CHAPTER 4

Basic biology of plasmid and phage vectors

Plasmid biology and simple plasmid vectors

Plasmids are widely used as cloning vehicles but, before discussing their use in this context, it is appropriate to review some of their basic properties. Plasmids are replicons which are stably inherited in an extrachromosomal state. Most plasmids exist as double-stranded circular DNA molecules. If both strands of DNA are intact circles the molecules are described as covalently closed circles or CCC DNA (Fig. 4.1). If only one strand is intact, then the molecules are described as open circles or OC DNA. When isolated from cells, covalently closed circles often have a deficiency of turns in the double helix, such that they have a supercoiled configuration.

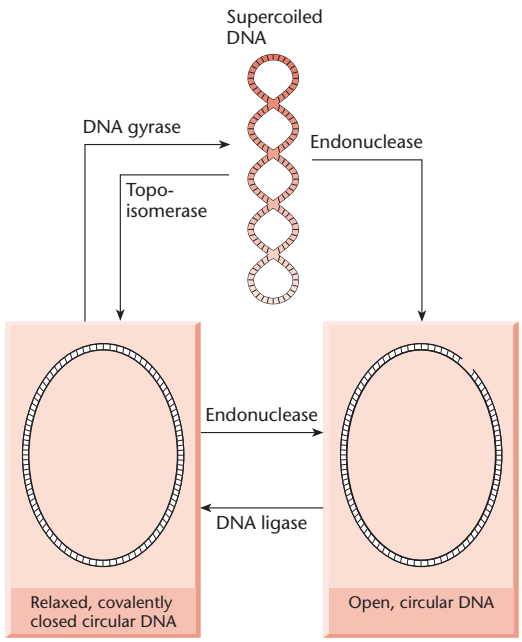


Fig. 4.1 The interconversion of supercoiled, relaxed covalently closed circular DNA and open circular DNA.

The enzymatic interconversion of supercoiled, relaxed CCC DNA* and OC DNA is shown in Fig. 4.1. Because of their different structural configurations, supercoiled and OC DNA separate upon electrophoresis in agarose gels (Fig. 4.2). Addition of an intercalating agent, such as ethidium bromide, to supercoiled DNA causes the plasmid to unwind. If excess ethidium bromide is added, the plasmid will rewind in the opposite direction (Fig. 4.3). Use of this fact is made in the isolation of plasmid DNA (see p. 48).

Not all plasmids exist as circular molecules. Linear plasmids have been found in a variety of bacteria, e.g. *Streptomyces* sp. and *Borrelia burgdorferi*. To prevent nuclease digestion, the ends of linear plasmids need to be protected and two general mechanisms have evolved. Either there are repeated sequences ending in a terminal DNA hairpin loop (*Borrelia*) or the ends are protected by covalent attachment of a protein (*Streptomyces*). For more details of linear plasmids the reader should consult Hinnebusch and Tilly (1993).

Plasmids are widely distributed throughout the prokaryotes, vary in size from less than 1×10^6 daltons to greater than 200×10^6, and are generally dispensable. Some of the phenotypes which these plasmids confer on their host cells are listed in Table 4.1. Plasmids to which phenotypic traits have not yet been ascribed are called *cryptic* plasmids.

Plasmids can be categorized into one of two major type – conjugative or non-conjugative – depending upon whether or not they carry a set of transfer genes, called the *tra* genes, which promote bacterial conjugation. Plasmids can also be categorized on the basis of their being maintained as multiple copies per

* The reader should not be confused by the terms *relaxed circle* and *relaxed plasmid*. Relaxed circles are CCC DNA that does not have a supercoiled configuration. Relaxed plasmids are plasmids with multiple copies per cell.

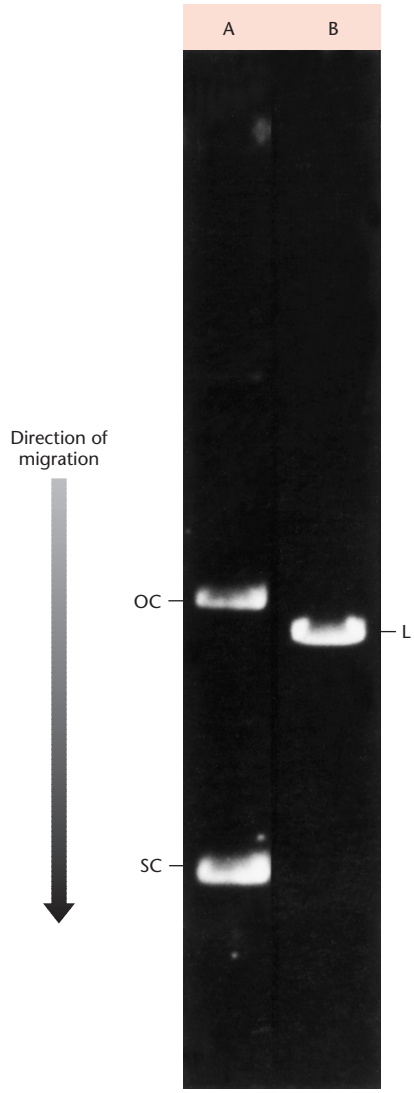


Fig. 4.2 Electrophoresis of DNA in agarose gels. The direction of migration is indicated by the arrow. DNA bands have been visualized by soaking the gel in a solution of ethidium bromide (which complexes with DNA by intercalating between stacked base pairs) and photographing the orange fluorescence which results upon ultraviolet irradiation. (A) Open circular (OC) and supercoiled (SC) forms of a plasmid of 6.4 kb pairs. Note that the compact supercoils migrate considerably faster than open circles (B). Linear plasmid (L) DNA is produced by treatment of the preparation shown in lane (A) with *Eco*RI, for which there is a single target site. Under the conditions of electrophoresis employed here, the linear form migrates just ahead of the open circular form.

cell (*relaxed* plasmids) or as a limited number of copies per cell (*stringent* plasmids). Generally, conjugative plasmids are of relatively high molecular weight and are present as one to three copies per chromosome, whereas non-conjugative plasmids are of low molecular weight and present as multiple copies per chromosome (Table 4.2). An exception is the conjugative plasmid R6K, which has a molecular weight of 25×10^6 daltons and is maintained as a relaxed plasmid.

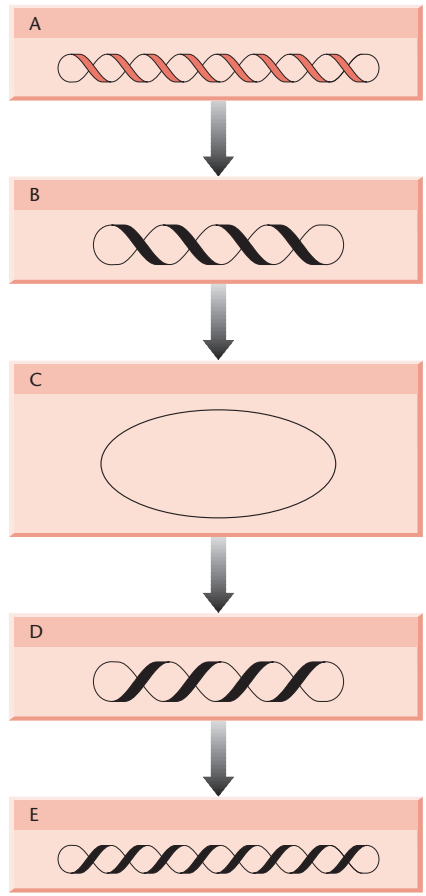


Fig. 4.3 Effect of intercalation of ethidium bromide on supercoiling of DNA. As the amount of intercalated ethidium bromide increases, the double helix untwists, with the result that the supercoiling decreases until the open form of the circular molecule is produced. Further intercalation introduces excess turns in the double helix, resulting in supercoiling in the opposite sense (note the direction of coiling at B and D). For clarity, only a single line represents the double helix.

Table 4.1 Some phenotypic traits exhibited by plasmid-carried genes.

Antibiotic resistance
Antibiotic production
Degradation of aromatic compounds
Haemolysin production
Sugar fermentation
Enterotoxin production
Heavy-metal resistance
Bacteriocin production
Induction of plant tumours
Hydrogen sulphide production
Host-controlled restriction and modification

Host range of plasmids

Plasmids encode only a few of the proteins required for their own replication and in many cases encode only one of them. All the other proteins required for replication, e.g. DNA polymerases, DNA ligase, helicases, etc., are provided by the host cell. Those replication proteins that are plasmid-encoded are located very close to the *ori* (origin of replication) sequences at which they act. Thus, only a small region surrounding the *ori* site is required for replication. Other parts of the plasmid can be deleted and foreign sequences can be added to the plasmid and replication will still occur. This feature of plasmids has greatly simplified the construction of versatile cloning vectors.

The host range of a plasmid is determined by its *ori* region. Plasmids whose *ori* region is derived from plasmid Col E1 have a restricted host range: they only replicate in enteric bacteria, such as *E. coli, Salmonella*, etc. Other *promiscuous* plasmids have a broad host range and these include RP4 and RSF1010. Plasmids of the RP4 type will replicate in most Gram-negative bacteria, to which they are readily transmitted by conjugation. Such promiscuous plasmids offer the potential of readily transferring cloned DNA molecules into a wide range of genetic backgrounds. Plasmids like RSF1010 are not conjugative but can be transformed into a wide range of Gram-negative and Gram-positive bacteria, where they are stably maintained. Many of the plasmids isolated from *Staphylococcus aureus* also have a broad host range and can replicate in many other Gram-positive bacteria. Plasmids with a broad host range encode most, if not all, of the proteins required for replication. They must also be able to express these genes and thus their promoters and ribosome binding sites must have evolved such that they can be recognized in a diversity of bacterial families.

Plasmid copy number

The copy number of a plasmid is determined by regulating the initiation of plasmid replication. Two major mechanisms of control of initiation have been recognized: regulation by antisense RNA and regulation by binding of essential proteins to iterons (for review, see Del Solar *et al.* 1998). Most of the cloning vectors in current use carry an *ori* region derived from plasmid Col E1 and copy-number control is mediated by antisense RNA. In this type of plasmid, the primer for DNA replication is a 555-base

Table 4.2 Properties of some conjugative and non-conjugative plasmids of Gram-negative organisms.

