



Cloning and Gene Manipulation

2

Aasna L. Parui, Lalith K. Chaganti, Rucha Kulkarni, and Kakoli Bose

Abstract

Gene cloning has continually been one of the many important discoveries in the areas of molecular and cell biology. The advent of recombinant DNA technology has revolutionized the field of biology in many ways, as it allowed the scientific community to understand the functional and cellular aspects for a wide array of essential genes. In brief, gene cloning involves inserting the DNA fragment of interest in a suitable vector and amplifying this recombinant molecule in host cells such as *E. coli*. However, the basic procedure also involves numerous steps such as restriction enzyme digestion and ligation reactions to be carried out in order to obtain a perfect recombinant DNA molecule. The purpose of this chapter is to present an overview of the cloning process, with the details of various steps involved and the techniques used in the identification of the recombinant clones.

Keywords

Cloning · Recombinant DNA · Restriction digestion · Ligation · PCR

A. L. Parui · R. Kulkarni · K. Bose (✉)

Integrated Biophysics and Structural Biology Lab, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Mumbai, India

Homi Bhabha National Institute, BARC Training School Complex, Mumbai, India

e-mail: kbose@actrec.gov.in

L. K. Chaganti

Integrated Biophysics and Structural Biology Lab, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Mumbai, India

© The Author(s) 2022

K. Bose (ed.), *Textbook on Cloning, Expression and Purification of Recombinant Proteins*, https://doi.org/10.1007/978-981-16-4987-5_2

2.1 Introduction

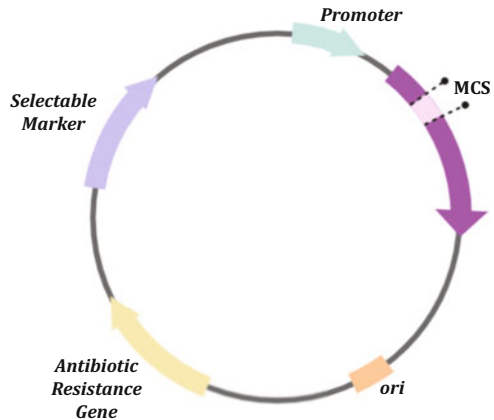
The word “cloning” refers to the asexual reproduction required to obtain organisms that are genetically identical to one another and to their parents. This process is in contrast to the sexual reproduction where offspring are not identical. The cloning method involves generation of large population of cells with identical DNA molecules obtained from a single living cell by a process of replication of the original single DNA molecule. The word “cloning” is also applied to genes, which is an extension of this concept in molecular biology [1].

Gene or DNA cloning is a common practice used by researchers to create exact copies (clones) of a particular gene or a DNA sequence using recombinant genetic engineering techniques. Major breakthrough in the cloning experiments was obtained by Herbert Boyer, Stanley Cohen, Paul Berg, and their colleagues in the early 1970s [2]. The traditional technique for gene cloning involves transfer of a target DNA or gene fragment from one organism to different cloning vectors (described in Chap. 3) or autonomously replicating genetic element, such as bacterial plasmids (small, circular piece of extra chromosomal DNA) and bacteriophages, which serve as mediums to propagate the cloned DNA within the cell. Apart from bacteria, plasmids are also naturally present in archaea and eukaryotes such as yeast and plants. They provide additional benefits to the organisms by conferring properties such as antibiotic resistance, virulence, and degradative abilities. The plasmids also contain an origin of replication (*ori*), which helps in controlling the host range and copy number of the plasmid within the host. For cloning purposes, the plasmids have been designed artificially in such a way that DNA of interest can be easily inserted into these vectors and propagated substantially for various purposes. Apart from the *ori* site and antibiotic resistance gene, the lab-engineered plasmids have a Multiple Cloning Site (MCS—short segment containing various restriction enzyme sites required for easy insertion of the gene of interest), a promoter region (allows transcription of the downstream gene, present especially in expression plasmids), and a selectable marker gene (Fig. 2.1). These segments in a plasmid help ease the task of modifying the plasmids based on different experimental requirements, thus making them an attractive tool for molecular cloning.

Further, the recombinant vector is transferred into suitable host cells such as *E. coli* for production of multiple DNA copies. Plasmids that have an antibiotic resistance gene are typically employed in DNA cloning or bacterial transformation. The presence of the antibiotic resistance gene allows for the selection of bacteria harboring the desired plasmid. Thus, bacteria with the recombinant plasmid will thrive in an antibiotic-containing media while plasmid deficient bacteria will not be resistant to the antibiotic and fail to survive (Fig. 2.2). This method helps the cells containing the desired recombinant DNA to be distinguished from the others, and their selection becomes possible [3].

Gene cloning technique is used for several downstream applications, such as DNA sequencing, mutagenesis, genotyping, or heterologous protein expression. However, precision in the development of the basic cloning steps plays key roles behind all these applications. The recent technological advancements in molecular

Fig. 2.1 Generalized plasmid map. A plasmid contains an *ori* (origin of replication) site, an antibiotic resistance gene (required for selective propagation of only plasmid-containing cells), and a selectable marker. The promoter region is present upstream of the MCS (multiple cloning site) where the gene of interest is inserted



genetics have allowed scientists to study, develop, and explore various modifications in the genomes for a wide range of organisms. Foreign DNA may now be introduced in bacterial plasmids with the use of restriction endonucleases (described in Sect. 2.4.1) and can be replicated further. Bacterial cells transformed with this foreign DNA can now express the genetic information and make suitable products encoded by the desired genes. Thus, by molecular cloning, we can learn a lot about the structure and *modus operandi* of different genes. Moreover, production of bulk amount of specific gene products including unique as well as rare proteins has become industrially feasible. We can also use such plasmids to transform the genetic constitution of other organisms. In this chapter, we will discuss the general strategies and principles of gene cloning as well as genetic engineering tools that can be used for a wide range of research purposes with a focus on their applications in recombinant DNA technology.

2.2 DNA Libraries

Acquiring the genetic information has become a major step for any field of biological sciences. This requires navigation through the complete genomic sequence of a specific organism, for either understanding the function of a particular gene of interest or the relevance of the entire genome. A DNA or gene library is a compilation of cloned DNA fragments that collectively represent the genes isolated from a particular organism. This DNA fragment collection may then be utilized for the identification of specific genes and other DNA sequences of interest, which is analogous to selecting desired books from any conventional library. Specific DNA fragments are generated by digesting the genome or genes with the help of specific restriction enzymes. The generated fragments are further cloned into specific plasmid vectors, and then transformed into suitable host cells [4]. The total number of all

