DNA Cloning

Michael Andrew Quail, *The Wellcome Trust Sanger Institute, Cambridge, UK*

Deoxyribonucleic acid (DNA) cloning is the art of creating recombinant DNA molecules that can be introduced into living cells, replicated and stably inherited, such that multiple 'clonal' copies of that DNA are produced. Techniques for cloning and the various vectors that can be used are described.

Introduction

Deoxyribonucleic acid (DNA) cloning is a cornerstone of molecular biology. Its power is such that without it a large proportion of modern biological science would simply not be possible.

A clone can be defined as an identical copy. DNA cloning involves joining, or 'ligating', DNA with a 'vector' that enables the resulting construct to be introduced into a cell, replicated and passed on to daughter cells as that cell divides (**Figure 1**). The result is a population of cells all containing clones of the original DNA. Thereafter that DNA is immortalized, since cells can be grown to order and DNA extracted for study or further manipulation whenever it is required.

DNA cloning has been made possible by the discovery of two types of protein, those that break or modify DNA such that they have suitable termini for ligation and those that are capable of ligating those molecules of DNA.

How to Clone

DNA exists as a double helix of antiparallel polymer strands with the nucleotide units in each joined by $5'-3'$ phosphodiester bonds. The ability to selectively break and join these bonds is the basis of DNA cloning. For the detailed experimental protocols described in this article the reader is referred to Sambrook and Russell (2001), an excellent three-volume laboratory manual of molecular biology.

Ligation

DNA strands can be joined through the action of the enzyme DNA ligase, the normal biological role of which is to join the series of Okazaki fragments that are generated on the lagging strand during replication. Typically the ligase from the *Escherichia coli* bacteriophage T4, known as 'T4 DNA ligase', is used in cloning experiments as it is both easily purified and highly active. Ligase can be purchased commercially and is normally supplied with a $10\times$ concentrate of reaction buffer, containing Mg^{2+} and adenosine $5'$ triphosphate (ATP) that are both required for the activity of this enzyme. Ligase utilizes the Mg^{2+} cation to contact the free $3'$ hydroxyl group at the end of a DNA strand and uses the energy stored in the phosphodiester bonds of ATP to covalently attach this position to the $5'$ phosphate at the terminus of another DNA strand, so joining the two strands. Ligating double-stranded DNA thus consumes two molecules of ATP and requires the two DNA molecules to have compatible termini. However, only one of the two $5'$ termini needs to have a phosphate group. This is useful because dephosphorylation of the vector DNA prevents its self-ligation. DNA can be ligated with either blunt ends or, more efficiently, as compatible staggered ('sticky'), cohesive ends (**Figure 2**), the conditions used being slightly different in each case.

For sticky-end ligation, insert DNA is incubated in $1\times$ ligase buffer in the presence of approximately equimolar amounts (typically 10–50 ng) of vector that has compatible termini, with 0.1–1 units of T4 DNA ligase in a total volume of 10–50 µl for 2–8 h at 16°C, or overnight at 4°C. DNA molecules possessing sticky ends transiently base pair while in solution so bringing their ligatable termini into close proximity. It is for this reason that sticky-end ligation is more efficient.

Blunt-end ligation can be driven by increasing the concentration of termini such that ligation is favoured. Typically vector is mixed with a threefold molar excess of insert DNA in $1\times$ ligase buffer, with 1 unit of T4 DNA ligase, in a total volume of 4–10 µl and incubated overnight at 16°C.