Plasmid	Size (MDa)	Conjugative	No. of plasmid copies/ chromosome equivalent	Phenotype
Col E1	4.2	No	10–15	Col E1 production
RSF1030	5.6	No	20–40	Ampicillin resistance
clo DF13	6	No	10	Cloacin production
R6K	25	Yes	13–38	Ampicillin and streptomycin resistance
F	62	Yes	1–2	–
RI	62.5	Yes	3–6	Multiple drug resistance
Ent P 307	65	Yes	1–3	Enterotoxin production

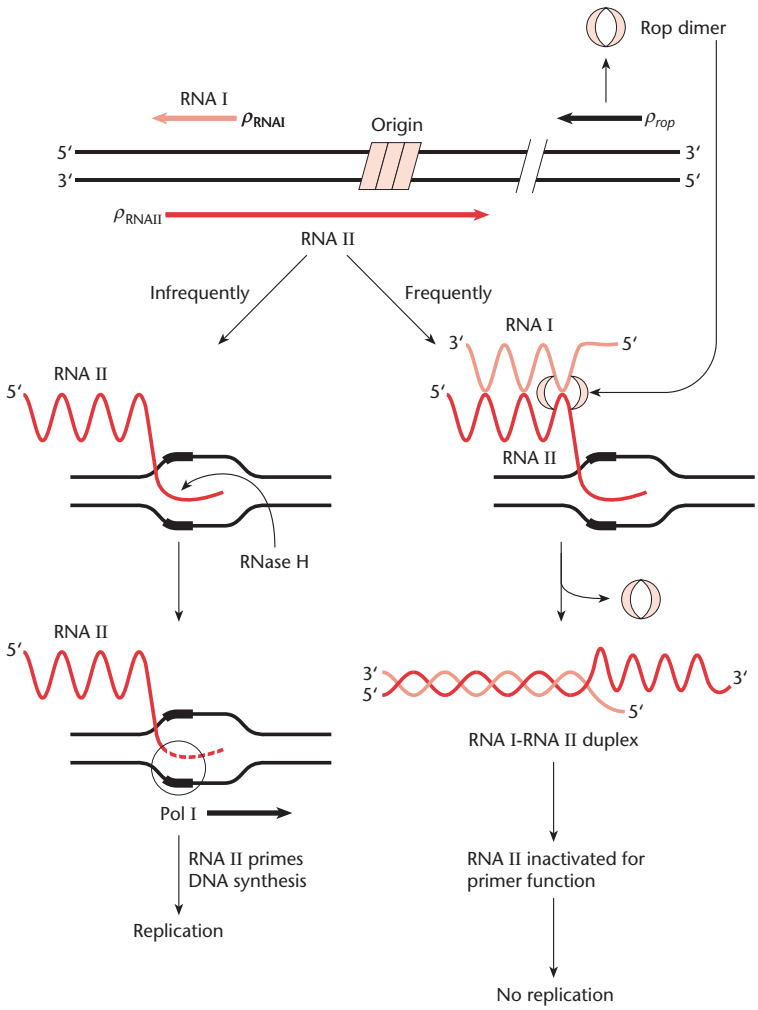


Fig. 4.4 Regulation of replication of Col E1-derived plasmids. RNA II must be processed by RNase H before it can prime replication. 'Origin' indicates the transition point between the RNA primer and DNA. Most of the time, RNA I binds to RNA II and inhibits the processing, thereby regulating the copy number. ρ_{RNAI} and $\rho_{RNA\,II}$ are the promoters for RNA I and RNA II transcription, respectively. RNA I is coloured pale red and RNA II dark red. The Rop protein dimer enhances the initial pairing of RNA I and RNA II.

ribonucleotide molecule called RNA II, which forms an RNA–DNA hybrid at the replication origin (Tomizawa & Itoh 1982), RNA II can only act as a primer if it is cleaved by RNase H to leave a free 3′ hydroxyl group. Unless RNA II is processed in this way, it will not function as a primer and replication will not ensue. Replication control is mediated by another small (108-base) RNA molecule called RNA I (Tomizawa & Itoh 1981), which is encoded by the same region of DNA as RNA II but by the complementary strand. Thus RNA I and RNA II are complementary to each other and can hybridize to form a double-stranded RNA helix. The formation of this duplex interferes with the processing of RNA II by RNase H and hence replication does not ensue (Fig. 4.4). Since RNA I is encoded by the plasmid, more of it will be synthesized when the copy number of the plasmid is high. As the host cell grows and divides, so the concentration of RNA I will fall and the plasmid will begin to replicate again (Cesarini *et al.* 1991, Eguchi *et al.* 1991).

In addition to RNA I, a plasmid-encoded protein called Rop helps maintain the copy number (Cesarini *et al.* 1982). This protein, which forms a dimer, enhances the pairing between RNA I and RNA II so that processing of the primer can be inhibited even at relatively low concentrations of RNA I. Deletion of the *ROP* gene (Twigg & Sherratt 1980) or mutations in RNA I (Muesing *et al.* 1981) result in increased copy numbers.

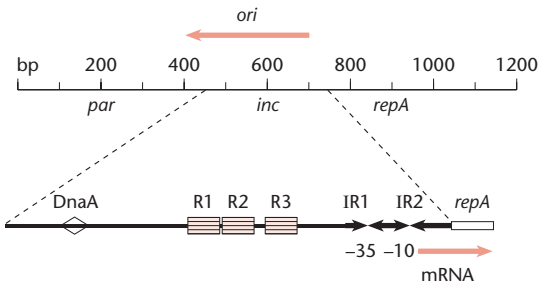


Fig. 4.5 The *ori* region of pSC101. R1, R2, and R3 are the three iteron sequences (CAAAGGTCTAGCAGCAGAATTTACAGA for R3) to which RepA binds to handcuff two plasmids. RepA autoregulates its own synthesis by binding to the inverted repeats IR1 and IR2. The location of the partitioning site *par* and the binding sites for the host protein DnaA are also shown.

In plasmid pSC101 and many of the broad-host-range plasmids, the *ori* region contains three to seven copies of an iteron sequence which is 17 to 22 bp long. Close to the *ori* region there is a gene, called *rep*A in pSC101, which encodes the RepA protein. This protein, which is the only plasmid-encoded protein required for replication, binds to the iterons and initiates DNA synthesis (Fig. 4.5).

Copy-number control is exerted by two superimposed mechanisms. First, the RepA protein represses its own synthesis by binding to its own promoter region and blocking transcription of its own gene (Ingmer & Cohen 1993). If the copy number is high, synthesis of RepA will be repressed. After cell division, the copy number and concentration of RepA will drop and replication will be initiated. Mutations in the RepA protein can lead to increased copy number (Ingmer & Cohen 1993, Cereghino *et al.* 1994). Secondly, the RepA protein can link two plasmids together, by binding to their iteron sequences, thereby preventing them from initiating replication. By this mechanism, known as *handcuffing* (McEachern *et al.* 1989), the replication of iteron plasmids will depend both on the concentration of RepA protein and on the concentration of the plasmids themselves.

Partitioning and segregative stability of plasmids

The loss of plasmids due to defective partitioning is called segregative instability. Naturally occurring plasmids are stably maintained because they contain a partitioning function, *par*, which ensures that they are stably maintained at each cell division. Such *par* regions are essential for stability of low-copy-number plasmids (for review, see Bingle & Thomas 2001). The higher-copy-number plasmid Col E1 also contains a *par* region but this is deleted in many Col E1-derived cloning vectors, e.g. pBR322. Although the copy number of vectors such as pBR322 is usually high, plasmid-free cells arise under nutrient limitation or other stress conditions (Jones *et al.* 1980, Nugent *et al.* 1983). The *par* region from a plasmid such as pSC101 can be cloned into pBR322, thereby stabilizing the plasmid (Primrose *et al.* 1983).

DNA superhelicity is involved in the partitioning mechanism (Miller *et al.* 1990). pSC101 derivatives lacking the *par* locus show decreased overall superhelical density as compared with wild-type pSC101. Partition-defective mutants of pSC101 and similar mutants of unrelated plasmids are stabilized in *Escherichia coli* by *top*A mutations, which increase negative DNA supercoiling. Conversely, DNA gyrase inhibitors and mutations in DNA gyrase increase the rate of loss of *par*-defective pSC101 derivatives.

Plasmid instability may also arise due to the formation of multimeric forms of a plasmid. The mechanism that controls the copy number of a plasmid ensures a fixed number of plasmid origins per bacterium. Cells containing multimeric plasmids have the same number of plasmid origins but fewer plasmid molecules, which leads to segregative instability if they lack a partitioning function. These multimeric forms are not seen with Col E1, which has a natural method of resolving multimers back to monomers. It contains a highly recombinogenic site (*cer*). If the *cer* sequence occurs more than once in a plasmid, as in a multimer, the host-cell Xer protein promotes recombination, thereby regenerating monomers (Summers & Sherratt 1984, Guhathakurta *et al.* 1996; for review, see Summers 1998).

Incompatibility of plasmids

Plasmid *incompatibility* is the inability of two different plasmids to coexist in the same cell in the absence of selection pressure. The term incompatibility can only be used when it is certain that entry of the second plasmid has taken place and that DNA

restriction is not involved. Groups of plasmids which are mutually incompatible are considered to belong to the same incompatibility (Inc) group. Over 30 incompatibility groups have been defined in *E. coli* and 13 for plasmids of *S. aureus*. Plasmids will be incompatible if they have the same mechanism of replication control. Not surprisingly, by changing the sequence of the RNA I/RNA II region of plasmids with antisense control of copy number, it is possible to change their incompatibility group. Alternatively, they will be incompatible if they share the same *par* region (Austin & Nordstrom 1990, Firsheim & Kim 1997).

The purification of plasmid DNA

An obvious prerequisite for cloning in plasmids is the purification of the plasmid DNA. Although a wide range of plasmid DNAs are now routinely purified, the methods used are not without their problems. Undoubtedly the trickiest stage is the lysis of the host cells; both incomplete lysis and total dissolution of the cells result in greatly reduced recoveries of plasmid DNA. The ideal situation occurs when each cell is just sufficiently broken to permit the plasmid DNA to escape without too much contaminating chromosomal DNA. Provided the lysis is done gently, most of the chromosomal DNA released will be of high molecular weight and can be removed, along with cell debris, by high-speed centrifugation to yield a *cleared lysate*. The production of satisfactory cleared lysates from bacteria other than *E. coli*, particularly if large plasmids are to be isolated, is frequently a combination of skill, luck and patience.