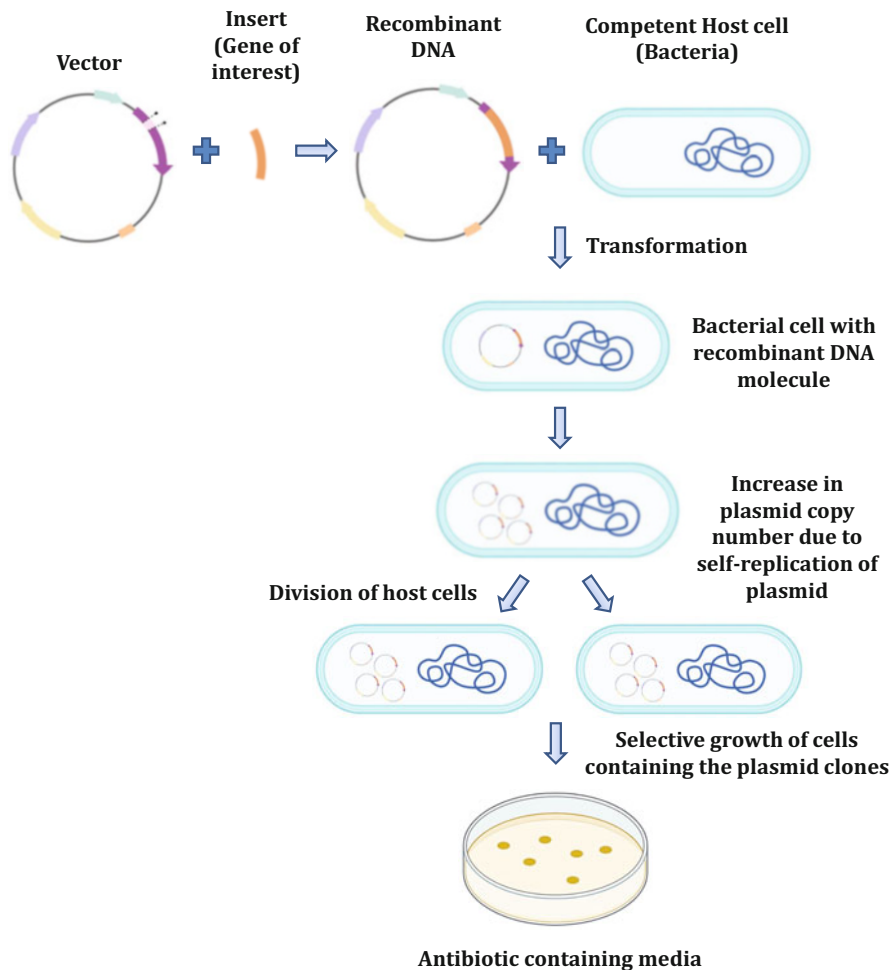


Fig. 2.2 Steps in gene cloning. The gene of interest is inserted into a suitable vector at the MCS site and this recombinant DNA molecule is then transformed into a compatible bacterial host. The *ori* site promotes the replication of recombinant plasmid inside the host. After numerous divisions of the transformed cells, the recombinant clones are obtained on the selective media supplemented with the antibiotic. Cells transformed with the recombinant plasmid will only grow on the selective medium

DNA molecules of a particular genome/field of interest makes up that particular library. The target or specific DNA from the library is further screened with a molecular probe. Once prepared, the library can be propagated indefinitely in the host cells and can be readily retrieved whenever a new probe is available to seek out a particular fragment from the entire library.

There are two types of DNA libraries that can be used to isolate specific DNAs: (1) genomic library and (2) cDNA library. The choice of the particular type of gene library depends on factors such as protein production from a specific gene and studying genetic architecture.

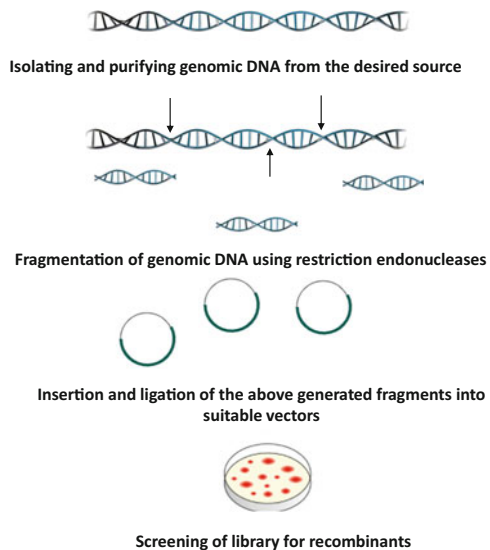
2.2.1 Genomic Library

A genomic library is a collection of clones that contain DNA fragments representing the total genomic DNA of a specific organism of interest. Depending upon the organism and size of its genome, this library can be prepared either in bacterial plasmids, phage vectors, cosmids, bacterial artificial chromosome (BAC), or a yeast artificial chromosome (YAC). Chapter 3 of this book elaborates on different vectors and discusses the importance of choosing them for distinct cloning purposes. A detailed outline of the construction of genomic library is schematically represented in Fig. 2.3. The steps involved in construction of a genomic library are:

1. Isolation and purification of genomic DNA:

The first step in construction of any genomic library requires isolation of complete genomic DNA from the organism of interest (bacteria, virus, plants, or animals). Depending on the type of organism, the procedures engaged in the isolation of genomic DNA vary widely. In eukaryotes, genomic DNA can be prepared either from nuclear DNA or any organelle-specific DNA. Nuclear genomic library is prepared by specifically isolating the DNA from the nucleus. The eukaryotic cell nuclei are purified by digestion with protease and organic (phenol-chloroform) extraction. In case of organelle genomic library, first the

Fig. 2.3 Preparation of a Genomic library. Genomic DNA is isolated from the organism of interest using methods such as Phenol-Chloroform extraction [7]. It is further subjected to random fragmentation using either physical (sonication) or enzymatic (restriction endonucleases) methods. The fragmented DNAs are cloned into suitable vector and the transformed recombinants are then selected under appropriate selection pressure conditions. The target DNA is screened from the recombinant clones using methods such as autoradiography and PCR



respective organelle is purified and then the DNA is isolated only from that particular purified organelle. Organelle separation procedures vary for different organelles.

2. Fragmentation and restriction digestion of genomic DNA:

The isolated genomic DNA is very long and needs to be cut into fragments of ideal sizes. This can be achieved either by fragmentation or by enzymatic digestion. Physical methods include pipetting the DNA molecule or applying intensified ultrasound waves (sonication), whereas the enzymatic method involves the use of restriction enzymes as described in Sect. 2.4. Generation of DNA fragments of various sizes depends on the distribution probability of specific restriction enzyme site within a gene. Therefore, complete digestion of the genomic DNA generates very short fragments of variable sizes depending on the presence of the restriction enzyme site in the entire sequence. As a consequence, the desired gene of interest in its complete form might not be represented within a library. Therefore, partial restriction enzyme digestion is usually employed to generate overlapping fragments containing one or more gene [4]. The generated fragments are then purified by either gel electrophoresis or density gradient centrifugation techniques, which are then further cloned into a suitable vector.

3. Ligation of the DNA fragments:

The third step is to insert the generated DNA fragments into a suitable vector as shown in Fig. 2.3 below. Different vectors such as plasmids, λ phage, YAC, and BAC (described in Chap. 3) are used for cloning the DNA fragments. YAC (up to 2000 kb) and BAC (up to 300 kb) are considered suitable vectors for cloning larger DNA molecules [5]. However, it is difficult to clone a large insert into these vectors; therefore, bacteriophage λ or cosmid vectors are usually employed for generating genomic libraries. Since a larger insert size (up to 40 kb) can be accommodated by these vectors compared to plasmids (~10 kb), there is a greater chance of cloning a gene sequence with both the coding sequence and regulatory elements in a single clone. T4 DNA ligase is typically used for ligating the selected DNA sequences into the vectors. Details about the steps involved in ligation are discussed in the Sect. 2.5.

Selection of the number of clones required for construction of a genomic library is the most important step. One should ensure that the constructed library is a representative of the entire genome. However, since any genomic insert generated by a particular restriction enzyme has an equal chance of being in the library compared to any other insert, the number of clones to be pooled depends on the size of the organism's genome " f " and the average insert size. The probability (P) of including any DNA sequence in a random library of (N) independent recombinants is represented by Eq. (2.1):

$$N = \ln(1 - P) / \ln(1 - f) \quad (2.1)$$

where

f = total genome length/average insert size.

N = necessary number of recombinants.

P = desired probability that any fragment in the genome will be present.

f = fractional proportion of the genome in a single recombinant.

Thus, bigger the library, greater is the chance of finding a gene in that particular library. On the contrary, increasing the insert size would allow fewer clones that are needed to represent a genome.

4. Library screening:

A common method employed to screen the library is colony hybridization. Each transformed host cell of a library will have just one vector with one insert of DNA. First colonies of host cells carrying the plasmid or phage libraries are plated onto an agar plate with a suitable antibiotic such as ampicillin. This will ensure growth of only those cells that are transformed with vectors containing antibiotic resistance gene. The colonies are then transferred onto a nitro cellulose membrane for further processing. Once the cells are attached to the membrane, they are lysed, deproteinized (to avoid protein contamination), and the released DNA is denatured by alkaline treatment. Later, hybridization is performed between the target DNA and labelled DNA probe (complementary sequence to the target DNA). The target DNA can then be identified by autoradiography. Polymerase chain reaction (PCR) and immunological screening can also be used as alternatives to colony hybridization [6].

PCR screening is generally used to identify uncommon DNA sequences among diverse cocktails of molecular clones by increasing the quantity of a specific sequence. The library is plated as plaques or colonies on agar plates and individually these colonies are inoculated into the wells of the multi-well plate. PCR reactions are performed with primers flanked by a unique target sequence to identify the clone of interest. This method is applicable upon availability of detailed gene sequence for designing of typical primers.

Immunological screening includes the use of antibodies that identify antigenic determinants on polypeptide specifically. It does not depend on the function of the foreign protein produced, instead requires a protein-specific antibody. This screening technique is similar to colony hybridization; however, instead of using labelled DNA probe, antibodies are used to specifically detect the target protein (Table 2.1).