Preparation of DNA for ligation

Prior to ligation, compatible insert and vector DNA must be prepared. Both need compatible termini and must be free of contaminating chemicals such as phenol, ethylenediaminetetraacetic acid (EDTA) or ethanol that may inhibit ligation. Vector DNA is typically prepared by digestion with the relevant
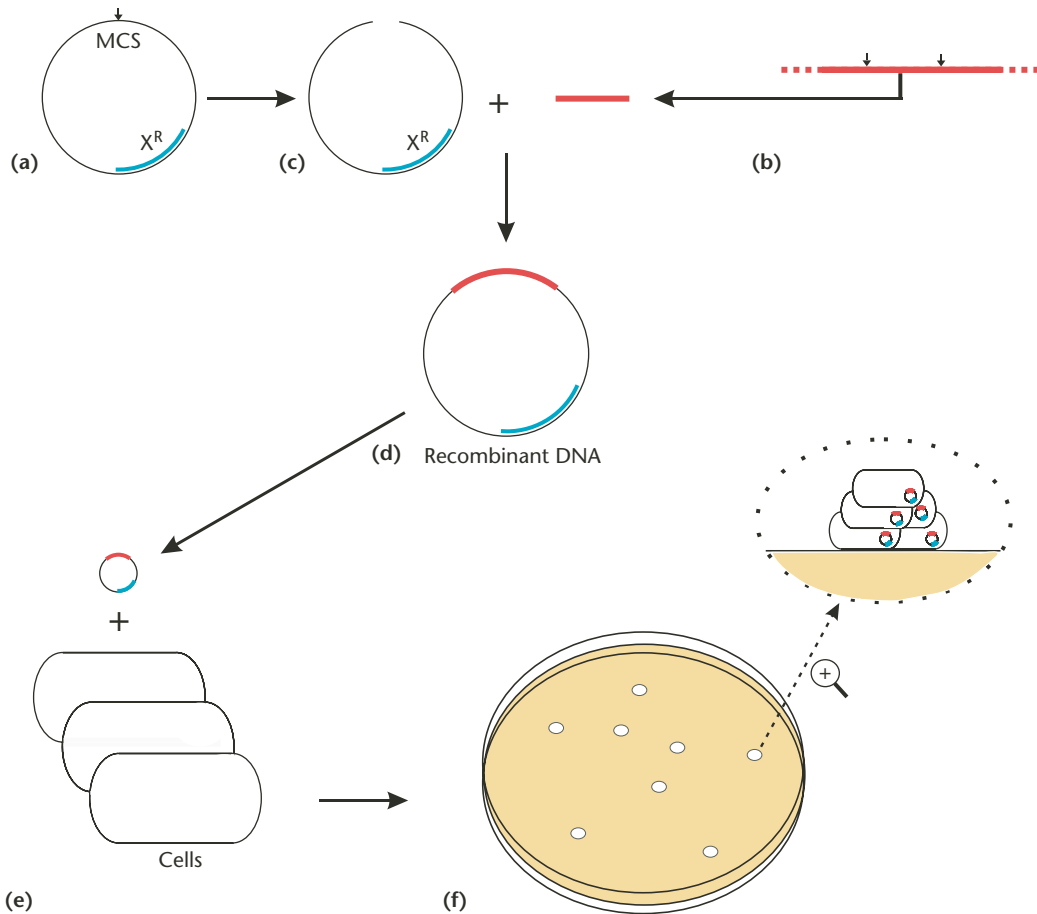
Figure 1 A schematic representation of a typical cloning experiment. (a) The vector is cut (↓) within its multicloning site (MCS). (b) The target DNA is cut (↓) so as to produce termini compatible with the vector. (c) The vector and insert are ligated to produce recombinant DNA. (d, e) Recombinant DNA is introduced into appropriate host cells. In this illustration the vector encodes resistance to an antibiotic, X. (f) If the cells are plated out onto medium containing X, only cells that have been transformed will grow and divide to form colonies (groups of around a million cells or 'clones' that have arisen from the same original transformed cell).
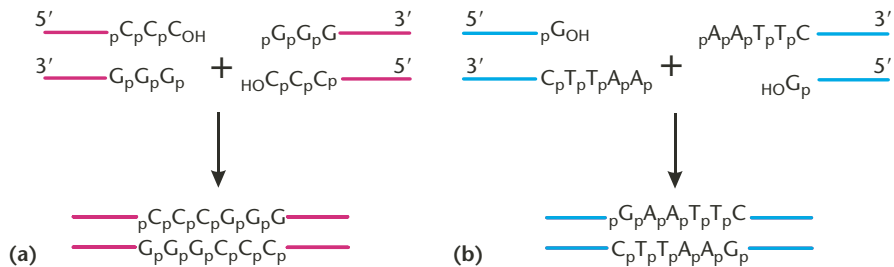

Figure 2 Schematic diagram depicting (a) blunt end ligation and (b) ligation of cohesive ('sticky') ends. (a) Two double-stranded molecules with blunt ends are ligated to produce a single molecule. (b) 5′ protruding sticky ends, which were created by digestion with the enzyme *Eco*RI, are ligated to produce a single double-stranded DNA molecule with the *Eco*RI recognition site being recreated.

restriction enzyme, then dephosphorylated by incubation with alkaline phosphatase to remove the terminal 5′ phosphate groups and so prevent vector self-ligation. Insert DNA is typically prepared using one of the methods described below, then if it is part of a mixture of DNA fragments, it is purified by separating that mixture on a preparative (normally low-melting-point) agarose gel, cutting out the slice of agarose containing the fragment one wishes to clone and extracting the DNA from the agarose matrix.

Extraction of DNA from agarose can be performed by one of a number of methods including melting followed by phenol extraction, agarase digestion of the agarose backbone or absorption and subsequent elution of the DNA from a silica matrix.

Generation of sticky ends

DNA molecules with sticky ends are usually generated through the use of type II restriction endonucleases. These enzymes, whose biological function is to digest or 'restrict' foreign DNA that enters the cell, recognize specific sequences, 'recognition sites' and catalyze the double-stranded cleavage of DNA within that site to generate either molecules with protruding 5′, protruding 3′ or blunt termini (**Table 1**).

Some restriction enzymes generate protruding termini that are compatible, for example *Bam*HI, *Bcl*I, *Bgl*II, *Dpn*II, *Mbo*I and *Sau*3AI all cleave to give GATC 5′ overhangs, and so DNA cleaved by any of these enzymes will ligate to that cleaved by any of the others. This is of particular use when one wishes to generate genomic libraries. A DNA library is a ligation that contains a mixture of DNA inserts. For whole-genome studies, libraries are often made so as to contain inserts representing all sequences in that genome. The patterns of cleavage of a restriction enzyme with a six-base (and longer) recognition sequence are often nonrandom across a whole genome. Therefore a common technique is to digest the DNA with an enzyme with a four-base recognition site such as *Sau*3AI (GATC). Normally such an enzyme would cut every 4^4 (that is, 256), bases, but if the digestion time is kept short or the amount of enzyme limited,

then a partial digestion will occur and so by controlling the conditions a spectrum of fragments in the desired size range can be produced. Since it is important to cleave the vector DNA only once, preferably within the desired cloning site region, and *Sau*3AI would digest a typical vector molecule into many pieces, the vector can be prepared using an enzyme such as *Bam*H I that cuts less frequently but leaves compatible sticky ends.

