cell is infected, the RNA is copied into double-stranded DNA by the action of a viral protein, *reverse transcriptase*. This protein is present in the virion and enters the cell along with the RNA. (Reverse transcriptase is formally an RNA-directed DNA polymerase, and we will encounter it again in Chapter 7 where we consider its use in the production of cDNA from mRNA templates.) This DNA then circularizes and is integrated into the host cell DNA by the action of another virion protein known as *integrase*. The efficiency of integration of the DNA into the genome is one of the main attractions of this system for genetic manipulation of animal cells.

The integrated DNA is bounded by sequences known as *long terminal repeats* (LTR) which include a strong promoter for transcription of the integrated viral genes *gag*, *pol* and *env*. Full-length transcripts provide the viral RNA which is assembled into virus particles; one region of the virus, known as the *psi* site, is essential for this process. The packaged virus particles acquire envelope glycoproteins from the host cell membrane as they bud off from the cell, without lysis. These glycoproteins determine the type of receptors the virus uses to infect further cells.

Development of vectors based on retroviruses rests on the knowledge that most of these functions can be provided *in trans*, e.g. by genes from a defective helper virus already integrated into the genome of the host cell. The main features that are *cis*-acting, and therefore need to be located on the vector itself, are the LTR sequences and the *psi* site.

The basic features of the use of such a vector are outlined in Figure 6.19. The vector is a shuttle plasmid, so *E. coli* is used for construction of the recombinant plasmid by inserting the required gene at the multiple cloning site. This construct is then used to transfect a culture of a special cell line (*helper cells*) that contains the *gag*, *pol* and *env* genes required for virus production, integrated into the genome. The transfected cells will therefore be able to produce virus particles containing an RNA copy of your construct. These particles are able to infect other cells that do not contain the integrated essential genes; since

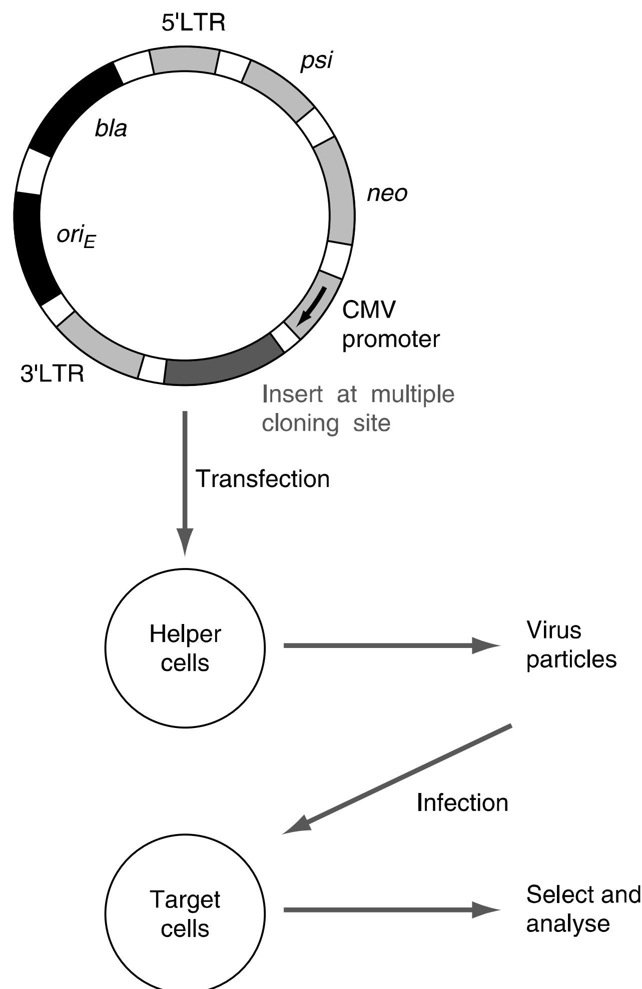


Figure 6.19 Structure and use of a retroviral vector: *bla* = selection in *E. coli*; ori_E = *E. coli* replication; *neo* = selection in infected cells; CMV promoter is for transcription of cloned gene; LTR, *psi* = essential *cis*-acting retroviral sequences (see text)

the viral particles carry preformed reverse transcriptase and integrase, the RNA will be copied into DNA in such cells, and the DNA will be efficiently integrated into the genome. However, since these cells do not carry the essential genes, no further production of viral particles will occur. However, your gene is now stably integrated into the chromosome and can be expressed from the adjacent promoter derived from the vector.

The specificity of the viral particles for other cells will be determined by the envelope gene carried by the helper cells. Replacing that gene by other genes for envelope glycoproteins from other viruses, in particular the VSV-G gene from vesicular stomatitis virus, enables a wider range of target cells to be used, not just mammalian cells but extending to, for example chickens, oysters, toads, zebrafish, mosquitoes – in fact cells from virtually all non-mammalian (and mammalian) species can be infected.

The expression of genes in mammalian cells, and the incorporation of foreign genes into the genome of whole animals (*transgenesis*) is considered further in Chapters 15 and 17 respectively.