Many methods are available for isolating pure plasmid DNA from cleared lysates but only two will be described here. The first of these is the 'classical' method and is due to Vinograd (Radloff *et al.* 1967). This method involves isopycnic centrifugation of cleared lysates in a solution of CsCl containing ethidium bromide (EtBr). EtBr binds by intercalating between the DNA base pairs, and in so doing causes the DNA to unwind. A CCC DNA molecule, such as a plasmid, has no free ends and can only unwind to a limited extent, thus limiting the amount of EtBr bound. A linear DNA molecule, such as fragmented chromosomal DNA, has no such topological constraints and can therefore bind more of the EtBr molecules. Because the density of the DNA–EtBr complex decreases as more EtBr is bound, and because more EtBr can be bound to a linear molecule than to a covalent circle, the covalent circle has a higher density at saturating concentrations of EtBr. Thus covalent circles (i.e. plasmids) can be separated from linear chromosomal DNA (Fig. 4.6).

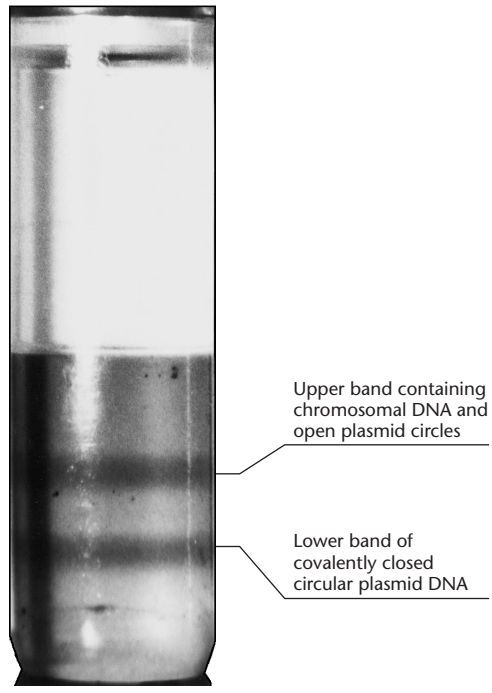


Fig. 4.6 Purification of Col E1 *Kan*^{R} plasmid DNA by isopycnic centrifugation in a CsCl–EtBr gradient. (Photograph by courtesy of Dr G. Birnie.)

Currently the most popular method of extracting and purifying plasmid DNA is that of Birnboim and Doly (1979). This method makes use of the observation that there is a narrow range of pH (12.0–12.5) within which denaturation of linear DNA, but not covalently closed circular DNA, occurs. Plasmid-containing cells are treated with lysozyme to weaken the cell wall and then lysed with sodium hydroxide and sodium dodecyl sulphate (SDS). Chromosomal DNA remains in a high-molecular-weight form but is denatured. Upon neutralization with acidic sodium acetate, the chromosomal DNA renatures and aggregates to form an insoluble network. Simultaneously,

the high concentration of sodium acetate causes precipitation of protein–SDS complexes and of high-molecular-weight RNA. Provided the pH of the alkaline denaturation step has been carefully controlled, the CCC plasmid DNA molecules will remain in a native state and in solution, while the contaminating macromolecules co-precipitate. The precipitate can be removed by centrifugation and the plasmid concentrated by ethanol precipitation. If necessary, the plasmid DNA can be purified further by gel filtration.

Recently a number of commercial suppliers of convenience molecular-biology products have developed kits to improve the yield and purity of plasmid DNA. All of them take advantage of the benefits of alkaline lysis and have as their starting-point the cleared lysate. The plasmid DNA is selectively bound to an ion-exchange material, prepacked in columns or tubes, in the presence of a chaotropic agent (e.g. guanidinium hydrochloride). After washing away the contaminants, the purified plasmid is eluted in a small volume.

The yield of plasmid is affected by a number of factors. The first of these is the actual copy number inside the cells at the time of harvest. The copy-number control systems described earlier are not the only factors affecting yield. The copy number is also affected by the growth medium, the stage of growth and the genotype of the host cell (Nugent *et al.* 1983, Seelke *et al.* 1987, Duttweiler & Gross 1998). The second and most important factor is the care taken in making the cleared lysate. Unfortunately, the commercially available kits have not removed the vagaries of this procedure. Finally, the presence in the host cell of a wild-type *end*A gene can affect the recovery of plasmid. The product of the *end*A gene is endonuclease I, a periplasmic protein whose substrate is double-stranded DNA. The function of endonuclease I is not fully understood. Strains bearing *end*A mutations have no obvious phenotype other than improved stability and yield of plasmid obtained from them.

Although most cloning vehicles are of low molecular weight (see next section), it is sometimes necessary to use the much larger conjugative plasmids. Although these high-molecular-weight plasmids can be isolated by the methods just described, the yields are often very low. Either there is inefficient release of the plasmids from the cells as a consequence of their size or there is physical destruction caused by shear forces during the various manipulative steps. A number of alternative procedures have been described (Gowland & Hardmann 1986), many of which are a variation on that of Eckhardt (1978). Bacteria are suspended in a mixture of Ficoll and lysozyme and this results in a weakening of the cell walls. The samples are then placed in the slots of an agarose gel, where the cells are lysed by the addition of detergent. The plasmids are subsequently extracted from the gel following electrophoresis. The use of agarose, which melts at low temperature, facilitates extraction of the plasmid from the gel.

Desirable properties of plasmid cloning vehicles

An ideal cloning vehicle would have the following three properties:

- low molecular weight;
- ability to confer readily selectable phenotypic traits on host cells;
- single sites for a large number of restriction endonucleases, preferably in genes with a readily scorable phenotype.

The advantages of a low molecular weight are several. First, the plasmid is much easier to handle, i.e. it is more resistant to damage by shearing, and is readily isolated from host cells. Secondly, low-molecular-weight plasmids are usually present as multiple copies (see Table 4.2), and this not only facilitates their isolation but leads to gene dosage effects for all cloned genes. Finally, with a low molecular weight there is less chance that the vector will have multiple substrate sites for any restriction endonuclease (see below).

After a piece of foreign DNA is inserted into a vector, the resulting chimeric molecules have to be transformed into a suitable recipient. Since the efficiency of transformation is so low, it is essential that the chimeras have some readily scorable phenotype. Usually this results from some gene, e.g. antibiotic resistance, carried on the vector, but could also be produced by a gene carried on the inserted DNA.

One of the first steps in cloning is to cut the vector DNA and the DNA to be inserted with either the

same endonuclease or ones producing the same ends. If the vector has more than one site for the endonuclease, more than one fragment will be produced. When the two samples of cleaved DNA are subsequently mixed and ligated, the resulting chimeras will, in all probability, lack one of the vector fragments. It is advantageous if insertion of foreign DNA at endonuclease-sensitive sites inactivates a gene whose phenotype is readily scorable, for in this way it is possible to distinguish chimeras from cleaved plasmid molecules which have self-annealed. Of course, readily detectable insertional inactivation is not essential if the vector and insert are to be joined by the homopolymer tailing method (see p. 40) or if the insert confers a new phenotype on host cells.

pBR322, a purpose-built cloning vehicle

In early cloning experiments, the cloning vectors used were natural plasmids, such as Col E1 and pSC101. While these plasmids are small and have single sites for the common restriction endonucleases, they have limited genetic markers for selecting transformants. For this reason, considerable effort was expended on constructing, *in vitro*, superior cloning vectors. The best, and most widely used of these early purpose-built vectors is pBR322. Plasmid pBR322 contains the Ap^R and Tc^R genes of RSF2124 and pSC101, respectively, combined with replication elements of pMB1, a Col E1-like plasmid (Fig. 4.7a). The origins of pBR322 and its progenitor, pBR313, are shown in Fig. 4.7b, and details of its construction can be found in the papers of Bolivar *et al.* (1977a,b).

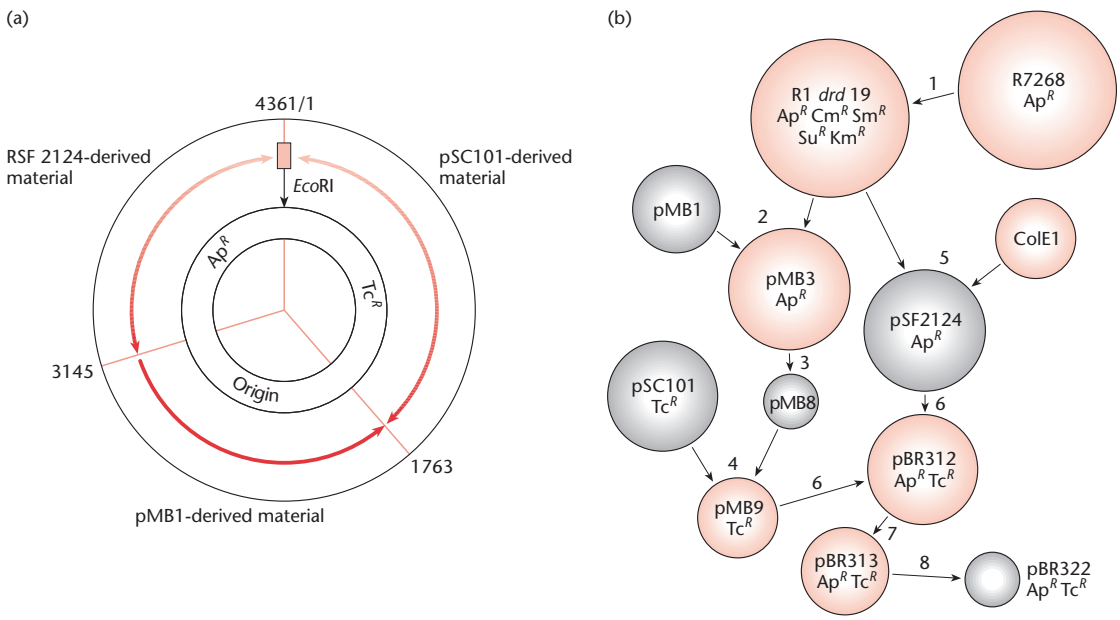


Fig. 4.7 The origins of plasmid pBR322. (a) The boundaries between the pSC101, pMB1 and RSF2124-derived material. The numbers indicate the positions of the junctions in base pairs from the unique *Eco*RI site. (b) The molecular origins of plasmid pBR322. R7268 was isolated in London in 1963 and later renamed R1. 1, A variant, R1*drd*19, which was derepressed for mating transfer, was isolated. 2, The Ap^R transposon, Tn*3*, from this plasmid was transposed on to pMB1 to form pMB3. 3, This plasmid was reduced in size by *Eco*RI* rearrangement to form a tiny plasmid, pMB8, which carries only colicin immunity. 4, *Eco*RI* fragments from pSC101 were combined with pMB8 opened at its unique *Eco*RI site and the resulting chimeric molecule rearranged by *Eco*RI* activity to generate pMB9. 5, In a separate event, the Tn*3* of R1*drd*19 was transposed to Col E1 to form pSF2124. 6, The Tn*3* element was then transposed to pMB9 to form pBR312. 7, *Eco*RI* rearrangement of pBR312 led to the formation of pBR313, from which (8) two separate fragments were isolated and ligated together to form pBR322. During this series of constructions, R1 and Col E1 served only as carries for Tn*3*. (Reproduced by courtesy of Dr G. Sutcliffe and Cold Spring Harbor Laboratory.)