Generation of blunt ends

DNA with blunt ends can be prepared using restriction enzymes that leave a blunt end, through cleavage with DNaseI or after end-repair of DNA molecules that have staggered termini. Although less efficient, this method has more versatility as there is no absolute requirement for the presence of a particular recognition site. DNA can be broken randomly by mechanical breakage or digested with any enzyme, then repaired. For ends with protruding 5′ termini, repair is commonly achieved by incubation with nucleotides and the Klenow fragment of *E. coli* DNA polymerase I. Protruding 3′ ends are removed by taking advantage of the high 3′–5′ exonuclease activity of T4 DNA polymerase which digests the single-stranded bases in the overhang. If the nature of the ends is unknown, as would be the case for random breakage, then ends are normally repaired by digesting protruding single-stranded termini by incubation with mung bean or *Bal*3I nuclease or by incubating with a mixture of both T4 and Klenow polymerase. (*See* DNA: Mechanical Breakage.)

Table 1 A selection of restriction endonucleases commonly used in DNA cloning

Enzyme[a]	Source	Recognition site[b]	Ends produced
*Bam*HI	*Bacillus amyloliquefaciens* H	5′ GˆGATCC 3′	5′ overhang
*Eco*RI	*Escherichia coli*	5′ GˆAATTC 3′	5′ overhang
*Hin*dIII	*Haemophilus influenzae*	5′ AˆAGCTT 3′	5′ overhang
*Nco*I	*Nocardia corallina*	5′ CˆCATGG 3′	5′ overhang
*Nde*I	*Neisseria denitrificans*	5′ CAˆTATG 3′	5′ overhang
*Not*I	*Nocardia otitidis-caviarum*	5′ GCˆGGCCGC 3′	5′ overhang
*Pst*I	*Providencia stuartii*	5′ CTGCAˆG 3′	3′ overhang
*Sau*3AI	*Staphylococcus aureus* 3A	5′ ˆGATC 3′	5′ overhang
*Sma*I	*Serratia marcescens*	5′ CCCˆGGG 3′	Blunt

[a]Enzymes commonly get their names from abbreviations of the organism from which they were first isolated. Roman numerals after the name depict the order in which enzymes were discovered from an organism, for example, *Hin*dIII was the third enzyme to be isolated from *Haemophilus influenzae*.
[b]The recognition sequence for one strand in the 5′–3′ direction is shown. Typically restriction sites have twofold dyad symmetry, so the same site is seen when reading in the 5′–3′ direction on the other strand. Enzymes cleave the DNA strand within each recognition site to give a 3′ hydroxyl group and a 5′ phosphate group, at the position denoted by ˆ.

Conversion of blunt to sticky ends

There are several methods by which blunt ends can be given sticky tails:

- Through the action of the enzyme terminal transferase, tracts of a single nucleotide base can be added to the 3′ hydroxyl group at each terminus. If different but complementary bases are added to both vector and insert, then the two will be compatible and can be ligated.
- Single base overhangs can be added to the 3′ position of blunt termini by the action of *Taq* DNA polymerase. *Taq* DNA polymerase, commonly used in the polymerase chain reaction (PCR), frequently adds an extra adenosine at the 3′ position of the amplified DNA. Thus PCR products can readily be cloned by ligating to a thymine tailed vector, so-called TA cloning. By incubating DNA with blunt termini in the presence of *Taq* DNA polymerase and a single deoxynucleotide triphosphate at 72°C, DNA with single base 3′ overhangs can easily be prepared.
- Short double-stranded adaptor or linker DNAs containing restriction enzyme recognition sequences can be ligated to blunt termini. Adaptors have one blunt end and one sticky end; thus when ligated they convert the blunt ends to sticky ones. Linkers have two blunt ends but contain internal restriction sites. After linker addition, the resulting molecules can be digested with the appropriate restriction enzyme and sticky ends created.

Generation of deoxyribonucleic acid for cloning by polymerase chain reaction

PCR is an extremely powerful technique as it allows for minute quantities of DNA (or messenger ribonucleic acid (mRNA)) to be amplified up to cloneable quantities. PCR products synthesized using enzymes such as *Taq* DNA polymerase (that lack 3′–5′ exonuclease activity) have a single adenosine 3′ overhang and those synthesized with 'proofreading' polymerases which possess 3′–5′ exonuclease activity are generated with blunt ends. Furthermore, since the primers used for PCR are incorporated into the products (and mismatches between primer and target sequence are often tolerated, especially at the 5′ end of the primer) PCR can be used to introduce restriction sites for cloning at the termini of the amplified fragments. Of particular use in this respect are the recognition sites for the enzymes *Nde*I and *Nco*I, which contain the initiation codon sequence ATG. Here PCR can be used to introduce an *Nde*I or *Nco*I site at the initiation codon of a particular gene so that it can be cloned in-frame into an expression vector.