2.2.1.1 Applications

Genomic libraries can be used for many purposes:

- The whole genomic sequence of an organism can be produced.
- Serves as a repository for genomic sequences for the development of transgenic animals.

Table 2.1 List of vectors used for generation of DNA libraries

Vectors	Origin	Insert size (kb)	Properties and applications
Plasmid	Naturally occurring multicopy Circular DNA	10	2–6 kb size, commonly used for molecular cloning, contains origin of replication, reporter and antibiotic resistance genes
Phage	Bacteriophage λ	10–20	Genome size is 47 kb, contains double-stranded DNA (dsDNA) and infects <i>E. coli</i> , efficient packaging system, used to study individual genes
Cosmid	Plasmid containing a bacteriophage λ cos site	40	Contains a λ phage cos site, which allows packaging, transmits as plasmids in <i>E. coli</i> , helpful for subcloning YAC, BAC, PAC, etc.
BAC	<i>E. coli</i> F factor plasmid	300	Circular DNA molecules, up to 7 kb length, have F-plasmid, high-capacity vectors used for investigation of bulky genomes
YAC	Yeast (<i>Saccharomyces cerevisiae</i>) centromere, telomere, and autonomously replicating sequence	2000	Linear DNA, have properties of a yeast chromosome, contains selection markers for identification of successful transformants that are used for analysis of large genomes

- The structure of a given chromosome can be investigated.
- Genomic libraries from higher eukaryotes are important to study untranslated regions (regulatory elements) of a gene, including promoters or introns.
- In prokaryotes, genomic libraries are used to clone relatively smaller gene fragments.

2.2.2 cDNA Library

cDNA library is a collection of complementary DNA (cDNA) fragments which have been cloned individually into separate vector molecules. In cDNA libraries, DNA copies complementary to the transcribed RNA sequences (usually the mature mRNA) of an organism are produced by the reverse transcription of RNA by the reverse transcriptase enzyme [8, 9]. Thus, these cDNA libraries contain only the coding sequences generated from the fully transcribed and spliced mRNA produced from the expressed genes (exons). Unlike the genomic library, these cDNA libraries lack repetitive sequences, introns (non-coding regions), regulatory regions, and enhancers of the gene. Hence, cDNA libraries are prepared primarily from the higher eukaryotes and not from the lower eukaryotes or prokaryotes, which lack these regulatory elements.

2.2.2.1 Construction of a cDNA Library

A detailed outline for construction of a cDNA library is described below:

(a) Initial extraction and purification of mRNA:

This step involves the isolation of total mRNA from the cells. Eukaryotic mRNA consists of 50–250 adenylate residues (poly-A tail) at the 3' end, which facilitates simple separation of mRNA through affinity chromatography using oligo(dT). Chromatographic column or Magnetic beads coupled with oligo(dT) are usually used to purify mRNA from the much more prevalent rRNAs and tRNAs in a cell lysate. The poly-A tail at the 3' end of the mRNA enables its efficient binding to the oligo(dT) beads. After providing sufficient washes to remove the impurities, these mRNA can then be eluted using strong magnetic force or low salt buffer; the bound mRNA is isolated from the total RNA content (Fig. 2.4). The recovered mRNA is then analyzed by agarose gel electrophoresis, before using it as a template for cDNA synthesis [10].

(b) Production of cDNA:

Once mRNA is extracted, the complementary DNA strand is synthesized using reverse transcriptase enzyme to make mRNA:DNA duplex. Herein, a short oligo (dT) primer with free 3'-OH is annealed to a poly-A tail of mRNA and the primer is extended by reverse transcriptase to generate the complementary DNA strand (Fig. 2.5). Now, the mRNA template from the mRNA:DNA hybrid is removed by alkaline hydrolysis using an RNase H enzyme and this generates a single-stranded cDNA (ss-cDNA). By producing a short hairpin loop at its 3'

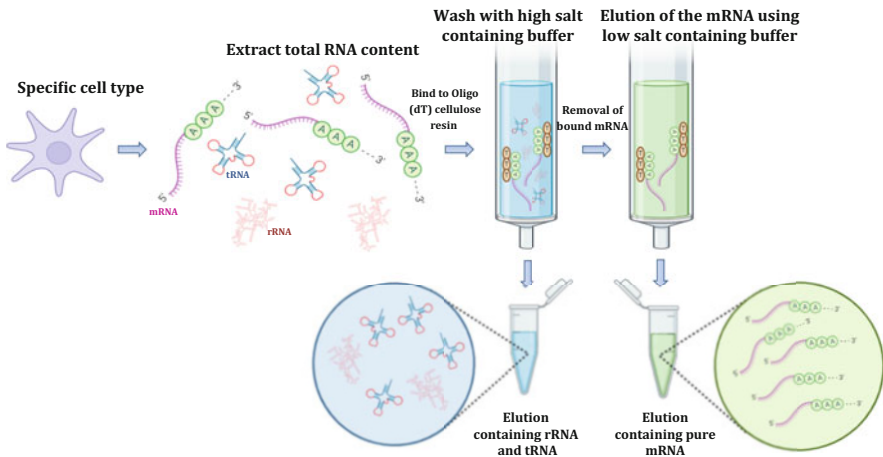


Fig. 2.4 Isolation of mRNA from total RNA content. The poly(A) region at the 3' end of the eukaryotic mRNA allows its selective isolation from total cellular RNA content. It is loaded on an oligo(dT) affinity chromatography column under high salt conditions that promotes hybridization between the 3' poly(A) tails of the mRNA and the oligo(dT)-coupled matrix. The rRNAs and tRNAs are washed out of the column after hybridization and the mRNA is then eluted with a low salt buffer

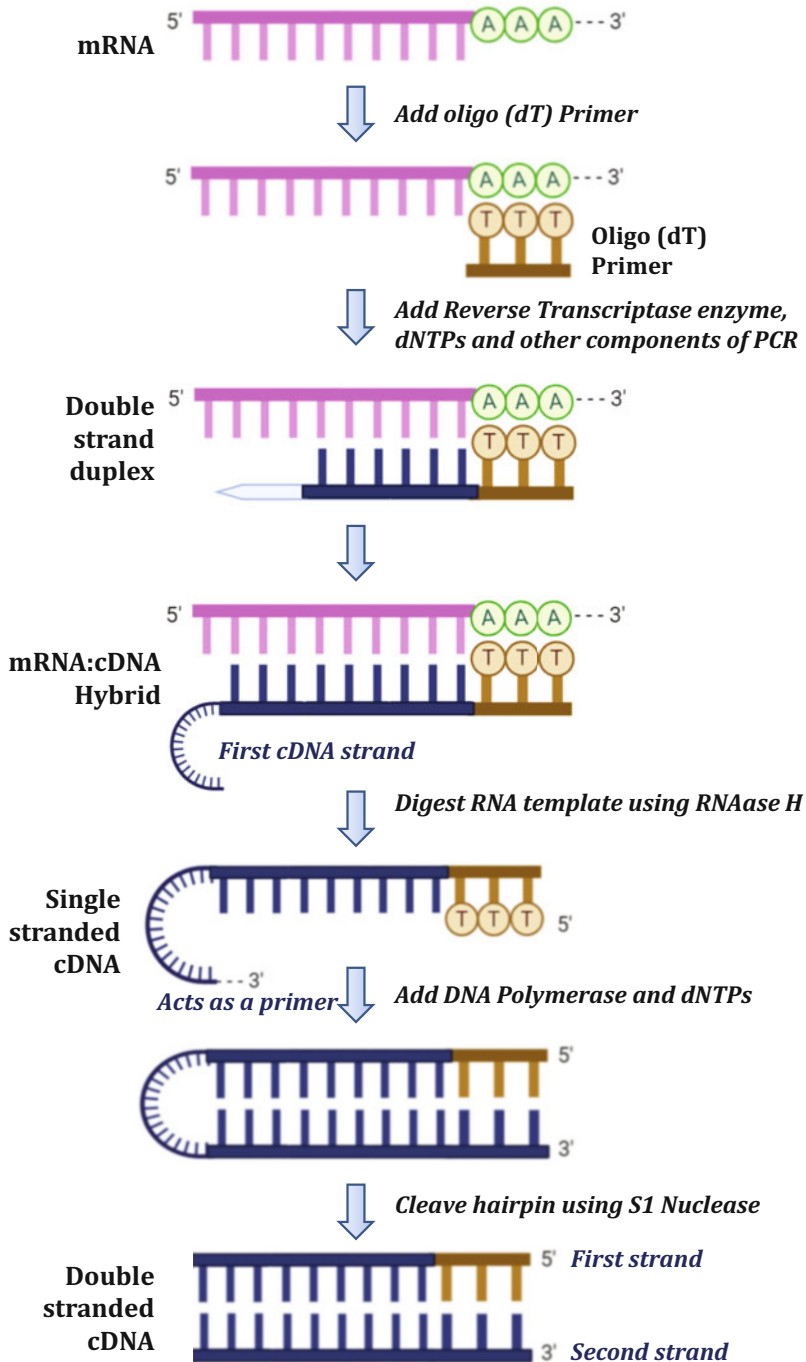


Fig. 2.5 cDNA synthesis. In the presence of dNTPs, the first strand of cDNA is synthesized by reverse transcriptase and oligo (dT) primer. A hybrid mRNA-cDNA is generated, followed by digestion of the mRNA template by alkaline hydrolysis and the enzyme ribonuclease H. The natural