Plasmid pBR322 has been completely sequenced. The original published sequence (Sutcliffe 1979) was 4362 bp long. Position 0 of the sequence was arbitrarily set between the A and T residues of the *Eco*RI recognition sequence (GAATTC). The sequence was revised by the inclusion of an additional CG base pair at position 526, thus increasing the size of the plasmid to 4363 bp (Backman & Boyer 1983, Peden 1983). More recently, Watson (1988) has revised the size yet again, this time to 4361 bp, by eliminating base pairs at coordinates 1893 and 1915. The most useful aspect of the DNA sequence is that it totally characterizes pBR322 in terms of its restriction sites, such that the exact length of every fragment can be calculated. These fragments can serve as DNA markers for sizing any other DNA fragment in the range of several base pairs up to the entire length of the plasmid.

There are over 40 enzymes with unique cleavage sites on the pBR322 genome (Fig. 4.8). The target sites of 11 of these enzymes lie within the Tc^R gene, and there are sites for a further two (*Cla*I and *Hin*dIII) within the promoter of that gene. There are unique sites for six enzymes within the Ap^R gene. Thus, cloning in pBR322 with the aid of any one of those 19 enzymes will result in insertional inactivation of either the Ap^R or the Tc^R markers. However, cloning in the other unique sites does not permit the easy selection of recombinants, because neither of the antibiotic resistance determinants is inactivated.

Following manipulation *in vitro, E. coli* cells transformed with plasmids with inserts in the Tc^R gene can be distinguished from those cells transformed with recircularized vector. The former are Ap^R and Tc^S, whereas the latter are both Ap^R and Tc^R. In practice, transformants are selected on the basis of their Ap resistance and then replica-plated on to Tc-containing media to identify those that are Tc^S. Cells transformed with pBR322 derivatives carrying inserts in the Ap^R gene can be identified more readily (Boyko & Ganschow 1982). Detection is based upon the ability of the β-lactamase produced by Ap^R cells to convert penicillin to penicilloic acid, which in turn binds iodine. Transformants are selected on rich medium containing soluble starch and Tc. When colonized plates are flooded with an indicator solution of iodine and penicillin, β-lactamase-producing (Ap^R) colonies clear the indicator solution whereas Ap^S colonies do not.

The *Pst*I site in the Ap^R gene is particularly useful, because the 3′ tetranucleotide extensions formed on digestion are ideal substrates for terminal transferase. Thus this site is excellent for cloning by the homopolymer tailing method described in the previous chapter (see p. 40). If oligo(dG.dC) tailing is used, the *Pst*I site is regenerated (see Fig. 3.11) and the insert may be cut out with that enzyme.

Plasmid pBR322 has been a widely used cloning vehicle. In addition, it has been widely used as a model system for the study of prokaryotic transcription and translation, as well as investigation of the effects of topological changes on DNA conformation. The popularity of pBR322 is a direct result of the availability of an extensive body of information on its structure and function. This in turn is increased with each new study. The reader wishing more detail on the structural features, transcriptional signals, replication, amplification, stability and conjugal mobility of pBR322 should consult the review of Balbás *et al.* (1986).

Example of the use of plasmid pBR322 as a vector: isolation of DNA fragments which carry promoters

Cloning into the *Hin*dIII site of pBR322 generally results in loss of tetracycline resistance. However, in some recombinants, Tc^R is retained or even increased. This is because the *Hin*dIII site lies within the promoter rather than the coding sequence. Thus whether or not insertional inactivation occurs depends on whether the cloned DNA carries a promoter-like sequence able to initiate transcription of the Tc^R gene. Widera *et al.* (1978) have used this technique to search for promoter-containing fragments.

Four structural domains can be recognized within *E. coli* promoters. These are:

- position 1, the purine initiation nucleotide from which RNA synthesis begins;
- position −6 to −12, the Pribnow box;
- the region around base pair −35;
- the sequence between base pairs −12 and −35.

Although the *Hin*dIII site lies within the Pribnow box (Rodriguez *et al.* 1979) the box is re-created on insertion of a foreign DNA fragment. Thus when insertional inactivation occurs it must be the region from −13 to −40 which is modified.

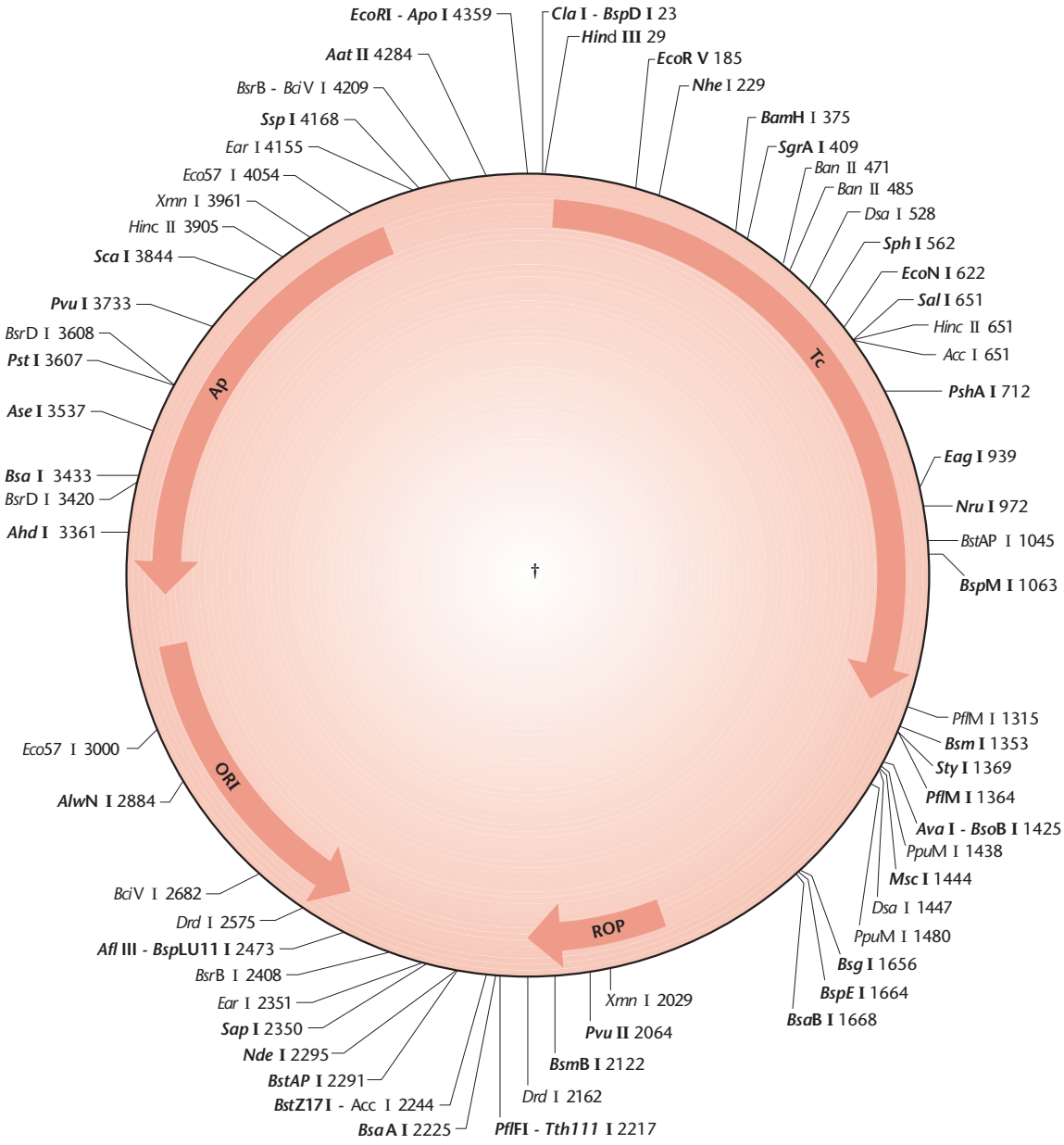


Fig. 4.8 Restriction map of plasmid pBR322 showing the location and direction of transcription of the ampicillin (Ap) and tetracycline (Tc) resistance loci, the origin of replication (*ori*) and the Col E1-derived Rop gene. The map shows the restriction sites of those enzymes that cut the molecule once or twice. The unique sites are shown in bold type. The coordinates refer to the position of the 5′ base in each recognition sequence with the first T in the *Eco*RI site being designated as nucleotide number 1. The exact positions of the loci are: Tc, 86–1268; Ap, 4084–3296; Rop, 1918–2105 and the origin of replication, 2535.

Improved vectors derived from pBR322

Over the years, numerous different derivatives of pBR322 have been constructed, many to fulfil special-purpose cloning needs. A compilation of the properties of some of these plasmids has been provided by Balbás *et al.* (1986).

Much of the early work on the improvement of pBR322 centred on the insertion of additional unique restriction sites and selectable markers, e.g. pBR325 encodes chloramphenicol resistance in addition to ampicillin and tetracycline resistance and has a unique *Eco*RI site in the Cm^R gene. Initially, each new vector was constructed in a series of steps analogous to those used in the generation of pBR322 itself (Fig. 4.7). Then the construction of improved vectors was simplified (Vieira & Messing 1982, 1987, Yanisch-Perron *et al.* 1985) by the use of *polylinkers* or *multiple cloning sites* (MCS), as exemplified by the pUC vectors (Fig. 4.9). An MCS is a short DNA sequence, 2.8 kb in the case of pUC19, carrying sites for many different restriction endonucleases. An MCS increases the number of potential cloning strategies available by extending the range of enzymes that can be used to generate a restriction fragment suitable for cloning. By combining them within an MCS, the sites are made contiguous, so that any two sites within it can be cleaved simultaneously without excising vector sequences.