Vectors

Vectors are DNA molecules that are capable of accepting selected fragments of DNA and replicating the resulting hybrid when it is introduced into living cells. Typically vectors are modified versions of naturally occurring independent replicons or viruses. Due to its extremely well-characterized genetic systems, the majority of vectors propagate in laboratory strains of the enteric bacterium *E. coli*, although 'shuttle vectors' that can also replicate in other organisms and vectors that exclusively replicate in other organisms are widely available. Vectors can be obtained commercially from molecular biology reagent suppliers, from authors of published papers or from collections such as ATCC (see Web Links).

Features of vectors

Required features

Cloning site

The vector must possess a recognition site for at least one restriction endonuclease so that the vector can accept DNA. Most common vectors have been constructed so as to have a multicloning site (MCS). Here a short piece of DNA containing the recognition sequences for several enzymes has been added into the region of the vector where foreign DNA is to be inserted.

Origin of replication

To be able to replicate and propagate the DNA construct, vectors must contain origin of replication (ori) sequences that are recognized by the host cell DNA replication systems. It is the ori that largely determines vector copy number.

Selectable marker

A vector must contain sequences that confer a positive advantage on cells that contain it. This is because maintenance and propagation of cloned DNA places a burden upon that cell, and without an advantage cells that did not contain the clone would grow preferentially resulting eventually in loss of the clone from the population. Furthermore, when DNA is introduced into a cell population only a small fraction of those cells take up the DNA and one needs a mechanism to select for those cells that contain the vector construct. Therefore vectors are designed with selectable markers. These typically fall into two main classes:

- genes that encode resistance to a certain antibiotic (e.g. ampicillin, G418, kanamycin, chloramphenicol, tetracycline, etc.) and

- auxotrophic markers which complement a host cell mutation in an essential biosynthetic pathway. For example, the yeast *URA3* gene is commonly used in yeast vectors in combination with a host cell that has a mutation resulting in a requirement for uracil. When cells are grown in a medium that does not contain uracil only those containing the vector with the *URA3* marker can grow.

Optional features

There are a great many vectors now available to the molecular biologist. Each has slightly different features allowing it to be more suitable for one application than another, and so a careful look at the features that a vector contains is essential. Common features are noted below.

Sequences that enable recombinant selection

Many vectors have their cloning sites within the coding region of a gene that confers a selectable phenotype, such that when DNA is inserted the function of that gene is disrupted. Since the vector can self-ligate to give nonrecombinant molecules, this is a valuable means of selecting for cells that contain recombinant clones. The most widely available selection is α-complementation, where the cloning site is located within coding sequences for the α fragment of the enzyme β-galactosidase (encoded by the *lacZ* gene). The α fragment is able to complement a host cell mutation (*lacZ*ΔM15) in this gene. Thereby on selective agar plates containing the colorimetric substrate for this enzyme (X-gal) and the inducer isopropyl-β-D-thiogalactopyranoside (IPTG), nonrecombinants have β-galactosidase activity, break down X-gal into a blue compound and appear blue whereas recombinants where this gene has been disrupted remain white/colorless. Another recombinant selection method utilizes *ccdB* (or the cell-death gene) the product of which strongly inhibits bacterial DNA gyrase and thereby DNA synthesis, resulting in cell death. If the cloning site is positioned within this gene, only cells containing vector with an insert will grow. Likewise the *sacB* gene (present in P1 and some bacterial artificial chromosome (BAC) vectors) encodes the enzyme levansucrase, which can convert the common sugar sucrose into a cytotoxic compound. When cells are grown on medium containing sucrose, only recombinants where DNA has been inserted into the cloning site to disrupt this gene will grow.

Primer binding sites

Sequences that bind well-characterized (often commercially available) primers are often found flanking the multicloning site. These are useful for sequencing, or amplification by PCR, of the inserted DNA.

Reporter genes

Reporter genes encode a product with an easily detectable activity such as *lacZ* (β-galactosidase), chloramphenicol acetyltransferase (CAT) or green fluorescent protein (GFP). The promoter activity of a given piece of DNA can be quantified by cloning that sequence upstream of a promoterless reporter gene. Likewise if a gene being studied is cloned in-frame with a reporter gene, the activity of the reporter gene will give a measure of the expression level of the corresponding gene product. In the case of GFP, information regarding the subcellular localization of the product can also be obtained.

Strong promoter sequences

1. *Riboprobe generation*. Strong RNA polymerase promoters (e.g. SP6, T3 or T7) often flank cloning sites; this enables single-stranded probes from the termini of the inserted DNA to be generated. This can be carried out *in vitro* when the construct is incubated with the corresponding polymerase and ribonucleotides.

2. *Overexpression*. For overexpression of a gene, for example, as a prelude to protein purification, that gene can be cloned into a vector that has a known strong promoter sequence, such as T7 polymerase promoter, P_{tac}, P_{araBAD} or P_{CMV}, just upstream of the cloning site. These promoters are often inducible such that expression can be induced at the right time. In some instances the overexpressed protein can accumulate to over 50% of the total cell protein.

