

Choosing a Cloning Vector

Andrew Preston

1. Introduction

Since the construction of the first generation of general cloning vectors in the early 1970s, the number of plasmids created has increased to an almost countless number. Thus, a critical decision facing today's investigator is that of which plasmid to use in a particular project? Despite the bewildering choice of commercial and other available vectors, the choice of which cloning vector to use can be decided by applying a small number of criteria: insert size, copy number, incompatibility, selectable marker, cloning sites, and specialized vector functions. Several of these criteria are dependent on each other. This chapter discusses these criteria in the context of choosing a plasmid for use as a cloning vector and **Table 1** displays the features of some commonly used cloning vectors.

2. Criteria for Choosing a Cloning Vector

2.1. Insert Size

For projects in which it is desired that a particular piece of DNA be cloned, one consideration is the size of the insert DNA. Most general cloning plasmids can carry a DNA insert up to around 15 kb in size. Inserts in excess of this place constraints on proper replication of the plasmids (particularly for high-copy-number vectors) and can cause problems with insert stability. Several types of vectors are available for cloning large fragments of DNA. These are most commonly used to construct libraries of clones that are often representative of entire genomes. Library clones are then screened to identify the particular clone that carries the DNA of interest.

2.1.1. Cosmids (1,2)

Cosmids are conventional vectors that contain a small region of bacteriophage λ DNA containing the cohesive end site (cos). This contains all of the *cis*-acting elements for packaging of viral DNA into λ particles. For cloning of DNA in these vec-

Table 1
Commonly Used Cloning Vectors

Plasmid	Features	Commercial source
pUC18, pUC19	Small size (2.7 kb) High copy number Multiple cloning site Ampicillin-resistance marker Blue/white selection (<i>see</i> Chapter 19)	NEB
pBluescript vectors	As pUC Single-stranded replication origin T7 and SP6 promoters flanking MCS ^a	Stratagene
pACYC vectors	Low copy number (15 copies per cell) p15A origin of replication	NEB
Supercos	Cosmid vector Two cos sites Insert size 30–42 kb Ampicillin-selectable marker T3 and T7 promoters flanking cloning site	Stratagene
EMBL3	λ replacement vector MCS sites: <i>SalI</i> , <i>BamHI</i> and <i>EcoRI</i>	Promega
λ ZAP	λ vector In vivo excision into pBluescript phagemid vector Cloning capacity 10 kb Blue/white selection	Stratagene
pBeloBAC11	BAC vector Inserts up to 1 Mb T7 and SP6 promoters flank insertion site Blue/white selection Cos site LoxP site	NEB

^aMCS = multiple-cloning site.

tors, linear genomic DNA fragments are ligated in vitro to the vector DNA and this is then packaged into bacteriophage particles (*see* Chapter 7). On introduction into *Escherichia coli* host cells, the vector is circularized to form a large plasmid containing the cloned DNA fragment. Cosmids are most commonly used to generate large insert libraries. Because of the constraints of packaging, the vector plus insert should comprise between 28 and 45 kb.

2.1.2. λ Vectors (see Chapter 18)

The bacteriophage λ genome comprises 48,502 bp. On entering the host cell, the phage adopts one of two life cycles: lytic growth or lysogeny. In lytic growth, approx 100 new virions are synthesized and packaged before lysing the host cell, releasing the progeny phage to infect new hosts. In lysogeny, the phage genome undergoes recombination into the host chromosome, where it is replicated and inherited along with the host DNA (3). Which of the two different life cycles is adopted is determined by the multiplicity of infection and the host cell nutritional status. The larger the multiplicity of infection and the poorer the nutritional state of the cell, the more lysogeny is favored (4).