The pUC vectors also incorporate a DNA sequence that permits rapid visual detection of an insert. The MCS is inserted into the *lacZ′* sequence, which encodes the promoter and the α-peptide of β-galactosidase. The insertion of the MCS into the lacZ′ fragment does not affect the ability of the α-peptide to mediate complementation, but cloning DNA fragments into the MCS does. Therefore, recombinants can be detected by blue/white screening on growth medium containing Xgal (see Box 3.2 on p. 35). The usual site for insertion of the MCS is between the iniator ATG codon and codon 7, a region that encodes a functionally non-essential part of the α-complementation peptide. Recently, Slilaty and Lebel (1998) have reported that blue/white colour selection can be variable. They have found that reliable inactivation of complementation occurs only when the insert is made between codons 11 and 36.

Bacteriophage λ

Essential features

Bacteriophage λ is a genetically complex but very extensively studied virus of *E. coli* (Box 4.1). Because it has been the object of so much molecular-genetical research, it was natural that, right from the beginnings of gene manipulation, it should have been investigated and developed as a vector. The DNA of phage λ, in the form in which it is isolated from the phage particle, is a linear duplex molecule of about 48.5 kbp. The entire DNA sequence has been determined (Sanger *et al.* 1982). At each end are short single-stranded 5′ projections of 12 nucleotides, which are complementary in sequence and by which the DNA adopts a circular structure when it is injected into its host cell, i.e. λ DNA naturally has cohesive termini, which associate to form the *cos* site.

Functionally related genes of phage λ are clustered together on the map, except for the two positive regulatory genes *N* and *Q*. Genes on the left of the conventional linear map (Fig. 4.10) code for head and tail proteins of the phage particle. Genes of the central region are concerned with recombination (e.g. *red*) and the process of lysogenization, in which the circularized chromosome is inserted into its host chromosome and stably replicated along with it as a prophage. Much of this central region, including these genes, is not essential for phage growth and can be deleted or replaced without seriously impairing the infectious growth cycle. Its dispensability is crucially important, as will become apparent later, in the construction of vector derivatives of the phage. To the right of the central region are genes concerned with regulation and prophage immunity to superinfection (*N*, *cro*, *c*I), followed by DNA synthesis (*O*, *P*), late function regulation (*Q*) and host cell lysis (*S*, *R*). Figure 4.11 illustrates the λ life cycle.

Promoters and control circuits

As we shall see, it is possible to insert foreign DNA into the chromosome of phage-λ derivative and, in some cases, foreign genes can be expressed efficiently via λ promoters. We must therefore briefly

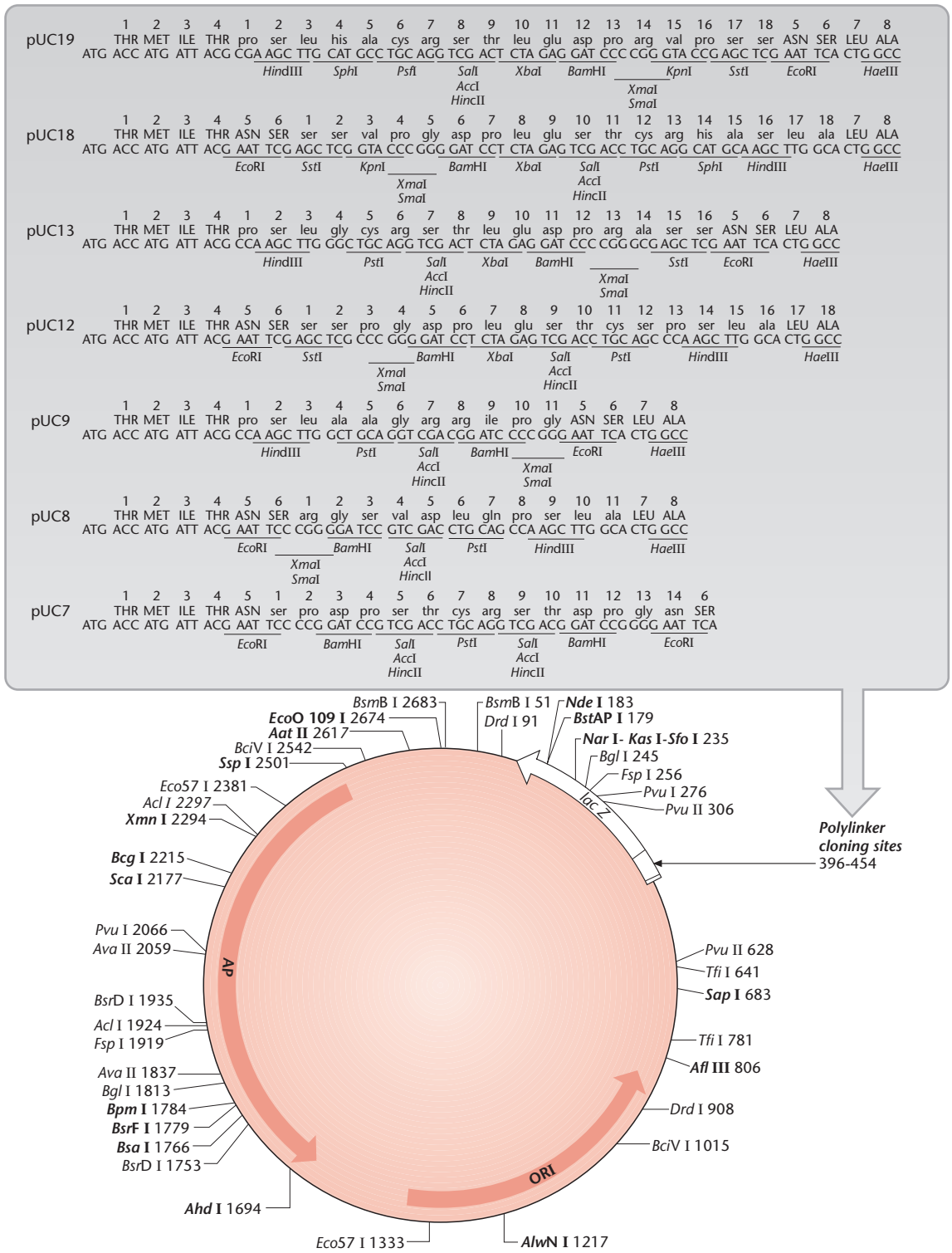
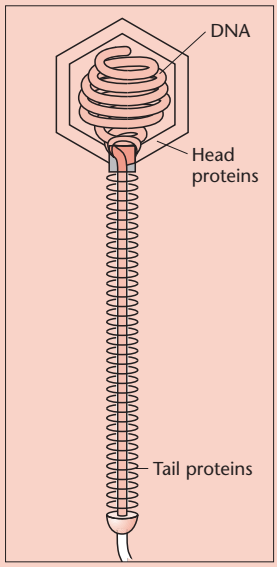


Fig. 4.9 Genetic maps of some pUC plasmids. The multiple cloning site (MCS) is inserted into the *lacZ* gene but does not interfere with gene function. The additional codons present in the *lacZ* gene as a result of the polylinker are labelled with lower-case letters. These polylinker regions (MCS) are identical to those of the M13 mp series of vectors (see p. 63).

Box 4.1 Bacteriophage λ: its important place in molecular biology and recombinant DNA technology



In the early 1950s, following some initial studies on *Bacillus megaterium*, André Lwoff and his colleagues at the Institut Pasteur described the phenomenon of lysogeny in *E. coli.* It became clear that certain strains of *E. coli* were lysogenized by phage, that is to say, these bacteria harboured phage λ in a dormant form, called a prophage. The lysogenic bacteria grew normally and might easily not have been recognized as lysogenic. However, when Lwoff exposed the bacteria to a moderate dose of ultraviolet light, the bacteria stopped growing, and after about 90 min of incubation the bacteria lysed, releasing a crop of phage into the medium.

The released phage were incapable of infecting more *E. coli* than had been lysogenized by phage λ (this is called immunity to superinfection), but non-lysogenic bacteria could be infected to yield another crop of virus. Not every non-lysogenic bacterium yielded virus; some bacteria were converted into lysogens because the phage switched to the dormant lifestyle – becoming prophage – rather than causing a lytic infection.

By the mid-1950s it was realized that the prophage consisted of a phage λ genome that had become integrated into the *E. coli* chromosome. It was also apparent to Lwoff's colleagues, Jacob and Monod, that the switching between the two states of the virus – the lytic and lysogenic lifestyles – was an example of a fundamental aspect of genetics that was gaining increasing attention, gene regulation.

Intensive genetic and molecular biological analysis of the phage, mainly in the 1960s and 1970s, led to a good understanding of the virus. The key molecule in maintaining the dormancy of the prophage and in conferring immunity to superinfection is the phage repressor, which is the product of the phage *cI* gene. In 1967 the phage repressor was isolated by Mark Ptashne (Ptashne 1967a,b). The advanced molecular genetics of the phage made it a good candidate for development as a vector, beginning in the 1970s and continuing to the present day, as described in the text. The development of vectors exploited the fact that a considerable portion of the phage genome encodes functions that are not needed for the infectious cycle. The ability to package recombinant phage DNA into virus particles *in vitro* was an important development for library construction (Hohn & Murray 1977).

A landmark in molecular biology was reached when the entire sequence of the phage λ genome, 48 502 nucleotide pairs, was determined by Fred Sanger and his colleagues (Sanger *et al.* 1982).

consider the promoters and control circuits affecting λ gene expression (see Ptashne (1992) for an excellent monograph on phage-λ control circuits).

In the lytic cycle, λ transcription occurs in three temporal stages: early, middle and late. Basically, early gene transcription establishes the lytic cycle (in competition with lysogeny), middle gene products replicate and recombine the DNA and late gene products package this DNA into mature phage particles. Following infection of a sensitive host, early transcription proceeds from major promoters situated immediately to the left (P_L) and right (P_R) of

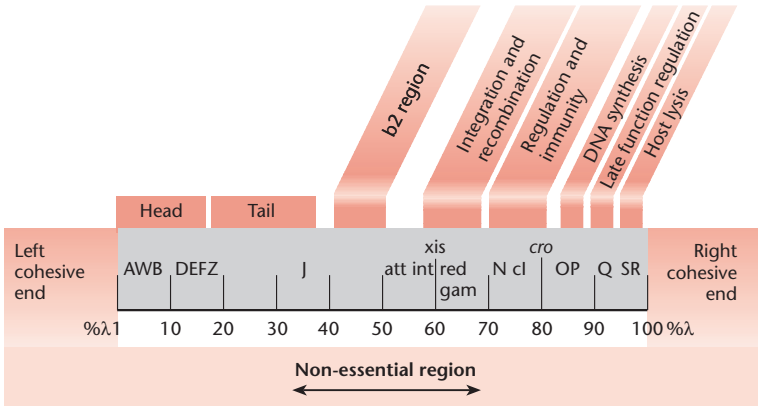


Fig. 4.10 Map of the λ chromosome, showing the physical position of some genes on the full-length DNA of wild-type bacteriophage λ. Clusters of functionally related genes are indicated.