end, ss-cDNA acts as its own primer. Due to the hydrophobicity of the bases, ss-nucleic acid molecules have a tendency to form such secondary structures. The free 3'-OH in the hairpin loop is essential for the generation of the complementary DNA strand. Thus, the ss-cDNA is converted into a double-stranded (ds) cDNA with the help of DNA polymerase. The generated ds-cDNA initially has a hairpin loop at one end. This is then removed by S1 nuclease treatment and the final product is blunt-ended ds-cDNA molecule [4].

(c) Ligation of cDNA into the vector:

The generated cDNAs are cloned into plasmid and bacteriophage vectors; however, plasmids are extensively used for cloning and isolation of the desired cDNAs. These ds-cDNAs are ligated into appropriate vector either by using a blunt end ligation or by adding linkers to ds-cDNA ends. Due to the inefficiency of blunt-end ligation, small restriction-site linkers are initially ligated to both ends before cloning into any suitable vector. In this method, 10–12 base pair (bp) long hybridizing complementary oligonucleotide linkers with a restriction enzyme site is ligated to the ds-cDNA ends using T4 DNA bacteriophage ligase. The resulting ds-cDNAs with linkers at both ends are digested with the respective restriction enzyme to generate cDNAs with sticky ends. Restriction digestion of cDNAs with internal restriction site can be overcome by efficient modification of ds-cDNA with the methylases before adding linkers. This methylation step ensures the protection from the action of restriction enzymes.

(d) Library screening:

Screening of colonies with cDNA is similar to genomic library. The most common methods used are hybridization methods and immunological assays, which are elaborated in the previous sections.

2.2.2.2 Applications of cDNA Library

cDNA libraries can be used for many purposes:

- Unlike the genomic DNA libraries, cDNA can be directly expressed in prokaryotic organisms.
- Discovery of Novel genes.
- Storage of less information as a result of elimination of the non-coding regions.
- cDNAs are used for in vitro study of gene function.
- A cDNA library is useful for isolating genes that code for specific mRNAs.
- cDNA libraries are also useful to identify the tissue-specific mRNAs, where certain genes are expressed only in one cell type but not in the other.
- cDNA libraries are important in reverse genetics, where more genomic information obtained from genomic libraries is of less use.



Fig. 2.5 (continued) hairpin of the first cDNA strand acts as a primer for the synthesis of the second strand. Using a self-priming method, DNA polymerase I catalyzes synthesis of the second strand and further the hairpin is cleaved using S1 nuclease. Double-strand cDNAs corresponding to the many different mRNAs extracted from the cell are formed at the end of this reaction

2.2.3 Difference Between Genomic and cDNA Library (Table 2.2)

Table 2.2 Difference between genomic and cDNA library

Genomic library	cDNA library
Prepared directly from the genomic DNA	Prepared using mRNA as a template
Represents entire genome of the organism	Represents only those genes of genome which express under specific conditions or tissues
Restriction endonucleases and ligases are important for its construction	Reverse transcriptase enzyme plays an important role in its construction
They carry introns, non-coding regions, and other regulatory elements	They lack introns and contain only the coding regions
Represent the DNA of both eukaryotic and prokaryotic organisms	Represent the DNA of only eukaryotic organisms
They are not capable of expression in prokaryotes (such as bacteria) because they carry introns, and prokaryotes do not have machinery to process non-coding regions	They are capable of expression in bacteria because they do not contain the non-coding regions

2.3 Polymerase Chain Reaction (PCR)

2.3.1 Background

PCR is a comparatively straightforward technology that amplifies the DNA template for producing specific DNA fragments *in vitro*. Practically, the conventional ways to clone a DNA sequence into a vector and to replicate it can involve days or weeks, while amplifying the DNA sequences using PCR just takes hours. While a large volume of biological materials are required for most of the biochemical analysis including nucleic acid detection with radioisotopes, the PCR method takes relatively less number of reagents and effort. In a reduced amount of time, PCR is able to accomplish higher sensitivity for detection and amplification levels of particular sequences. The technical characteristics make it highly helpful for use in fundamental as well as commercial research and also in genetic identification testing, forensics, industrial quality control, and *in vitro* diagnostics. Basic PCRs are widely employed in many molecular biology laboratories where DNA fragments are amplified and DNA or RNA sequences are detected from a cell or a particular environment sample. Furthermore, PCR has expanded well ahead of basic amplification and detection, and several extensions were recently made to the original PCR method [4].

2.3.2 Components of PCR

DNA template: The double-stranded DNA (dsDNA) sample containing the specific target sequence for amplification.

DNA polymerase: It is an enzyme that synthesizes new strands of DNA complementary to the target sequence. Of the different types of DNA polymerase enzymes, the first and most commonly used are *Taq* DNA polymerase (from *Thermis aquaticus*) and *Pfu* DNA polymerase (from *Pyrococcus furiosus*). The latter is currently being used widely because of its higher fidelity in copying DNA. These enzymes may be slightly different, yet each possesses two key features that put them pertinent for PCR:

1. They can amplify new DNA strands from a DNA template using specific primers.
2. They are susceptible to higher temperatures.

Primers: These are small ss-DNA sequences that are complementary to the target sequence. The polymerase begins synthesizing new DNA from the 3' free hydroxyl group of the primer.

Nucleotides (dNTPs or deoxynucleotide triphosphates): The four single units of the nucleotide bases, viz., A (Adenine), T (Thymine), G (Guanine), and C (Cytosine) that are essentially the “building blocks” of new DNA strands.

2.3.3 PCR Protocol

Following steps are to be included in a particular PCR experiment (Fig. 2.6):

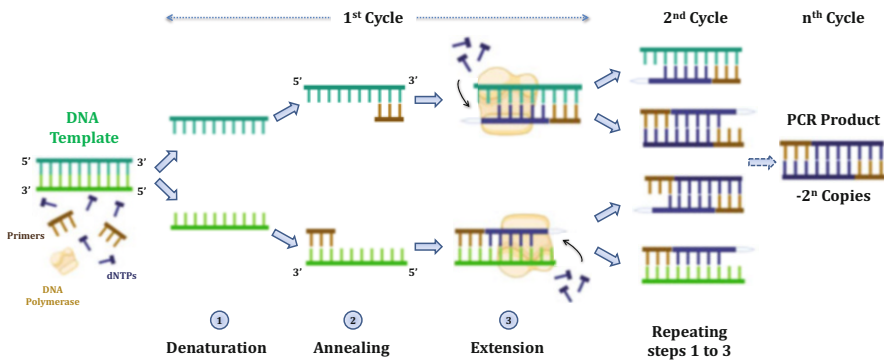


Fig. 2.6 Basic steps of PCR. The dsDNA is denatured into two ssDNA and the respective primers bind at their 3' ends in the annealing step. Extension of the new strand occurs with the help of DNA polymerase and dNTPs. The resulting DNA fragments are again denatured in the next cycle and the three steps are repeated for specific number of cycles to obtain the amplified PCR product

1. **Initial denaturation:** The initial step of PCR includes denaturation of the target DNA by heating it to 95 °C for 5 min. It involves separation of the two intertwined strands of DNA to produce the essential single-stranded DNA (ssDNA) templates.
2. **Annealing:** In the second step of PCR, the reaction temperature is decreased to ~40–60 °C for 15–60 s so that the oligonucleotide primers can bind to the denatured specific target DNA by forming stable and specific associations. Further, these primers serve as the docking site for the DNA polymerase.
3. **Extension:** During this step of PCR, the DNA polymerase synthesizes new complementary DNA strands by binding to the primer. The temperature in extension is usually raised to 72 °C as this is an optimum temperature for most of the DNA polymerases such as *Taq* or *Pfu* that is present in the reaction mixture. Instead of two, a total number of four DNA stands are obtained after the extension step.
4. **Amplification:** During this step, the temperature is increased to 95 °C again. Each of the ds-DNA molecules, comprising one strand of the original molecule and one newly synthesized strand of DNA that were obtained from the previous step, again get denatured into single strands. This begins the second cycle of denaturation–annealing–extension, at the end of which there are eight DNA strands that are obtained.