Use of λ as a cloning vector involves only the lytic cycle. This renders the middle third of the λ genome, which encodes functions for gene expression regulation and establishment of lysogeny, redundant for these purposes. It is the ability to replace this portion of the genome with foreign DNA without affecting the lytic life cycle that makes λ useful as a cloning vector. Insertion λ vectors have the nonessential DNA deleted and contain a single site for insertion of DNA. Typically 5–11 kb of foreign DNA can be accommodated in these vectors. Replacement vectors contain specific restriction sites flanking the nonessential genes. Digestion of linear vector DNA at these sites produces two “arms” that are ligated to the foreign DNA. Many commercially available λ vectors are sold as predigested and purified arms. Replacement vectors typically can accommodate between 8 and 24 kb of foreign DNA, depending on the vector.

During the early phase of infection, λ DNA replicates bidirectionally, in circular form from a single origin of replication before shifting to replication via a rolling-circle mode. This produces a concatamer of genomes in a head-to-tail arrangement that is then processed to give individual genomes for packaging. The shift to rolling-circle replication depends on the interplay between host- and phage-encoded recombination functions. As the recombination proficiency of different λ vectors can vary, the investigator is urged to ensure that the *E. coli* strain used for infection is capable of properly replicating the phage. This information is generally supplied with commercially available vectors. A great many features that aid cloning into and screening of recombinant phage have also been incorporated into λ vectors. Often, the use of these features also necessitates the use of particular host strains.

2.1.3. Bacterial Artificial Chromosomes (5,6)

Bacterial artificial chromosomes (BACs) are circular DNA molecules. They contain a replicon that is based on the F factor comprising *oriS* and *repE* encoding an ATP-driven helicase along with *parA*, *parB*, and *parC* to facilitate accurate partitioning (see Chapter 1). The F factor is capable of carrying up to one quarter of the *E. coli* chromosome and, thus, BACs are capable of maintaining very large DNA inserts (often up to 350 kb); however, many BAC libraries contain inserts of around 120 kb. Newer versions of BAC vectors contain sites to facilitate recovery of cloned DNA (e.g., *loxP*) (7). A DNA fragment is cloned into BAC vectors in a similar fashion to cloning into general cloning vectors; DNA is ligated to a linearized vector and then introduced into an *E. coli* cloning strain by electroporation.

2.2. Copy Number

Different cloning vectors are maintained at different copy numbers, dependent on the replicon of the plasmid (*see* Chapter 1). In a majority of cases in which a piece of DNA is cloned for maintenance and amplification for subsequent manipulation, the greater the yield of recombinant plasmid from *E. coli* cultures, the better. In this scenario, a high-copy-number vector is desirable such as those whose replication is driven by the ColE1 replicon (8). The original ColE1-based plasmids have a copy number of 15–20. However, a mutant ColE1 replicon, as found in the pUC series of plasmids (9), produces a copy number of 500–700 as a result of a point mutation within the RNAI regulatory molecule (*see* Chapter 1) that renders it more resistant to inhibition by RNAI (10). It should be noted that this mutation is temperature sensitive. Mutant RNAI is resistant to RNAI inhibition at 37°C or 42°C but not at 30°C, at which temperature the copy number of pUC plasmids returns to that of nonmutated ColE1 plasmids.

In some cases, a high-copy-number may cause problems for cloning DNA. For example, the cloned DNA may encode proteins that are toxic to the cell when present at high levels. This is particularly true of membrane proteins. Even if the protein is expressed poorly from the cloned DNA, the presence of many hundreds of copies of the gene on the plasmid may raise the level of protein to toxic levels. In these cases, using a plasmid with a lower copy number may reduce the gene dosage below a level at which toxicity occurs. For example, pBR322 is based on the original ColE1 replicon and thus has a copy number of 15–20 (11). The pACYC series of plasmids are based on the p15A replicon, which has a copy number of 18–22 (12). Low-copy-number plasmids include pSC101 (copy number around 5) (13), whereas BACs are maintained at one copy per cell (5).