In-frame genes encoding easily purifiable proteins

To facilitate purification of an overexpressed protein, the corresponding gene can be cloned in-frame with sequences that code for an easily purifiable protein or peptide such as glutathione-*S*-transferase (GST), maltose binding protein (MalE) or polyhistidine peptide (His-TAG). This results in the overexpression of a fusion protein comprising the protein under study and the easy to purify protein, which can be at either the *N*- or *C*-terminus depending on vector and cloning site used. Subsequently this fusion can be readily purified by virtue of the characteristics of the particular fusion partner. His-TAG, which is probably the most commonly used fusion, readily binds nickel ions, and so His-TAGged fusions can be purified by affinity chromatography on immobilized nickel, washed, then subsequently eluted with imidazole. The fusion partner often has no effect on the activity of the protein and so is not normally removed, though some vectors incorporate protease cleavage sites such as that for factor Xa, at the fusion junction.

loxP sites

loxP is the recognition sequence for the bacteriophage P1 recombinase enzyme, Cre, which catalyzes recombination between sequences that are flanked by such sites. By cloning DNA into vectors that possess loxP sites that DNA can, for example, be recombined into another vector containing loxP sites or into the host chromosome, when in the presence of either the Cre recombinase *in vitro* or a plasmid carrying the *cre* gene *in vivo*.

cos sites

cos sites are the recognition sequences for phage packaging systems. The DNA present between two such sites, provided it is within the packaging limits for that system, will be packaged into phage particles.

Types of vectors

An outline of the features of each vector type is given in **Table 2**.

Plasmids

Plasmids are the most widely used vectors. This is probably because of the ease with which plasmid DNA can be isolated from the host cell (see Sambrook and Russell, 2001) and the variety of plasmids that are available. They are double-stranded independent replicons that range in size from approximately 2 to 10 kb and which are capable of accepting DNA inserts up to about 12 kb, though up to 5 kb is perhaps more common. **Figure 3** shows the structure of pUC18, a typical plasmid that replicates in *E. coli*.

M13 bacteriophage

M13 is a filamentous bacteriophage, with a 6.4 kb single-stranded genome, that infects *E. coli*. An infected cell produces 100–200 double-stranded (or replicative form (RF)) copies of the phage genome. These encode genes that coordinate packaging of one of the phage genome DNA strands into a proteinaceous viral coat and their extrusion from the cell where they are then free to infect further cells. A series of

Table 2 A summary of the properties of the various classes of vectors used in DNA cloning

| Vector	Cloning capacity	Use	Common examples	Method of introduction into cell	Type of growth on agar plate	Copy number	Original reference
Plasmids	Up to 12 kb	Multiple (e.g. mutagenesis, overexpression, sequencing, screening, manipulation)	pUC, pBR322, pGEM	Transformation or electroporation into bacteria. Various methods for eukaryotes (see text)	Colonies	1–700	Cohen *et al.* (1973) *PNAS* **70**: 3240–3244
M13	Normally up to 6 kb	Probe generation, sequencing, *in vitro* mutagenesis	M13mp18	Transfection	Plaques	100–200	Messing *et al.* (1977) *PNAS* **74**: 3642–3646
Bacteriophage lambda (λ)	9–23 kb	Genomic and cDNA libaries, expression screening	λZAP, λFIX	Packaging of construct followed by infection	Plaques	N/A	Murray and Murray (1974) *Nature* **251**: 476–481
Cosmids	Up to 47 kb	Genomic libraries, genome mapping	Supercos, lawrist, tropist	Packaging of construct followed by infection	Colonies	Up to 50	Collins and Hohn (1978) *PNAS* **75**: 4242–4246
Fosmids	Up to 43 kb	Genomic libraries, genome mapping	pFOS1	Packaging of construct followed by infection	Colonies	1–2	Kim *et al.* (1992) *NAR* **20**: 1083–1085
P1 bacteriophage	30–100 kb	Genomic libraries, genome mapping	pAd10*sac*BII	Packaging of construct followed by infection	Colonies	1–2	Sternberg (1990) *PNAS* **87**: 103–107
BACs	Up to 300 kb	Genomic libraries, genome mapping	pBACe3.6, pBeloBACII	Electroporation	Colonies	1–2	Shizuya *et al.* (1992) *PNAS* **89**: 2629–2633
PACs	Up to 300 kb	Genomic libraries, genome mapping	pCYPAC2	Electroporation	Colonies	1–2	Ioannou *et al.* (1994) *Nature Genetics* **6**: 84–89
YACs	20–2000 kb	Genomic libraries, genome mapping	pYAC4	Electroporation, spheroplast absorption, or lithium-mediated transformation	Colonies	1	Burke *et al.* (1987) *Science* **236**: 806–812
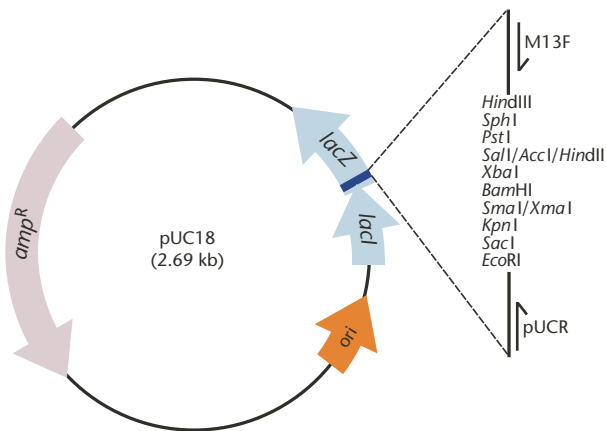
Figure 3 Diagram of the plasmid cloning vector, pUC18. This 2.69 kb double-stranded, closed-circular DNA molecule has an origin of replication (ori) allowing multicopy propagation in *E. coli*, a β-lactamase gene (*amp^R*) giving resistance to the antibiotic ampicillin and a *lacI* gene, the product of which induces the expression of the *lacZ* gene, which in turn encodes β-galactosidase. The multicloning site (dark bar, shown expanded to the right) is close to the 5′ end of the *lacZ* gene, such that insertion at any of the points within the multicloning site (MCS) should disrupt β-galactosidase expression, thereby allowing blue/white recombinant selection on media containing the colorimetric substrate X-gal and inducer isopropyl-β-D-thiogalactopyranoside (IPTG). The MCS contains closely nested restriction sites for a number of enzymes, in the order shown. It is flanked by sequences to which the M13 forward and pUC reverse primers bind, facilitating polymerase chain reaction amplification or sequencing of the inserted DNA.