However, it is to be noted that for every template and primer permutation, each step of the cycle should be optimized individually. If the temperature is comparable between annealing and extension, these two steps can be merged in one step in which both primer annealing and extension can be done. The amplified products may be evaluated for sizes, quantities, and sequences after 20–40 cycles and subsequently employed in other experimental methods.

2.4 Restriction Digestion

Gene cloning requires the recombinant DNA molecule to be cut in a very precise manner such that insertion of the new DNA fragment is only at one particular site. Restriction digestion is a procedure where DNA is cut in appropriate sites using restriction endonucleases [3]. These sites are present only at a particular region in the entire vector (called MCS region) to avoid any unnecessary cuts that would generate various fragments of the same DNA molecule. The target DNA molecule is mixed under specific reaction conditions with restriction enzymes for digestion. These enzymes distinguish and attach to the particular DNA sequences, then cleave at specified nucleotides sequence. Restriction digestion may lead to formation of blunt ends (ends of a DNA molecule that finish in a base pair) or sticky ends (ends of a DNA molecule that have a nucleotide overhang) (Sect. 2.4.1). Restriction digestion is usually the step preceding insertion of a foreign gene into a vector via a process called ligation. The results of a restriction digestion can be analyzed by gel electrophoresis, a process wherein the digested products are separated on the basis of their

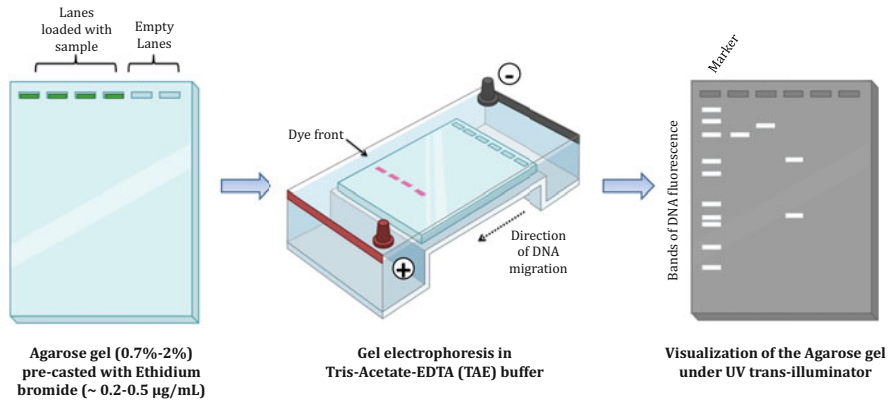


Fig. 2.7 Separation of DNA by Agarose gel electrophoresis. An agarose gel matrix (depending on the DNA size to be separated) containing ethidium bromide (EtBr) is pre-casted on a plastic tray. The DNA samples are mixed with the tracking dye (to determine the extent of DNA migration) and loaded into the wells of the gel. When visualized under UV transilluminator, the intercalated ethidium bromide fluoresces and the molecular weight can be determined from extend of migration

molecule length in a polymer-based gel (agarose). The gel is run against an electric field, where the negatively charged DNA molecules are allowed to travel from the anode to cathode and thus the separation occurs (Fig. 2.7). Visualization of the DNA is done with the help of a fluorescent dye such as ethidium bromide (EtBr) that intercalates into the DNA major grooves and fluoresces under UV light.

The constituents required for a restriction digestion are a DNA template, suitable restriction enzyme, a digestion buffer, and at times bovine serum albumin (BSA) to avoid sticking of enzymes to the tube surfaces and for stabilizing enzymes in overnight reactions [11]. At a certain temperature the reaction is incubated for the optimal activity of the restriction enzyme and after desired time the reaction is stopped by heat deactivation.

2.4.1 Restriction Enzymes (Endonucleases)

Many molecular biology methods are ingrained upon the skill to digest DNA molecules in a precise and predictable way (also known as “cutting” or “cleaving”). The advancement of this technology relies upon the discovery of bacterial restriction enzymes or endonucleases. Bacterial species contain restrictive enzymes that detect “nucleotide” patterns of DNA called palindromic (inverted repeat) sites of restriction [12]. Restriction sites are usually 4 to 8 base pairs (bp) long. The enzymes recognize and cleave at this site, generating a 5' phosphate and a 3' hydroxyl group at cleavage point. The restriction enzymes are usually named after the bacteria of which they are isolated. The initial letter of the genus is used, followed by the first two letters of the species. The type of strain or sub-strain sometimes follows the species designation in the name. Roman numerals are usually used to show if the specific enzyme was the

first, the second, the third, etc. For example, the first enzyme extracted from *Escherichia coli* strain RY13 is named EcoRI. So far, hundreds of restrictive enzymes accessible commercially have been discovered and isolated [12].

Restriction enzymes are of mainly three types—Type I, Type II, and Type III [12, 13], Type I and III being the complex ones having only limited role in recombinant DNA technology. Despite the fact that Type I and Type II both identify particular restriction sites, there is a significant variation between them. Type I restriction enzymes cleave ds-DNA at random locations away from their restriction recognition sites, resulting in indistinguishable restriction fragments. As a result, Type I restriction enzymes are of no use in molecular genetics. On the other hand, Type II restriction enzymes produce distinct and predictable restriction fragments by digesting the ds-DNA inside (or very near to) their restriction sites. Type II restriction enzymes can be further classified based on the type of cuts they make in the DNA leading to generation of either a *sticky* or a *blunt* end [13].

Some restriction enzymes digest DNA asymmetrically along their recognition sequence, leading to a single-stranded overhang on the digested end of the DNA segment. These overhangs, called “sticky ends,” consist of unpaired nucleotides that are produced at both the 5′ and 3′ ends. Cohesive ends are the ones produced by longer overhangs. The sticky overhangs are usually palindromic sequences, those that read the same from both 5′ to 3′ and 3′ to 5′ directions. The sticky ends make it possible for the vector and the insert to bind together. When the sticky ends are compatible, i.e., when the base pairs are complementary on the vector and the insert, the two parts of DNA are joined by a ligation process. Another advantage of sticky end generating enzymes is that less amount of enzyme is required when ligating the vector and the insert while cloning [14]. EcoRI, for example, identifies the sequence 5′GAATTC 3′ and makes a staggered cut, resulting in sticky ends with base pair overhangs. The formation of sticky overhangs is schematically explained in Fig. 2.8a below.

The second class of Type II restriction endonucleases (depending on the type of cut) includes the “blunt end” (also termed as the non-cohesive ends) generating restriction enzymes. These types of ends are generated when the enzyme gives a straight cut, thus terminating both strands into base pairs. This means there are no unpaired DNA strands or overhangs generated at the ends. Also, more amount of ligase enzyme as well as DNA is required for ligating the blunt-ended DNA molecules efficiently since there are no complementary ends produced [15]. For example, enzyme SmaI recognizes the sequence 5′GGGCCC 3′ and cuts both strands of the DNA between the same nucleotide pairs to produce blunt ends (Fig. 2.8b). However, for ligation purposes (described in Sect. 2.5), more amounts of ligase enzyme as well as DNA are required for ligating the blunt-ended DNA molecules efficiently [8]. One may also use additional tools such as adaptors (explained in Sect. 2.5.3) for efficient ligation of blunt end DNA.