2.3. Incompatibility

Incompatibility refers to the fact that different plasmids are sometimes unable to coexist in the same cell. This occurs if the two different plasmids share functions required for replication and/or partitioning into daughter cells. Direct competition for these functions often leads to loss of one of the plasmids from the cell during growth of a culture. Plasmid size can also influence maintenance within a culture, as larger plasmids require longer for replication and, thus, may be outcompeted by faster replicating of smaller plasmids. Thus, ColE1-based plasmids are incompatible with other ColE1-based plasmids but are compatible with R6K- or p15A-based plasmids. Incompatibility only becomes an issue if it requires that two plasmids be maintained-together (e.g., if cloning into an *E. coli* strain that contains a helper plasmid) (*see* Chapter 3).

2.4. Selectable Marker

Introduction of plasmids into *E. coli* cells is an inefficient process. Thus, a method of selecting those cells that have received a plasmid is required. Furthermore, cells that do not contain a plasmid are at a growth advantage over those that do and, thus, have to replicate both the chromosome and additional plasmid DNA. This is of particular consequence when dealing with high-copy-number or large plasmids. In this case, a selective pressure must be imposed for maintenance of the plasmid. Almost all conventional plasmids use an antibiotic resistance gene as a selectable marker, carried

on the backbone of the vector. Thus, the addition of the appropriate antibiotic to the growth medium will kill those cells that do not contain the plasmid and produce a culture in which all cells do contain a plasmid. In many cases, the choice of antibiotic is not restricted. However, some cloning strains of *E. coli* are inherently resistant to some antibiotics and, thus, the same antibiotic cannot be used as a selection for those cells carrying a particular plasmid. The genotype of the desired cloning strain should be checked prior to cloning (see Chapter 3). In some situations, downstream applications render some antibiotics as unsuitable choices. For example, the mutation of genes in cloned DNA fragments is often achieved by the disruption of the gene by insertion of an antibiotic-resistance cassette. This both mutates the gene and acts as a marker for the mutation. Often, the mutation is introduced into the organism from which the DNA is derived. In this case, only some antibiotics are suitable for use because of restrictions on introducing particular antibiotic resistances into some bacteria against which the antibiotic is used as a therapeutic or because of the inherent resistance of the original organism to the antibiotic. In these projects, the vector should not confer resistance to the antibiotic to be used in the downstream application.

Some plasmid vectors contain two antibiotic-resistance cassettes. For example, pACYC177 contains both ampicillin- and kanamycin-resistance genes. Many of the cloning sites in this vector lie in these genes and cloning into one of these sites inactivates that particular antibiotic resistance. Profiling the antibiotic resistances of recombinant clones is a way of selecting for those carrying insert DNA fragments. Many newer vectors now carry specialized cloning sites (polylinker, multiple-cloning site; see **Subheading 2.5.**) for which cloning of insert DNA does not interfere with inherent vector functions. The most common antibiotic resistances carried on vectors used in *E. coli* are resistance to ampicillin, kanamycin, tetracycline, and chloramphenicol.

2.4.1. Ampicillin

This drug inhibits the bacterial transpeptidase involved in peptidoglycan biosynthesis and thus inhibits cell wall biosynthesis (**14**). As such, ampicillin inhibits log-phase bacteria but not those in a stationary phase. Resistance to ampicillin is conferred by a β -lactamase, which cleaves the β -lactam ring of ampicillin (**14**). The β -lactamase most commonly expressed by cloning vectors is that encoded by the *bla* gene (**15**).

2.4.2. Kanamycin

A member of the aminoglycoside family of antibiotics, kanamycin was first isolated from *Streptomyces kanamyceticus* in Japan in 1957. This polycation is taken into the bacterial cell through outer-membrane pores but crosses the cytoplasmic membrane in an energy-dependent process utilizing the membrane potential. The molecule interacts with three ribosomal proteins and with rRNA in the 30S ribosomal subunit, to prevent the transition of an initiating complex to a chain-elongating complex, and thus inhibits protein synthesis. Resistance to kanamycin is conferred by aminophosphotransferases. Those commonly encoded by vectors are Aph (3')-I from Tn903 and Aph (3')-II from Tn,5 which transfer phosphate from ATP to the kanamycin to inactivate it (**16**). It is important to note that these two resistance genes have differing

DNA sequences and, thus, different restriction maps. They will not cross-hybridize under stringent conditions in Southern hybridizations.