M13-based cloning vectors have been developed (Messing, 1983), and found capable of accepting inserts several times larger than the M13 genome. However, DNA in M13 can be unstable and so insert sizes are normally kept below that of the vector and this system is not used for the longer-term maintenance of DNA. M13 has no selectable marker; instead infected cells are identified by their ability to form zones of clearing, or 'plaques'. If bacterial cells are mixed with a soft agar and allowed to grow on top of a normal agar plate, the cells in the soft agar will grow and give the agar layer an opaque appearance. Infected cells grow a lot more slowly than noninfected ones and so can be visualized as a plaque in the bacterial lawn.

Since the extruded phage contain copies of one particular DNA strand, this system has proved very useful for techniques that utilize a single-stranded template such as sequencing, site-directed mutagenesis and producing radiolabeled probe DNA for hybridization.

Bacteriophage lambda

Lambda (λ) is a naturally occurring phage virus of *E. coli*. It has a 48 kb linear double-stranded genome, with 12-base cohesive ends (cos sites) at each 5′ end, inside an icosahedral proteinaceous head to which is attached a long noncontractile tail. Infection occurs as a result of the tail binding to the outer membrane maltose receptor and injection of the phage genome through the tail and into the cell. Once in the cell the DNA can follow a lytic or lysogenic pathway depending on environmental or intracellular conditions and the genotype of the individual phage. In the lysogenic pathway, the phage DNA is integrated into the *E. coli* genome providing the geneticist with a tool to integrate DNA into genomic DNA. In the lytic pathway, the DNA is replicated under a rolling circle mechanism to produce a long concatemer of around 100 lambda genome units. Lambda genes encode head and tail proteins and an enzyme that recognizes the cos sites and cleaves the concatemer into individual genomes. Subsequently each 48 kb unit is packaged into lambda heads, the tails attach and the infected cell lyses to release approximately 100 infective phage. In order to be packaged the cos sites need to be around 38–52 kb apart and so different lambda vectors have been produced where nonessential regions have been deleted to make space for DNA to be inserted. A typical lambda vector λZAPII is shown in **Figure 4**. There are two main advantages to using lambda vectors:

- the packaging and infection process is very efficient (perhaps 100 times more efficient than transformation/transfection); and
- since the cells die during lysis, lambda can maintain some inserts that carry cytotoxic or unstable/poorly tolerated sequences.

Cosmids/fosmids

Cosmids were created by engineering cos sites into plasmid vectors to create hybrid vectors with the high-efficiency packaging/infection advantages of the lambda system and the ease of isolation advantage of plasmids. In recent years, cosmids have been superseded as the vector of choice for creating genomic DNA libraries by artificial chromosome vectors that can accommodate larger inserts. Since they are maintained at quite a high copy number in the cell, cosmid clones can become unstable and are often prone to deletion. To overcome this, fosmids, which are based on the single-copy bacterial F factor, have been developed.