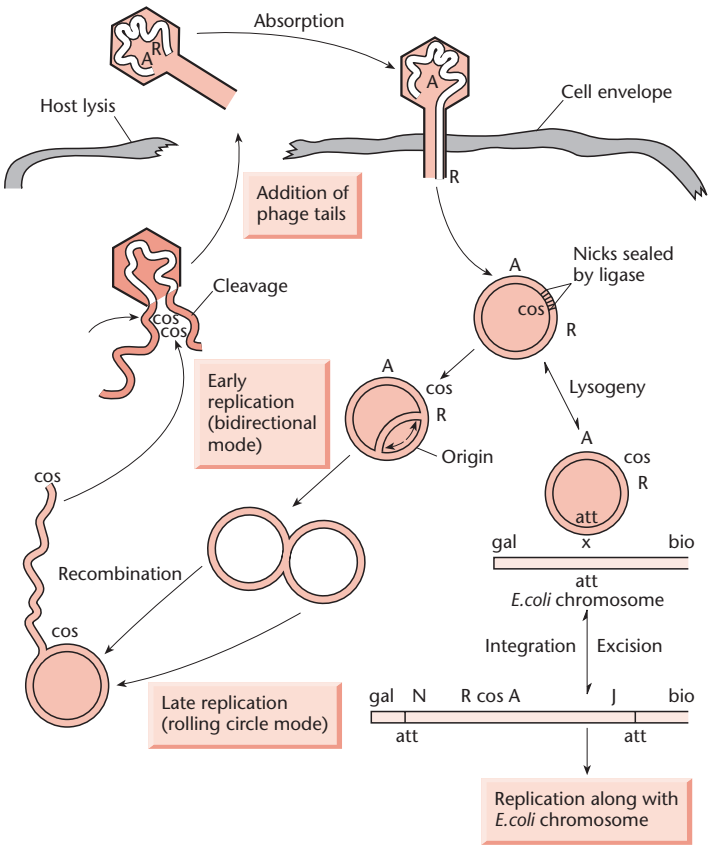


Fig. 4.11 Replication of phage-λ DNA in lytic and lysogenic cycles.

the repressor gene (*cI*) (Fig. 4.12). This transcription is subject to repression by the product of the *cI* gene and in a lysogen this repression is the basis of immunity to superinfecting λ. Early in infection, transcripts from P_L and P_R stop at termination sites t_L and t_{R1}. The site t_{R2} stops any transcripts that escape beyond t_{R1}. Lambda switches from early- to middle-stage transcription by anti-termination. The *N* gene product, expressed from P_L, directs this switch. It interacts with RNA polymerase and, antagonizing the action of host termination protein ρ, permits it to ignore the stop signals so that P_L and P_R transcripts extend into genes such as *red*, *O* and *P* necessary for the middle stage. The early and middle transcripts

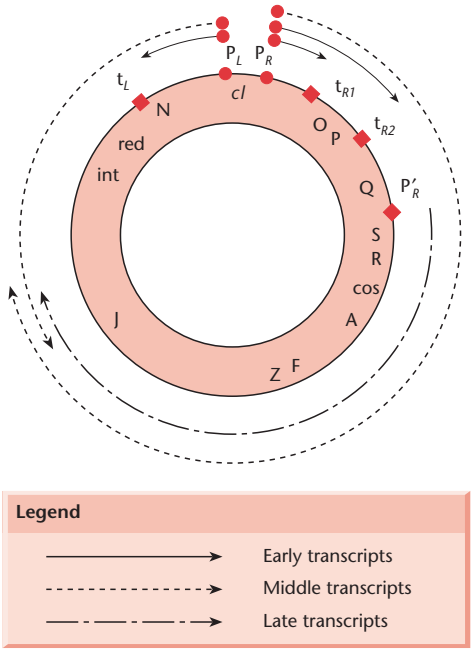


Fig. 4.12 Major promoters and transcriptional termination sites of phage λ. (See text for details.)

and patterns of expression therefore overlap. The *cro* product, when sufficient has accumulated, prevents transcription from P_L and P_R. The gene *Q* is expressed from the distal portion of the extended P_R transcript and is responsible for the middle-to-late switch. This also operates by anti-termination. The *Q* product specifically anti-terminates the short P_R transcript, extending it into the late genes, across the cohered *cos* region, so that many mature phage particles are ultimately produced.

Both *N* and *Q* play positive regulatory roles essential for phage growth and plaque formation; but an N^- phage can produce a small plaque if the termination site t_{R2} is removed by a small deletion termed *nin* (*N*-independent) as in λN^- *nin*.

Vector DNA

Wild-type λ DNA contains several target sites for most of the commonly used restriction endonucleases and so is not itself suitable as a vector. Derivatives of the wild-type phage have therefore been produced that either have a single target site at which foreign DNA can be inserted (*insertional* vectors) or have a pair of sites defining a fragment that can be removed (*stuffer*) and replaced by foreign DNA (*replacement* vectors). Since phage λ can accommodate only about 5% more than its normal complement of DNA, vector derivatives are constructed with deletions to increase the space within the genome. The shortest λ DNA molecules that produce plaques of nearly normal size are 25% deleted. Apparently, if too much non-essential DNA is deleted from the genome, it cannot be packaged into phage particles efficiently. This can be turned to advantage for, if the replaceable fragment of a replacement-type vector is either removed by physical separation or effectively destroyed by treatment with a second restriction endonuclease that cuts it alone, then the deleted vector genome can give rise to plaques only if a new DNA segment is inserted into it. This amounts to positive selection for recombinant phage carrying foreign DNA.

Many vector derivatives of both the insertional and replacement types were produced by several groups of researchers early in the development of recombinant DNA technology (e.g. Thomas *et al.* 1974, Murray & Murray 1975, Blattner *et al.* 1977, Leder *et al.* 1977). Most of these vectors were constructed for use with *Eco*RI, *Bam*HI or *Hin*dIII, but their application could be extended to other endonucleases by the use of linker molecules. These early vectors have been largely superseded by improved vectors for rapid and efficient genomic or complementary DNA (cDNA) library construction (see Chapter 6).

Improved phage-λ vectors

As with plasmid vectors, improved phage-vector derivatives have been developed. There have been several aims, among which are the following.

- To increase the capacity for foreign DNA fragments, preferably for fragments generated by any one of several restriction enzymes (reviewed by Murray 1983).
- To devise methods for positively selecting recombinant formation.
- To allow RNA probes to be conveniently prepared by transcription of the foreign DNA insert; this facilitates the screening of libraries in chromosome walking procedures. An example of a vector with this property is λZAP (see p. 93).
- To develop vectors for the insertion of eukaryotic cDNA (p. 93) such that expression of the cDNA, in

the form of a fusion polypeptide with β-galactosidase, is driven in *E. coli*; this form of expression vector is useful in antibody screening. An example of such a vector is λgt11.

The first two points will be discussed here. The discussion of improved vectors in library construction and screening is deferred until Chapter 6.

The maximum capacity of phage-λ derivatives can only be attained with vectors of the replacement type, so that there has also been an accompanying incentive to devise methods for positively selecting recombinant formation without the need for prior removal of the stuffer fragment. Even when steps are taken to remove the stuffer fragment by physical purification of vector arms, small contaminating amounts may remain, so that genetic selection for recombinant formation remains desirable. The usual method of achieving this is to exploit the Spi$^-$ phenotype.

Wild-type λ cannot grow on *E. coli* strains lysogenic for phage P2; in other words, the λ phage is Spi$^+$ (sensitive to P2 inhibition). It has been shown that the products of λ genes *red* and *gam*, which lie in the region 64–69% on the physical map, are responsible for the inhibition of growth in a P2 lysogen (Herskowitz 1974, Sprague *et al.* 1978, Murray 1983). Hence vectors have been derived in which the stuffer fragment includes the region 64–69%, so that recombinants in which this has been replaced by foreign DNA are phenotypically Spi$^-$ and can be positively selected by plating on a P2 lysogen (Karn *et al.* 1986, Loenen & Brammar 1980).

Deletion of the *gam* gene has other consequences. The *gam* product is necessary for the normal switch in λ DNA replication from the bidirectional mode to the rolling-circle mode (see Fig. 4.11). Gam$^-$ phage cannot generate the concatemeric linear DNA which is normally the substrate for packaging into phage heads. However, *gam*$^-$ phage do form plaques because the *rec* and *red* recombination systems act on circular DNA molecules to form multimers, which can be packaged. *gam*$^-$ *red*$^-$ phage are totally dependent upon *rec*-mediated exchange for plaque formation on *rec*$^+$ bacteria. λ DNA is a poor substrate for this *rec*-mediated exchange. Therefore, such phage make vanishingly small plaques unless they contain one or more short DNA sequences called *chi* (crossover hot-spot instigator) sites, which stimulate *rec*-mediated exchange. Many of the current replacement vectors generate *red*$^-$ *gam*$^-$ clones and

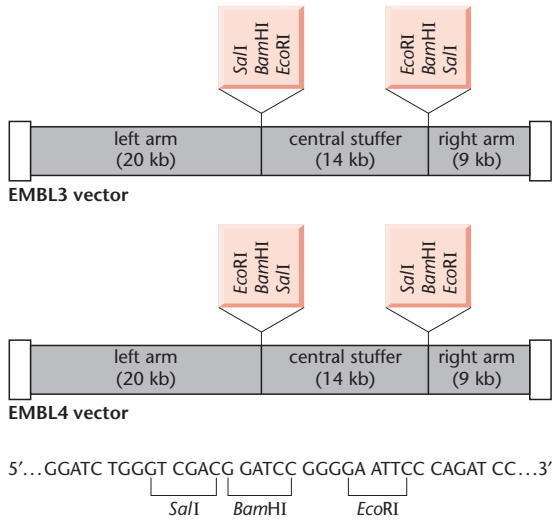


Fig. 4.13 The structure of bacteriophage λ cloning vectors EMBL3 and EMBL4. The polylinker sequence is present in opposite orientation in the two vectors.

so have been constructed with a *chi* site within the non-replaceable part of the phage.