2.4.3. Chloramphenicol

First isolated from a soil actinomycete in 1947, chloramphenicol was widely used as a broad-spectrum antibiotic although its clinical use has been curtailed because of drug-induced bone-marrow toxicity and the emergence of bacterial chloramphenicol resistance. Chloramphenicol inhibits the activity of ribosomal peptidyl transferase and thus inhibits protein synthesis (17). Chloramphenicol resistance is conferred by chloramphenicol acetyl transferase (*cat*), which transfers an acetyl group from acetyl CoA to chloramphenicol and inactivates it (18).

2.4.4. Tetracycline

Originally isolated from *Streptomyces aureofaciens* in 1948, there are now many tetracycline derivatives available. They bind to a single site on the 30S ribosomal subunit to block the attachment of aminoacyl tRNA to the acceptor site and thus inhibit protein synthesis (19). Tetracycline resistance is conferred by efflux proteins, TetA (A–E), which catalyze the energy-dependent export of tetracycline from the cell against a concentration gradient (19).

2.5. Cloning Sites

The cloning of DNA into a vector usually involves ligation of the insert DNA fragment to vector DNA that has been cut with a restriction endonuclease. This is facilitated by the insert and vector DNA fragments having compatible cohesive ends. Thus, the vector of choice may be one that has a restriction endonuclease site that is compatible with the insert fragment-generating enzyme. It should be noted, however, that any blunt-end fragment can be ligated to any other blunt-end fragment and that even DNA-fragments generated by restriction enzymes that generate overhangs can be made blunt ended (*see* Chapter 15). In many older vectors, the restriction endonuclease sites were dispersed around the plasmid and were often in one of the vector genes. For example, many of the cloning sites in the pACYC series of vectors are located within one of the antibiotic-resistance genes of these plasmids. Cloning into these sites inactivated the resistance gene and the subsequent sensitivity to the antibiotic was used as a screen for recombinant plasmids containing the insert DNA.

More modern vectors often contain an artificial stretch of DNA that has a high concentration of restriction endonuclease sites that do not occur elsewhere on the plasmid. These multiple-cloning sites (MCSs) or polylinkers give a wide choice of restriction endonucleases for use in the cloning step. They also limit the cloning site to one small region of the vector and thus allow the specific positioning of the insert DNA close to other features of the vector. For example, the MCSs of many vectors such as the pUC series are flanked by sequences complementary to a universal series of primers, the M13 forward and reverse primers. These priming sites are oriented such that extension of the primers annealed to these sites allows sequencing of both ends of an insert DNA in the MCS. In this fashion, one set of universal primers can

be used to sequence any insert DNA regardless of which site the DNA was inserted at within the MCS.

Many plasmids contain MCSs that lie within the coding sequence of the α fragment of *lacZ*. This feature (blue/white selection) facilitates the identification of recombinant constructs that carry a cloned fragment by distinguishing them from clones that arise from religation of the cloning vector. This feature is discussed in Chapter 19.

2.6. Specialized Plasmid Functions

Some projects will involve specific downstream applications that will require specialized plasmid functions that are only present on some plasmids. For example, both the pUC and pBluescript series of vectors are high-copy-number, ampicillin-resistance-conferring plasmids that contain MCSs that facilitate the use of a wide range of restriction endonucleases in the cloning step. However, one feature present on pBluescript vectors that is not present on pUC vectors is promoters flanking the MCS that permit transcription of the insert DNA on either strand. The two promoters T7 and SP6 are recognized by bacteriophage RNA polymerases that must be supplied *in trans*. They do not transcribe host genes or other plasmid genes, enabling specific transcription of the insert DNA (*see* Chapter 27).

pBluescript vectors are phagemids. They contain a single-stranded filamentous bacteriophage origin of replication (M13 phage) and, thus, are useful for generating single-stranded DNA (*see* Chapter 13) for applications such as DNA sequencing or site-directed mutagenesis. Single-stranded replication is initiated by infecting with a helper phage encoding the necessary functions. These vectors can also replicate as conventional double-stranded plasmids. The single-stranded origin can exist in two orientations. Those versions in which it is in same orientation as the plasmid origin are denoted as “+,” whereas those with the origin in the opposite orientation are denoted as “-.”