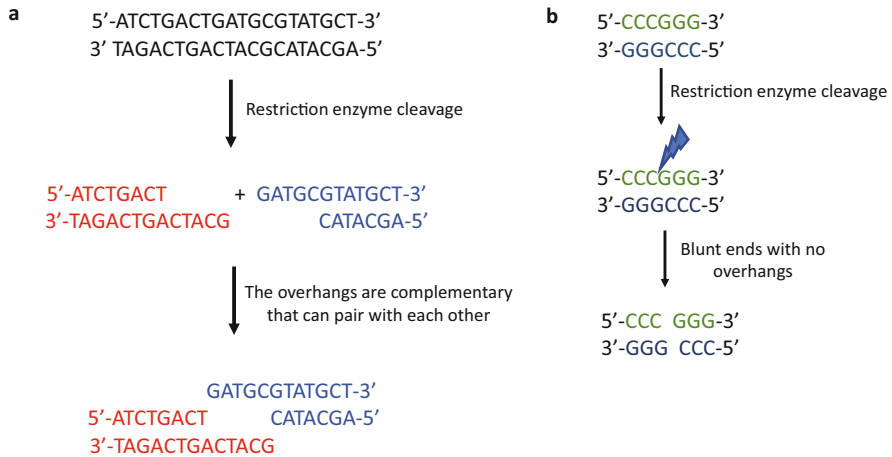


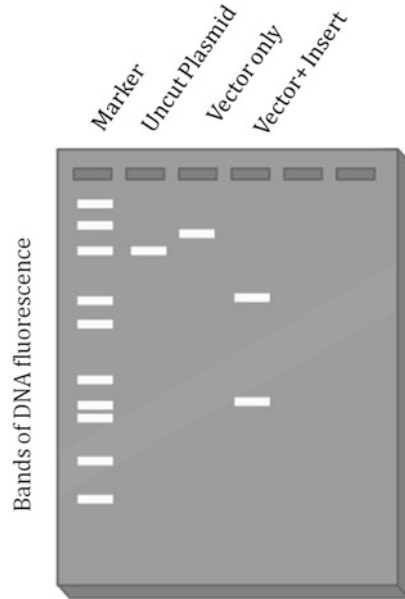
Fig. 2.8 Types of restriction enzyme cuts. **(a)** Generation of sticky or cohesive. Digestion of the DNA with a sticky end-generating restriction enzyme results in the formation of complementary staggered ends that have the capacity to pair up with each other. **(b)** Generation of blunt ends. Digestion of the DNA with a blunt end-generating restriction enzyme results in straight-cut cleavage and terminates both the strands in a base pair. There are no unpaired bases at the 5' and 3' prime ends

2.4.2 Steps and Tips for Restriction Digestion

General instructions:

1. The DNA for restriction digestion must be pure and devoid of impurities like EDTA, ethanol, and phenol, which are usually used to purify DNA.
2. Restriction enzymes should always be stored in a freezer. During the laboratory work, they can be kept in a benchtop cooler only for a limited amount of time.
3. To ensure optimal activity, restriction enzymes are used with appropriate buffers that are provided by the manufacturers. Some restriction enzymes, in addition to the buffer, require bovine serum albumin (BSA) for their optimal activity. BSA is usually supplied by the manufacturers at 100× concentration, which is then diluted to 10× in autoclaved distilled water before use.
4. The incubation temperature for most restriction enzymes is 37 °C. However, carefully read the reagent datasheet before incubating the reactions. Set the incubator or water bath at 37 °C or the recommended temperature for the restriction digestion reaction.
5. Incubation time varies depending on the amount of the enzyme used and the source of the DNA template. Usually 45 min to 3 h incubation is sufficient to digest any viral or bacterial DNA under the optimal conditions of incubation; however, eukaryotic DNA requires an overnight incubation.
6. Double digestion is a common procedure in restriction digestion, during which a piece of DNA is digested by two enzymes at the same time (Fig. 2.9). By using

Fig. 2.9 Analyzing digested product-size on a gel. The generalized expected results after restriction enzyme digestion of the recombinant DNA product are depicted above. An appropriate molecular weight standard is used as a reference to determine the correct size of the vector and insert. A single digestion should result into linearization of the vector (linear DNA travels slower than supercoiled plasmid in the uncut lane), while a successful double digestion should result in the release of a lower molecular weight insert



Visualization of the Agarose gel under UV trans-illumination

one enzyme, the vector gets linearized and a single band is observed. However, digestion with two restriction enzymes (in a sequential manner) releases the insert and two bands corresponding to vector backbone as well as insert are observed. In double digestion, it is essential to choose a buffer that ensures optimal activity for both the enzymes used. Furthermore, if BSA is required for either of the enzymes, it must be added to the double digestion reaction. The advantage of using BSA is that it will not inhibit the activity of the other enzyme that does not require it. The information regarding suitable buffer for setting the double digestion reaction can be obtained from the website of the manufacturer. In a case where no single buffer is found for a double digestion reaction, the digestion must be done sequentially. First, the reaction is digested with one enzyme + buffer combination, followed by a second digestion step with the second enzyme + buffer combination.

Protocol: Setting restriction enzyme digestion

1. Thaw all reagents on ice.
2. Prepare the reaction mixture of about 50 μL in a microfuge tube.
3. Add reagents in following order: molecular grade nuclease-free water, buffer, BSA (if mentioned), DNA template, and restriction enzyme.
4. Gently mix by tapping the tube. Briefly centrifuge to settle the contents of the tube.

5. If required, prepare positive control reaction with DNA template of known restriction site corresponding to the respective restriction enzyme of your choice.
6. Typical incubation time and temperature is 37 °C for 1 h, though the time and temperature may vary depending on the restriction enzyme used.
Note: Incubation time and temperature will vary depending on the enzyme as well as the concentration of the DNA template taken.
7. Restriction enzymes are then inactivated by incubation at high temperature (65–70 °C for 10–20 min).
8. Analyze the results of your restriction digestion using agarose gel electrophoresis (Fig. 2.9).
9. Typical restriction digestion reaction conditions are:

10× buffer	2 µl (1×)
DNA template	10 µl (2–4 µg)
Restriction enzyme 1 unit	1 µl
Autoclaved distilled water	7 µl
Final reaction volume	20 µl

Note: 10× denotes the concentration of the stock solution of any reagent; it is generally 10 times the concentration of the reagent that is supposed to be used in a particular reaction

2.5 Ligation

2.5.1 Introduction

Ligation of DNA is an important and final step in the construction of a recombinant plasmid. It involves joining of the DNA fragments (insert) to a compatible vector backbone that is digested with proper restriction enzymes. Both the insert and the vector need to have complementary overhanging base pairs or sticky ends (generated with the use of restriction enzymes during digestion) for the ligation reaction to take place (Fig. 2.10). Usually, digestion using two different restriction enzymes (one at the 5' end and the other at the 3' end) is preferred before ligating an insert into a vector. The pair of restriction enzymes used for the digestion should be the same for the vector as well as insert digestion so as to generate complementary overhangs. This allows the insert to be joined in the correct orientation to the vector and it also prevents the vector from self-ligating during the ligation process.

Apart from its application in cloning, *Non-Cloning Ligation* reactions have also found some popularity in other techniques. This form of ligation is basically adapted in Library preparation for Next Generation Sequencing (NGS) wherein a ligation step is typically incorporated to add bar-coded adapters to fragmented DNA [11]. It is also used in many novel detection or diagnostic methodologies, where the ligation of DNA probes followed by PCR amplification (Ligase Chain Reaction – LCR) has been used to detect single nucleotide polymorphisms (SNPs) [11].