P1 bacteriophage vectors

P1 is another naturally occurring *E. coli* phage. It has a linear 110 kb genome that upon infection is circularized through the action of the P1 encoded Cre recombinase which catalyzes recombination at loxP recognition sites. Vectors based upon P1 that can
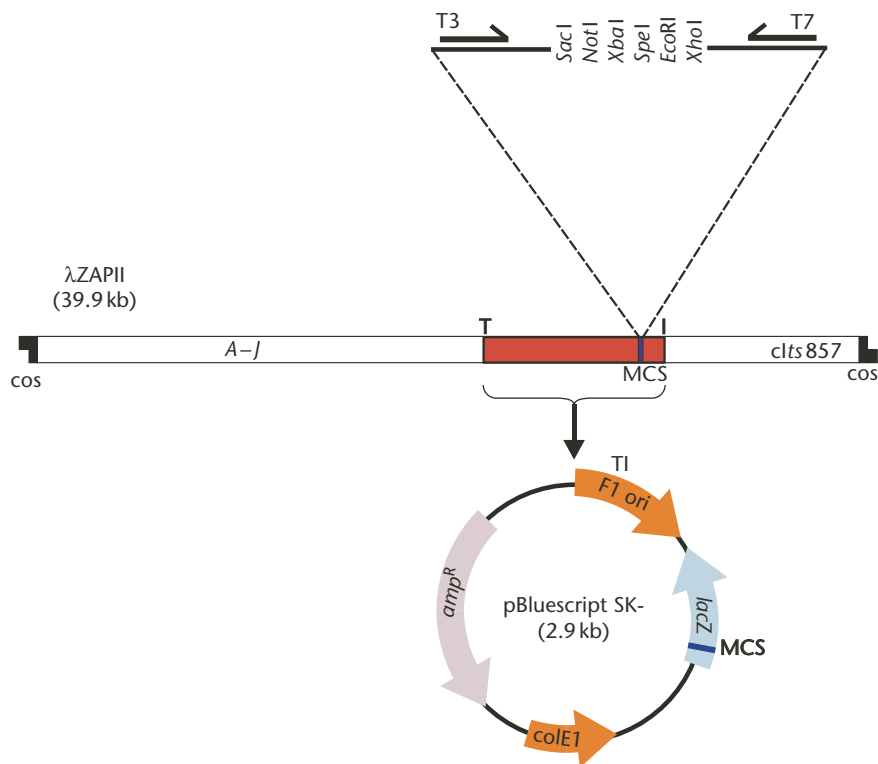
Figure 4 Diagrammatic illustration of the bacteriophage lambda cloning vector λZAPII (Stratagene). This 39.9 kb linear DNA molecule can accept inserts up to 10 kb. It has a multicloning site (MCS) (expanded above) flanked by T7 and T3 polymerase promoters to which T7 and T3 primers bind. The MCS is within the *lacZ* gene allowing blue/white selection of inserts. The region of the vector containing the insert can be excised to the plasmid pBluescript (SK-) as shown, by infection of cells harboring λZAPII with a helper phage. λZAPII also contains terminal cos sites, lambda genes A–J and a temperature-sensitive mutation within the cI repressor that represses lysis.

maintain inserts of 70–100 kb have been developed (Sternberg, 1990). Advantages of P1 vectors include:

- stability of inserts, because the vector is single copy;
- DNA is present as a closed circle within the cell and so can be isolated using a standard alkaline lysis miniprep; and
- constructs are introduced into the cell by the highly efficient process of packaging and infection.

Bacterial artificial chromosomes

BACs are vectors based on the *E. coli* F factor, which is normally present at one to two copies per cell. They can accept inserts of up to 300 kb and can be directly introduced into *deoR* strains of *E. coli* by electroporation, thus avoiding the use of packaging extracts. Their stability and the ease by which clone DNA can be miniprepped for analysis, or end sequencing, make BACs and P1 artificial chromosomes (PACs) (see below) the current vectors of choice for most whole-genome libraries. For a detailed experimental description of BAC library construction see Osoegawa *et al.*

(1998). A diagram of a typical BAC vector, pBACe3.6, is shown in **Figure 5**.

P1 artificial chromosomes

PACs are a hybrid of P1 bacteriophage and BAC vectors. They are low copy number, have good recombinant selection and can be introduced into the cell by electroporation.

Yeast artificial chromosomes

Yeast artificial chromosome (YAC) vectors are shuttle vectors containing yeast telomeres, centromere, an origin of replication and selectable markers allowing replication in yeast and *E. coli*. They can accommodate inserts of up to 2 megabases (Mb). The resulting constructs are transformed into yeast by spheroplast absorption and therein behave like an extra chromosome. *Saccharomyces cerevisiae* is quite tolerant of extreme base composition and repeats, and so YACs are often useful for cloning sequences that are unclonable in *E. coli*.
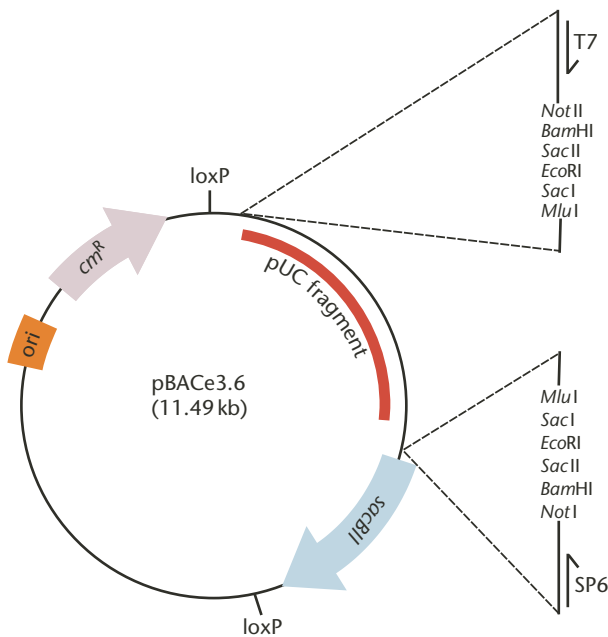
Figure 5 Diagrammatic illustration of the bacterial artificial chromosome (BAC) cloning vector pBACe3.6. This 11.49 kb double-stranded, closed-circular DNA molecule has an origin of replication (ori) allowing single-copy propagation in *E. coli*, a gene encoding resistance to the antibiotic chloramphenicol (*cm^R*) and loxP recombination sites that flank the region into which DNA is inserted. Dual multicloning sites (MCS) flank a pUC stuffer fragment. The parent vector possesses this region to allow multicopy growth in *E. coli* allowing easy preparation of the vector. Restriction of the vector with any of the MCS restriction enzymes excises this fragment. The MCS is flanked by T7 and T3 polymerase promoters, to which T7 and T3 primers bind, and is located between the *sacBII* gene and its promoter so that inserted DNA disrupts expression of its cytotoxic product, ensuring that only recombinants grow on media containing sucrose.