Many plasmids have been designed to achieve high-level expression of recombinant proteins from the cloned DNA. These expression vectors are discussed in Chapter 28.

3. Summary

When choosing a cloning vector for use in a cloning project, the investigator is faced with an enormous choice. However, the application of a small number of criteria can quickly guide the selection of a suitable vector. Many plasmids contain sufficient features that render them suitable for a wide range of projects. Thus, the investigator needs to be equipped with only a small number of vectors in order to satisfy most needs.

References

1. Hohn, B., Koukolikova-Nicola, Z., Lindenmaier, W., et al. (1988) Cosmids. *Biotechnology* **10**, 113–127.
2. Collins, J. and Hohn, B. (1978) Cosmids: a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage lambda heads. *Proc. Natl. Acad. Sci. USA* **75**, 4242–4246.
3. Ptashne, M. (1986) *A Genetic Switch: Gene Control and Phage λ* . Blackwell Scientific, Palo Alto, CA.

4. Herskowitz, I. and Hagen, D. (1980) The lysis–lysogeny decision of phage λ : explicit programming and responsiveness. *Annu. Rev. Genet.* **14**, 399–445.
5. Shizuya, H., Birren, B., Kim, U. J., et al. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797.
6. Monaco, A. P. and Larin, Z. (1994) YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends Biotechnol.* **12**, 280–286.
7. Palazzolo, M. J., Hamilton, B. A., Ding, D. L., et al. (1990) Phage lambda cDNA cloning vectors for subtractive hybridization, fusion-protein synthesis and Cre–loxP automatic plasmid subcloning. *Gene* **88**, 25–36.
8. Kahn, M., Kolter, R., Thomas, C., et al. (1979) Plasmid cloning vehicles derived from plasmids ColE1, F, R6K, and RK2. *Methods Enzymol.* **68**, 268–280.
9. Vieira, J. and Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259–268.
10. Lin-Chao, S., Chen, W. T., and Wong, T. T. (1992) High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Mol. Microbiol.* **6**, 3385–3393.
11. Bolivar, F., Rodriguez, R. L., Greene, P. J., et al. (1977) Construction and characterization of new cloning vehicles, II: a multipurpose cloning system. *Gene* **2**, 95–113.
12. Chang, A. C. and Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. *J. Bacteriol.* **134**, 1141–1156.
13. Stoker, N. G., Fairweather, N. F., and Spratt, B. G. (1982) Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. *Gene* **18**, 335–341.
14. Donowitz, G. R. and Mandell, G. L. (1988) Beta-lactam antibiotics (1). *N. Engl. J. Med.* **318**, 419–426.
15. Sutcliffe, J. G. (1978) Nucleotide sequence of the ampicillin resistance gene of *Escherichia coli* plasmid pBR322. *Proc. Natl. Acad. Sci. USA* **75**, 3737–3741.
16. Umezawa, H. (1979) Studies on aminoglycoside antibiotics: Enzymic mechanism of resistance and genetics. *Jpn. J. Antibiot.* **32(Suppl)**, S1–S14.
17. Drainas, D., Kalpaxis, D. L., and Coutsogeorgopoulos, C. (1987) Inhibition of ribosomal peptidyltransferase by chloramphenicol: kinetic studies. *Eur. J. Biochem.* **164**, 53–58.
18. Shaw, W. V. (1983) Chloramphenicol acetyltransferase: enzymology and molecular biology. *CRC Crit. Rev. Biochem.* **14**, 1–46.
19. Schnappinger, D. and Hillen, W. (1996) Tetracyclines: antibiotic action, uptake, and resistance mechanisms. *Arch. Microbiol.* **165**, 359–369.