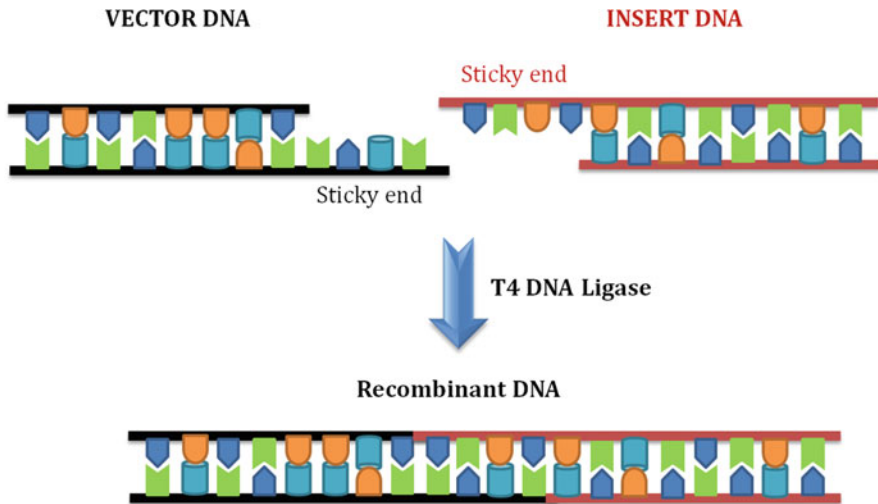


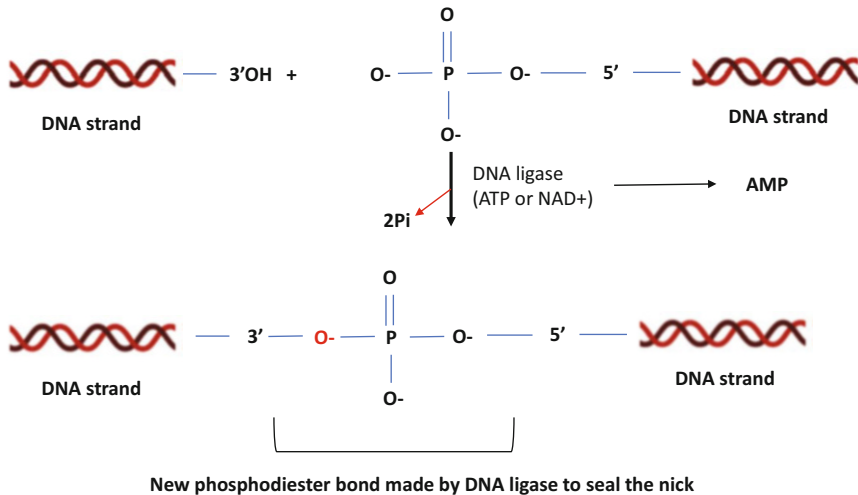
Fig. 2.10 Ligation reaction. The vector and insert are digested with the same pair of restriction enzymes prior to setting a ligation reaction. The complementary overhangs in both the DNA molecules, in the presence of the DNA ligase T4 enzyme, help in efficient ligation of the insert into the vector in correct gene orientation

In both the forms of ligations, cloning as well as non-cloning, the ligation reactions are primarily been catalyzed by enzymes called DNA ligases. However, this chapter will focus mainly on ligation pertaining to cloning genes of interest for producing recombinant proteins mainly in the bacterial host system for their further characterization in a laboratory setup.

2.5.2 DNA Ligases

For decades, DNA ligases have been studied for their role in joining the gaps that form DNA replication, recombination, and DNA repair. DNA ligases catalyze the formation of a phosphodiester bond between the 3' hydroxyl and 5' phosphate of the adjacent nucleotides resulting in the concomitant hydrolysis of ATP to AMP and inorganic phosphate [16]. A ligation reaction proceeds in three stages, where initially there is a transfer of an adenylyl group (AMP) from ATP to the ϵ -amine group of a lysine residue in the ligase enzyme. This results in the formation of an enzyme-nucleotide intermediate, with the release of pyrophosphate from ATP. In the second step, the adenylyl group is transferred from the enzyme to the 5'-phosphorylated end of the “donor” DNA strand, thus activating the enzyme. The third step involves a nucleophilic attack of the 3' hydroxyl group of the acceptor DNA to the adenylylated donor end of the other DNA strand resulting into the formation of the phosphodiester bond between the two strands with concomitant release of AMP (Fig. 2.11). However, DNA ligases can only form this covalent

Step 1 shows the transfer of AMP from ATP onto the ligase enzyme leading to the formation of enzyme nucleotide intermediate



Step 2 shows the formation of a phosphodiester bond between the two strands with the release of a pyrophosphate moiety

Fig. 2.11 Phosphodiester bond formation by DNA ligase. AMP is transferred from ATP to the ligase enzyme, resulting in the release of pyrophosphate from ATP. A nucleophilic attack of the 3' – OH group of the acceptor DNA strand to the adenylated donor strand results into the formation of the phosphodiester bond

linkage in a duplex molecule (i.e., when joining a nick in dsDNA or joining of an RNA to either a DNA or another RNA in a duplex form), but will not join single-stranded nucleic acids. For decades, molecular biologists have been exploiting DNA ligases for their efficiency in ligating DNA. T4 DNA Ligase, derived from bacteriophage T4, is the most commonly used DNA ligase and is found to be 400-fold more active than the bacterially derived *E. coli* DNA ligase. Hence, it is the enzyme of choice for most of the molecular cloning experiments [16].

2.5.3 Ligation Using Linkers and Adaptors

Although *E. coli* DNA ligase is an extremely popular enzyme for pasting a foreign gene into a vector, its application is somewhat limited by its inability to join a blunt-ended DNA cuts. To circumvent this problem, very large concentration of recombinant DNA molecules was used earlier. The presence of a highly concentrated DNA insert would increase the probability of its interaction with the ligase enzyme and hence ligation. This phenomenon, also known as “molecular crowding” [14, 17], however did not provide any promising solution to the problem that researchers encountered for ligating blunt-ended DNA molecules.

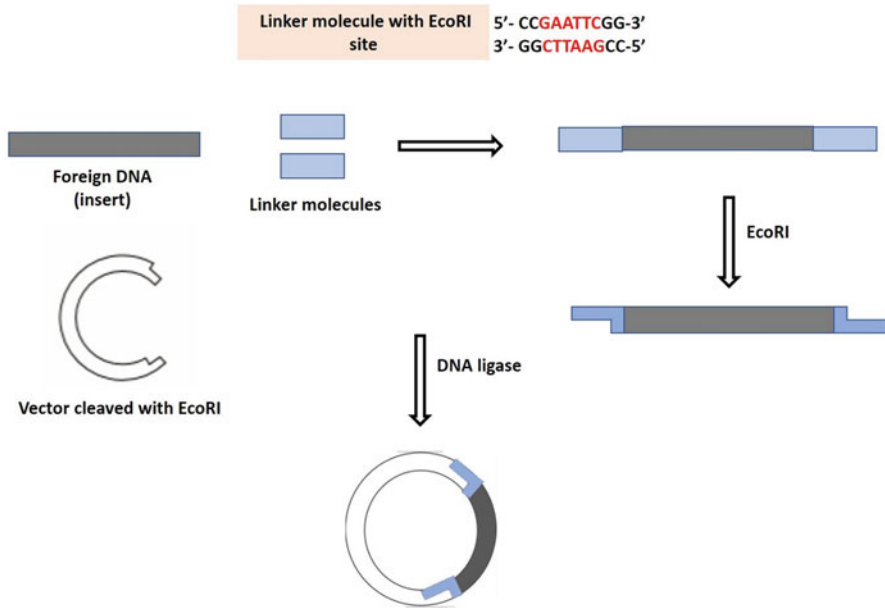


Fig. 2.12 Schematic representation showing attachment of linkers to DNA. A decameric linker molecule containing a site for restriction enzyme EcoRI is attached to a blunt-ended DNA insert. Upon individually digesting the vector and insert with EcoRI, it produces cohesive ends in both. The compatible overhanging ends of insert and vector facilitate the ligation required to produce the recombinant clone

Eventually, with the advancement in the recombinant DNA technology, a better approach to this problem was formulated where a linker sequence is attached to the blunt-ended DNA molecule (Fig. 2.12). This linker has a recognition site for a restriction enzyme that would produce sticky ends when cleaved. Once sticky ends are produced, ligation becomes easier [14, 15].

Another popular method of blunt-ended ligation is the use of adaptor molecules. Unlike linkers, adaptors are pre-formed cohesive-ended DNA fragments that are attached to the ends of blunt-ended DNA molecules, thus easing the ligation reaction [14]. Adaptor molecules with a free 5' hydroxyl (-OH) group (Fig. 2.13) are used initially while ligating them to the DNA. Since the free 5' phosphate end is a trigger for self-polymerization of DNA, it is replaced with -OH group. Once they are ligated to the DNA, 5' phosphate group is added to the adaptor ends in order to facilitate the next step of ligation reaction. The phosphate moiety is then added with the help of an enzyme polynucleotide kinase that uses phosphate group from ATP [18].