The most recent generation of λ vectors, which are based on EMBL3 and EMBL4 (Frischauf *et al.* 1983; Fig. 4.13), have a capacity for DNA of size 9–23 kb. As well as being *chi*$^+$, they have polylinkers flanking the stuffer fragment to facilitate library construction. Phages with inserts can be selected on the basis of their Spi$^-$ phenotype, but there is an alternative. The vector can be digested with *Bam*HI and *Eco*RI prior to ligation with foreign DNA fragments produced with *Bam*HI. If the small *Bam*HI–*Eco*RI fragments from the polylinkers are removed, the stuffer fragment will not be reincorporated.

Packaging phage-λ DNA *in vitro*

So far, we have considered only one way of introducing manipulated phage DNA into the host bacterium, i.e. by transfection of competent bacteria (see Chapter 2). Using freshly prepared λ DNA that has not been subjected to any gene-manipulation procedures, transfection will result typically in about 10^5 plaques/μg of DNA. In a gene-manipulation experiment in which the vector DNA is restricted and then ligated with foreign DNA, this figure is reduced to about 10^4–10^3 plaques/μg of vector DNA. Even with perfectly efficient nucleic acid

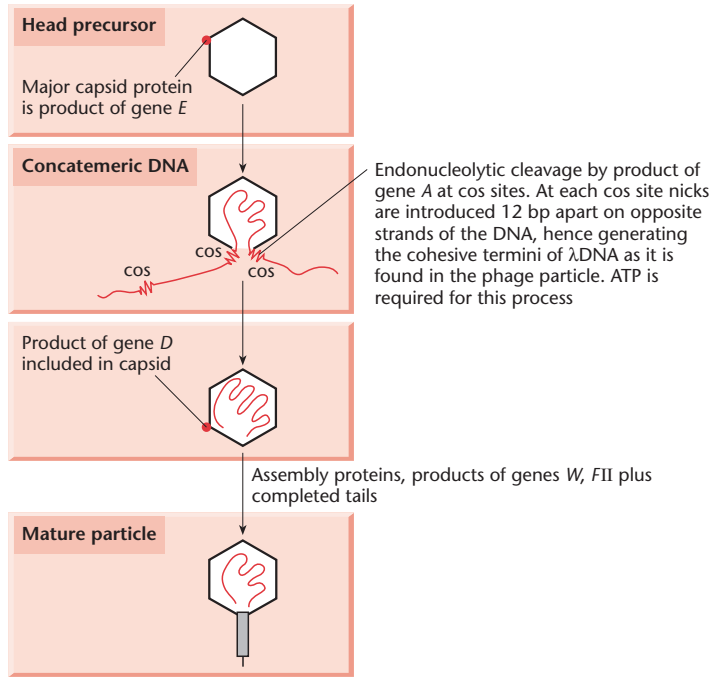


Fig. 4.14 Simplified scheme showing packaging of phage-λ DNA into phage particles.

biochemistry, some of this reduction is inevitable. It is a consequence of the random association of fragments in the ligation reaction, which produces molecules with a variety of fragment combinations, many of which are inviable. Yet, in some contexts, 10^6 or more recombinants are required. The scale of such experiments can be kept within a reasonable limit by packaging the recombinant DNA into mature phage particles *in vitro*.

Placing the recombinant DNA in a phage coat allows it to be introduced into the host bacteria by the normal processes of phage infection, i.e. phage adsorption followed by DNA injection. Depending upon the details of the experimental design, packaging *in vitro* yields about 10^6 plaques/μg of vector DNA after the ligation reaction.

Figure 4.14 shows some of the events occurring during the packaging process that take place within the host during normal phage growth and which we now require to perform *in vitro*. Phage DNA in concatemeric form, produced by a rolling-circle replication mechanism (see Fig. 4.11), is the substrate for the packaging reaction. In the presence of phage head precursor (the product of gene *E* is the major capsid protein) and the product of gene *A*, the concatemeric DNA is cleaved into monomers and encapsidated. Nicks are introduced in opposite strands of the DNA, 12 nucleotide pairs apart at each *cos* site, to produce the linear monomer with its cohesive termini. The product of gene *D* is then incorporated into what now becomes a completed phage head. The products of genes *W* and *FII*, among others, then unite the head with a separately assembled tail structure to form the mature particle.

The principle of packaging *in vitro* is to supply the ligated recombinant DNA with high concentrations of phage-head precursor, packaging proteins and phage tails. Practically, this is most efficiently performed in a very concentrated mixed lysate of two induced lysogens, one of which is blocked at the pre-head stage by an amber mutation in gene *D* and therefore accumulates this precursor, while the other is prevented from forming any head structure by an amber mutation in gene *E* (Hohn & Murray 1977). In the mixed lysate, genetic complementation occurs and exogenous DNA is packaged (Fig. 4.15). Although concatemeric DNA is the substrate for packaging (covalently joined concatemers are, of course, produced in the ligation reaction by association of the natural cohesive ends of λ), the *in vitro* system will package added monomeric DNA, which presumably first concatemerizes non-covalently.

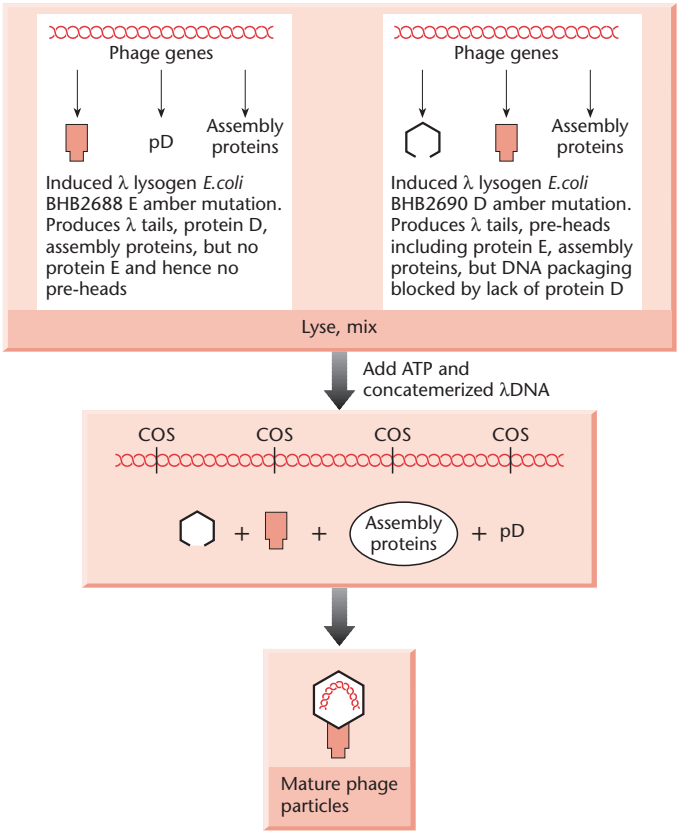


Fig. 4.15 *In vitro* packaging of concatemerized phage-λ DNA in a mixed lysate.

There are two potential problems associated with packaging *in vitro*. First, endogenous DNA derived from the induced prophages of the lysogens used to prepare the packaging lysate can itself be packaged. This can be overcome by choosing the appropriate genotype for these prophages, i.e. excision upon induction is inhibited by the *b2* deletion (Gottesmann & Yarmolinsky 1968) and *imm 434* immunity will prevent plaque formation if an *imm 434* lysogenic bacterium is used for plating the complex reaction mixture. Additionally, if the vector does not contain any amber mutation a non-suppressing host bacterium can be used so that endogenous DNA will not give rise to plaques. The second potential problem arises from recombination in the lysate between exogenous DNA and induced prophage markers. If troublesome, this can be overcome by using recombination-deficient (i.e. *red*$^-$ *rec*$^-$) lysogens and by UV-irradiating the cells used to prepare the lysate, so eliminating the biological activity of the endogenous DNA (Hohn & Murray 1977).

DNA cloning with single-stranded DNA vectors

M13, f1 and fd are filamentous coliphages containing a circular single-stranded DNA molecule. These coliphages have been developed as cloning vectors, for they have a number of advantages over other vectors, including the other two classes of vector for *E. coli*, plasmids and phage λ. However, in order to appreciate their advantages, it is essential to have a basic understanding of the biology of filamentous phages.

The biology of the filamentous coliphages

The phage particles have dimensions 900 nm × 9 nm and contain a single-stranded circular DNA molecule, which is 6407 (M13) or 6408 (fd) nucleotides long. The complete nucleotide sequences of fd and M13 are available and they are 97% homologous. The differences consist mainly of isolated nucleotides

here and there, mostly affecting the redundant bases of codons, with no blocks of sequence divergence. Sequencing of f1 DNA indicates that it is very similar to M13 DNA.

The filamentous phages only infect strains of enteric bacteria harbouring F pili. The adsorption site appears to be the end of the F pilus, but exactly how the phage genome gets from the end of the F pilus to the inside of the cell is not known. Replication of phage DNA does not result in host-cell lysis. Rather, infected cells continue to grow and divide, albeit at a slower rate than uninfected cells, and extrude virus particles. Up to 1000 phage particles may be released into the medium per cell per generation (Fig. 4.16).

The single-stranded phage DNA enters the cell by a process in which decapsidation and replication are tightly coupled. The capsid proteins enter the cytoplasmic membrane as the viral DNA passes into the cell while being converted to a double-stranded replicative form (RF). The RF multiplies rapidly until about 100 RF molecules are formed inside the cell. Replication of the RF then becomes asymmetric, due to the accumulation of a viral-encoded single-stranded specific DNA-binding protein. This protein binds to the viral strand and prevents synthesis of the complementary strand. From this point on, only viral single strands are synthesized. These progeny single strands are released from the cell as filamentous particles following morphogenesis at the cell membrane. As the DNA passes through the membrane, the DNA-binding protein is stripped off and replaced with capsid protein.

Why use single-stranded vectors?

For several applications of cloned DNA, single-stranded DNA is required. Sequencing by the original dideoxy method required single-stranded DNA, as do techniques for oligonucleotide-directed mutagenesis and certain methods of probe preparation. The use of vectors that occur in single-stranded form is an attractive means of combining the cloning, amplification and strand separation of an originally double-stranded DNA fragment.