6.7 Supervectors: YACs and BACs

Although cosmids were the first vectors that made the production and use of mammalian gene libraries feasible, their limited capacity would still not have sufficed for decoding the human genome. This was made possible by the development of novel supervectors that were able to carry 100 kb or more. The first one of these was the yeast artificial chromosome (YAC). Similarly to cosmids, these could be constructed through knowledge of what features were necessary to enable the vector to be carried by its host (in the case of yeast, telomeres, a centromere, and an origin of replication), as well as selectable markers and cloning sites (Figure 6.20). As with the shuttle plasmids we have referred to earlier, the YAC vector is propagated as a circular plasmid in *E coli*. Restriction enzyme digestion removes the stuffer fragment between the two telomeres, and cuts the remaining vector molecule into two linear arms, each carrying a selectable marker. The insert is then ligated between these arms, as in the case of phage lambda, and transformed into a yeast cell, with selection for complementation of both auxotrophic markers. This ensures that the recombinants contain both arms. Furthermore a successful recombinant must contain the *tel* sequences at each end, so that the yeast transformant can use these sequences to build functional telomeres. The titans amongst vectors, YACs are routinely used to clone 600 kb fragments, and specialized versions are available which can accommodate inserts close to 2 Mb. As such, they will not only easily accommodate any eukaryotic gene in its entirety, but also complete within their framework of three-dimensional structure and distant regulatory sequences. They have therefore been very useful in the production of transgenic organisms (see Chapter 17).

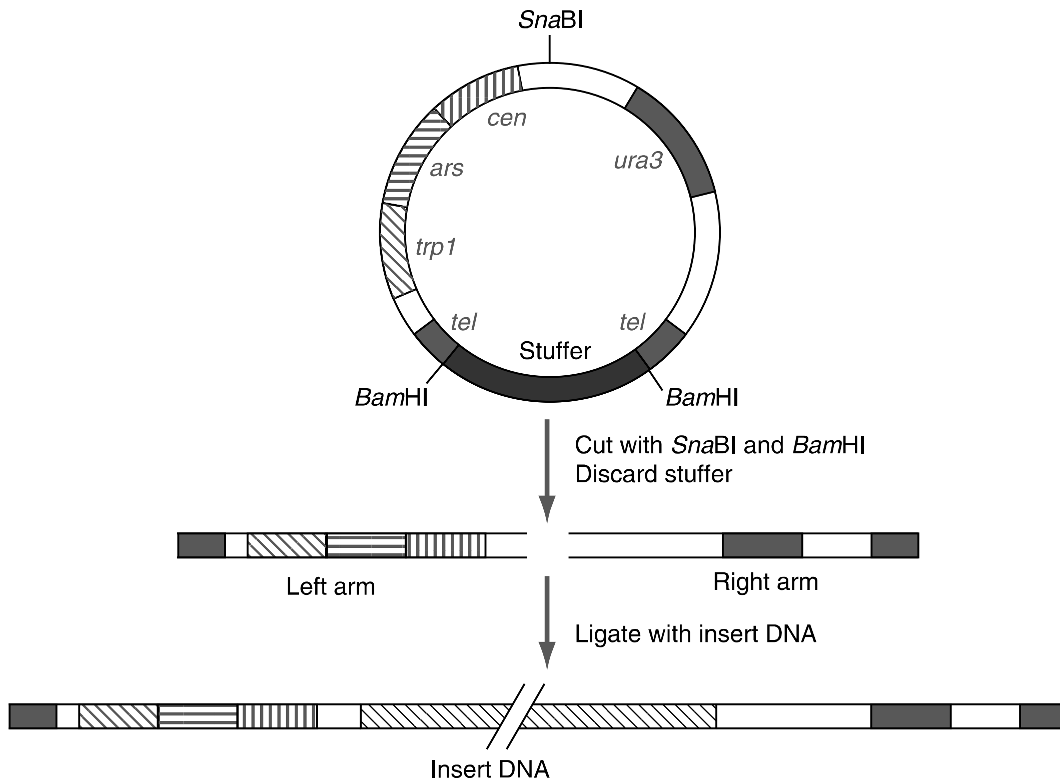


Figure 6.20 Structure and use of a yeast artificial chromosome vector: *trpl, ura3* = selectable markers; *cen/ars* = centromere and autonomously replicating sequence enabling replication in *S. cerevisiae; tel* = telomere; for simplification the *E. coli* origin of replication and selectable markers for *E. coli* are not shown

However, YACs have problems with the stability of the insert, especially with very large fragments which can be subject to rearrangement by recombination. Furthermore, apart from the fact that many laboratories are not set up for the use of yeast vectors, the recombinant molecules are not easy to recover and purify. Thus, larger bacterial vectors are used more than YACs even though their capacity is lower. These include vectors based on bacteriophage P1, which are able to accommodate inserts in excess of 100 kb, and bacterial artificial chromosomes (BACs), which are based on the F plasmid and can accommodate 300 kb of insert. These vectors lack the instability problems found in YACs, and play an important role in genome sequencing projects (see Chapter 10).

6.8 Summary

Earlier in this chapter we considered the advantage that can be gained in constructing gene libraries using vectors with higher cloning capacity such as

lambda replacement vectors and cosmids. Even larger fragments can be cloned by employing other vector systems. For relatively small genomes, such as those of bacteria, there is not usually a need for such large inserts. If we simply find out that our gene is somewhere in a 300 kb insert, there is still a lot of work to do to find out exactly where it is. We might as well go straight to a library in a lambda replacement vector; with inserts of 15–20 kb, a library of a few thousand clones will provide adequate coverage of a bacterial genome. Screening a library of this size is quite straightforward. However, libraries of larger fragments are extremely useful for establishing physical maps of the chromosome and as an adjunct to genome sequencing projects (see Chapter 10).

However, for larger genomes, such as those of mammalian cells, a reasonably complete gene library in a lambda replacement vector would require hundreds of thousands of clones, and screening then becomes laborious. In these cases, the use of larger inserts is advantageous, and the use of additional systems such as yeast artificial chromosomes (YAC vectors) becomes appropriate.