Human artificial chromosomes

With the goal of creating stable transgenic mammalian cell lines, researchers have been developing mammalian artificial chromosomes (MACs) and human artificial chromosomes (HACs). This was achieved in 1997 when Harrington *et al.* (1997) reported the construction and maintenance, in a human cell line, of human artificial chromosomes containing 0.3–2 Mb of inserted α-satellite DNA.

Introduction of Deoxyribonucleic Acid into the Cell

Although some bacteria can spontaneously take up DNA, this is not a widespread occurrence and so a number of methods have been developed for the directed transfer of cloned DNA into the cell.

Transfer into bacteria

There are four common methods for DNA transfer into bacteria (see Sambrook *et al.*, 1989; Hanahan *et al.*, 1991).

Transformation

Transformation is the process of plasmid uptake by bacteria. This process is typically chemically induced with the most common method involving treatment with calcium chloride. Here log phase cells are washed and resuspended in 0.1 M calcium chloride. These cells are incubated on ice with the DNA whereupon a calcium ion–DNA complex is formed on the outside of the cell. If this mixture is subsequently heat-shocked by incubating at 42°C for 60–90 s, the cell then takes up this complex. Typically 0.01–0.1% of the DNA is introduced generating up to 10^8 transformants per microgram of pUC18 DNA.

Electroporation

Application of a brief high-voltage electric shock to a suspension of cells transiently induces the presence of pores in the cell membrane through which DNA can enter the cell. Electroporation can be highly efficient giving more than 10^{10} transformants per microgram of pUC18 DNA.

Transfection

Transfection is defined as the transformation by viral or phage DNA, and can be achieved by either of the previous methods.

Infection

Infection of bacteria can reach efficiencies close to 100%. Typically cells are first grown under conditions that favor induction of the phage receptor; for example *E. coli* to be infected with phage lambda are cultured in media containing magnesium ions and maltose. Prepared cells can simply be incubated with phage particles, typically at 37°C for 20–30 min, during which time the phage are absorbed.

Transfer into eukaryotic cells

The introduction of DNA into eukaryotic cells is technically more demanding and often less efficient than transfer into bacteria. Nevertheless, a number of methods have been developed (Ausubel *et al.*, 1999; James and Grosveld, 1987) involving calcium phosphate-mediated transfection, DEAE dextran transfection, electroporation, microinjection, infection (viral vectors), electroporesis, protoplast/spheroplast fusion and lipofection. The method used is dependent on the type of genetic material being handled, the

reason that DNA is being introduced and more critically the type of cell being transformed.

Epilog

Cloning is at the heart of molecular biology. It can be immensely rewarding, or just as easily extremely frustrating. The key to success is usually to be found in gaining a full understanding of the system to be used and in careful experimental planning. As a starting point the beginner is directed to the references and suggestions for further reading cited below. Happy cloning!

See also

Mammalian Artificial Chromosomes (MACs)
Recombinant DNA
Yeast Artifical Chromosome (YAC) Clones

References

Ausubel FM, Brent R, Kingston RE, *et al*. (eds.) (1999) Introduction of DNA into mammalian cells. *Current Protocols in Molecular Biology*, chap. 9. New York, NY: John Wiley.

Hanahan D, Jessee J and Bloom FR (1991) Plasmid transformation of *Escherichia coli* and other bacteria. *Methods in Enzymology* **204**: 63–113.

Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K and Willard HF (1997) Formation of *de novo* centromeres and construction of first-generation human artificial minichromosomes. *Nature Genetics* **15**: 345–355.

James RFL and Grosveld FG (1987) DNA-mediated gene transfer into mammalian cells. In: Walker JM and Gaastraxs W (eds.) *Techniques in Molecular Biology*, pp. 187–202. London, UK: Croom Helm.

Messing J (1983) New M13 vectors for cloning. *Methods in Enzymology* **101**: 20–78.

Osoegawa K, Yeong Woon P, Zhao B, *et al.* (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**: 1–8.

Sambrook J and Russell DW (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sternberg N (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 103–107.

Further Reading

Birren B, Green ED, Klapholz S, Myers RM and Roskams J (eds.) (1997) *Genome Analysis: A Laboratory Manual. Volume 1, Analysing DNA*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Brown TA (2000) *Essential Molecular Biology: A Practical Approach*. Oxford, UK: Oxford University Press.

Markie D (ed.) (1996) YAC protocols. *Methods in Molecular Biology*, vol. 54. Totowa, NJ: Humana Press Inc.

Web Links

ATCC (American Type Culture Collection). A global bioresource center
http://www.atcc.org