As single-stranded vectors, the filamentous phages have a number of advantages. First, the phage DNA is replicated via a double-stranded circular DNA (RF) intermediate. This RF can be purified and manipulated *in vitro* just like a plasmid. Secondly, both RF and single-stranded DNA will transfect competent *E. coli* cells to yield either plaques or infected colonies, depending on the assay method. Thirdly, the size of the phage particle is governed by the size of the viral DNA and therefore there are no packaging constraints. Indeed, viral DNA up to six times the length of M13 DNA has been packaged (Messing *et al.* 1981). Finally, with these phages it is very easy to determine the orientation of an insert. Although the relative orientation can be determined from restriction analysis of RF, there is an easier method (Barnes 1980). If two clones carry the insert in opposite directions, the single-stranded DNA from them will hybridize and this can be detected by agarose gel electrophoresis. Phage from as little as 0.1 ml of culture can be used in assays of this sort, making mass screening of cultures very easy.

In summary, as vectors, filamentous phages possess all the advantages of plasmids while producing particles containing single-stranded DNA in an easily obtainable form.

Development of filamentous phage vectors

Unlike λ, the filamentous coliphages do not have any non-essential genes which can be used as cloning sites. However, in M13 there is a 507 bp intergenic region, from position 5498 to 6005 of the DNA sequence, which contains the origins of DNA replication for both the viral and the complementary strands. In most of the vectors developed so far, foreign DNA has been inserted at this site, although it is possible to clone at the carboxy-terminal end of gene IV (Boeke *et al.* 1979). The wild-type phages are not very promising as vectors because they contain very few unique sites within the intergenic region: *Asu*I in the case of fd, and *Asu*I and *Ava*I in the case of M13.

The first example of M13 cloning made use of one of 10 *Bsu*I sites in the genome, two of which are in the intergenic region (Messing *et al.* 1977). For cloning, M13 RF was partially digested with *Bsu*I and linear full-length molecules isolated by agarose gel electrophoresis. These linear monomers were blunt-end-ligated to a *Hin*dII restriction fragment comprising the *E. coli lac* regulatory region and the genetic information for the α-peptide of β-galactosidase. The complete ligation mixture was used to

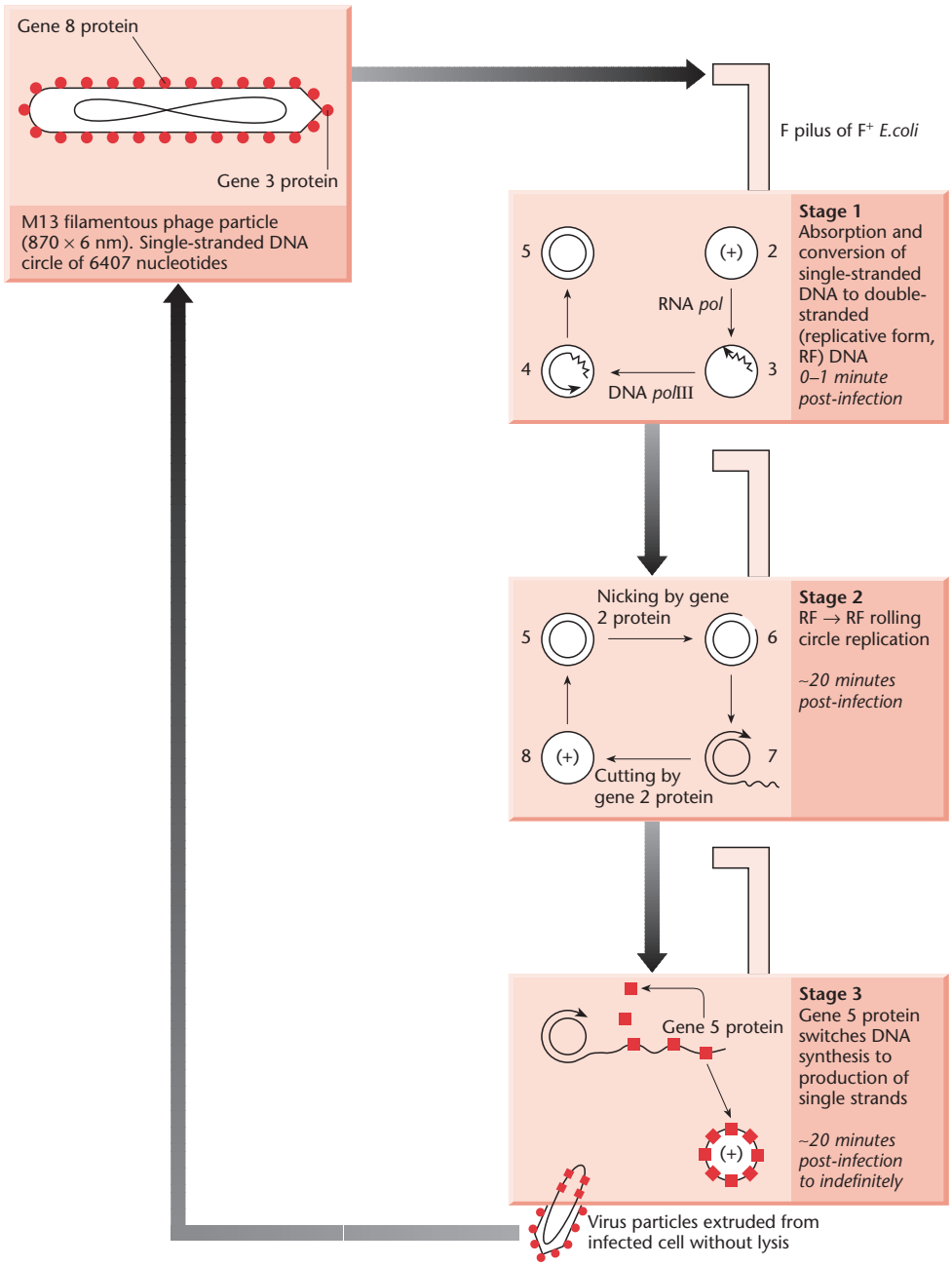


Fig. 4.16 Life cycle and DNA replication of phage M13.

transform a strain of *E. coli* with a deletion of the β-galactosidase α-fragment and recombinant phage detected by intragenic complementation on media containing IPTG and Xgal (see Box 3.2 on p. 35). One of the blue plaques was selected and the virus in it designated M13 mp1.

Insertion of DNA fragments into the *lac* region of M13 mp1 destroys its ability to form blue plaques,

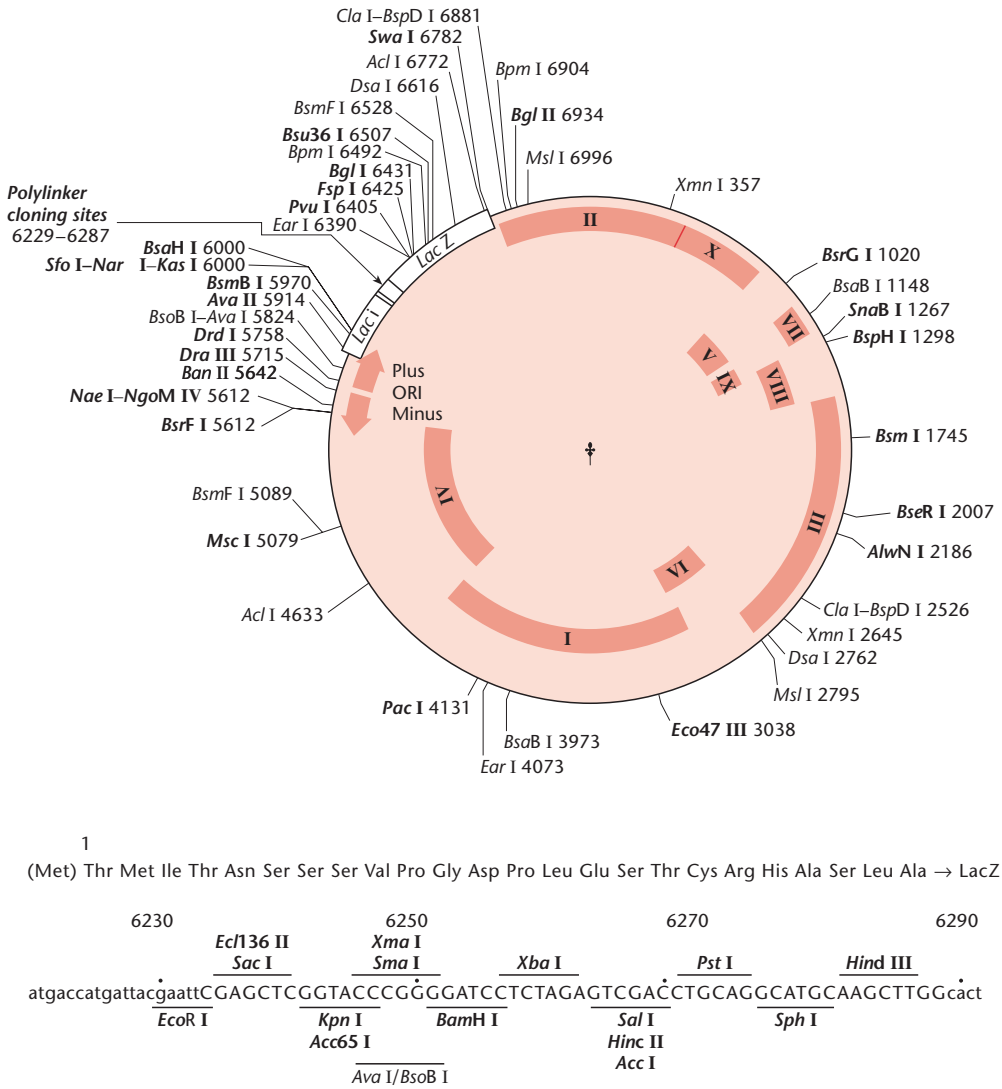


Fig. 4.17 Restriction map of cloning vector M13mp18. Phages M13mp18 and M13mp19 are 7249 bases in length and differ only in the orientation of the 54-base polylinker that they carry. The map shows the restriction sites of enzymes that cut the molecule once or twice. The unique sites are shown in bold type.

making detection of recombinants easy. However, the *lac* region only contains unique sites for *Ava*II, *Bgl*I and *Pvu*I and three sites for *Pvu*II, and there are no sites anywhere on the complete genome for the commonly used enzymes such as *Eco*RI or *Hin*dIII. To remedy this defect, Gronenborn and Messing (1978) used *in vitro* mutagenesis to change a single base pair, thereby creating a unique *Eco*RI site within the *lac* fragment. This variant was designated M13 mp2. This phage derivative was further modified to generate derivatives with polylinkers upstream of the lac α-fragment (Fig. 4.17). These derivatives (mp7–mp11, mp18, mp19) are the exact M13 counterparts of the pUC plasmids shown in Fig. 4.9.