Homopolymer tailing is another approach that can be used for blunt end ligation. Polymeric tails of the same nucleotides are added to the population of DNA molecules. If there are two different populations of DNA molecules to work with, opposing homopolymer tails are added (for example, poly d(A) tailing on one set of molecules and poly (T) on the other), thus facilitating the annealing of DNA

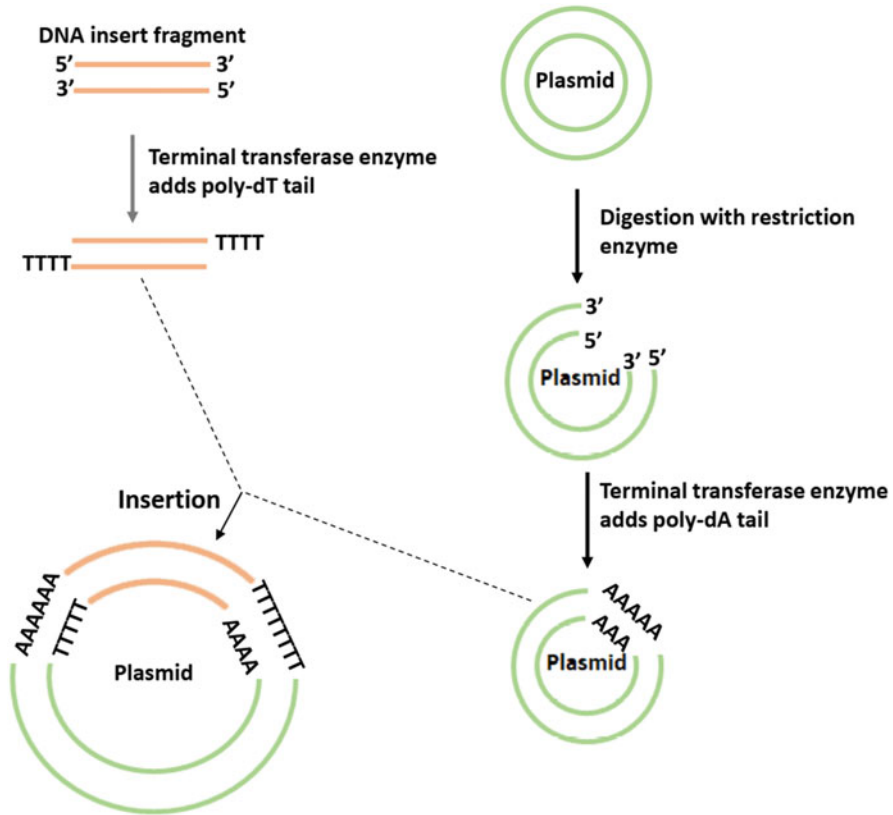


Fig. 2.14 Schematic representation of homopolymer tailing. When the gene insert is treated with exonuclease enzyme, it exposes the 3'-OH group of the insert. This region then is acted upon by deoxynucleotidyl transferase that adds specific nucleotides to generate homopolymer tails. Vector and insert consists of opposing homopolymer tails required for compatible blunt end ligation [14]

2.5.4 Standardizing the Ligation Reaction

The most important step in a ligation reaction is to optimize the amount of cut insert and vector to be used for the reaction. The vector to insert ratio used for a particular ligation reaction depends on the types of vectors used such as cDNA and genomic cloning vectors, as well as on the size and concentration of the vector and the insert used. For most standard cloning and ligation reactions (where the insert is smaller than the vector), a molar ratio of 1:3 of the vector to the digested insert DNA is usually recommended; however, one can also work with 1:1 and 1:2 molar ratio of vector to insert. In case of complicated cloning, where these ratios are not working, the amount of insert and vector can be optimized to improve the ligation efficiency [19]. For a standard ligation reaction of DNA fragments with blunt or sticky ends,

about 100 ng of digested vector is recommended, and the following formula is used to calculate the amount of insert to be used:

$$\frac{\text{Amount of vector (ng)} \times \text{Size of insert (kb)}}{\text{Size of vector (kb)}} \times \text{Molar ratio of } \frac{\text{insert}}{\text{vector}} \\ = \text{Amount of Insert (ng)}$$

For example,

The amount of insert DNA of 1 kb size required for the ligation with a 4 kb digested vector (50 ng) in 1:3 vector to insert molar ratio will be as shown in Eq. (2.2):

$$\frac{50\text{ng vector} \times 1\text{kb insert}}{4\text{kb vector}} \times \frac{3}{1} = 37.5\text{ng insert} \quad (2.2)$$

One can also use different ligation calculators such as NEBioCalculator [11] to calculate the molar ratios and estimate the amount of DNA to be used.

2.5.5 Steps Involved in Ligation

1. Assemble the following reaction (20 μ l) in a sterile microfuge tube kept on ice.

T4 DNA Ligase Buffer (10 \times)	2 μ l
Vector DNA (50 ng/ μ l)	2 μ l
Insert DNA	Appropriate amount depending on the concentration and molar ratio
T4 DNA ligase ^a (20 NEB units/ μ l)	1 μ l
Nuclease-free water	Add to a volume of 20 μ l

^aNOTE: T4 DNA Ligase is usually supplied in concentrated solutions (e.g., 400,000 units/ml, from New England Biolabs—NEB) by most manufacturers. Therefore, initially, it should be diluted in T4 DNA ligase dilution buffer and stored at a concentration of 20,000 NEB units/ml (60 NEB units corresponds to 1 Weiss unit) as aliquots at -20°C . As described by Bernard Weiss, Charles Richardson, and his colleagues, one Weiss unit is defined as the amount of enzyme required to catalyze the ATP-PPi exchange in the ligation reaction [20]. It is important to note that, while setting up the reaction, the aliquot should be kept in a benchtop cooler to prevent damage due to rapid freeze/thaw and should only be added at the end in the reaction mixture

2. Gently mix the contents by pipetting the solution and microfuge briefly for a few seconds.
3. Incubate the reaction according to the following conditions or according to the manufacturer's instructions:

16 $^{\circ}\text{C}$ overnight or room temperature for 10 min	Cohesive (sticky) ends
16 $^{\circ}\text{C}$ overnight or room temperature for 2 h	Blunt ends or single base overhangs

Table 2.3 Different controls used in ligation reaction

Control	Ligase	Requirement
Uncut vector	–	To check the viability of competent cells and to verify the antibiotic resistance of the plasmid as well as the quality of antibiotic used in the medium
Cut vector	–	To compare the background (number of colonies) obtained due to uncut vector
Cut vector	+	To compare the background obtained due to vector re-circularization—mostly useful in case of phosphatase treated vector (used for blunt end ligation strategies)
Insert or nuclease-free water	+	To recognize the contamination of the intact plasmid used in ligation or in the transformation reagents

4. Heat-inactivate the reaction at 65 °C for 10 min, if required.
5. Proceed with transforming 1–5 µl of the reaction mixture into competent cells of choice.

General tips:

1. It is always preferable to have appropriate controls for each of the ligation reactions as tabulated in Table 2.3 below.
2. T4 DNA ligase buffer contains ATP to drive the ligation reaction. To avoid degradation of ATP due to multiple freeze/thaw cycles, dispense the buffer into smaller aliquots of 5–10 µl and use one aliquot at a time. The whole step is to be performed on ice.
3. Polyethylene glycol (PEG) is usually known to promote ligation of blunt-ended fragments through macromolecular crowding [21]. Addition of about 2 µl of 50% (w/v) PEG 4000 in a 20 µl ligation reaction can be considered for blunt-ended ligations. However, while cloning cDNAs, one has to be careful with the concentrations used, as PEG can lead to formation of undesirable concatemers as well as residual PEG can be inhibitory to lambda packaging reactions (in vitro reactions used in the construction of cDNA libraries and genomic cloning of methylated DNA into λ-phage or cosmid vectors).

2.6 Ligation Independent Cloning (LIC)

2.6.1 Background

Conventional cloning steps such as restriction enzyme digestion and the subsequent ligation can become tedious at times. Ligation Independent Cloning (LIC) is a form of a cloning method that helps preclude the usage of ligase enzyme and thus evades the need for performing the tricky ligation step as involved in the abovementioned traditional cloning steps. The ligation-independent cloning method was developed in

the early 1990s, and since then it serves to be a quick, easy, and relatively cheap method for producing protein expression constructs [22].

The primary aim of this method is to generate long complementary overhangs at the ends of the template/insert DNA. These overhangs are required for establishing a stable and stronger association between the two fragments of interest without any external use of ligase enzyme. This method makes use of the T4 DNA polymerase enzyme for this purpose. The 3' exonuclease activity of this polymerase occurs in the presence of only a particular dNTP. Because of this property, it can create overhangs of varying length (typically 10–12 bp) in a sequence-specific manner. However, at the site of the first occurrence of the nucleotide (same as the added dNTP), equilibrium between 3' → 5' exonuclease and 5' → 3' polymerase activity is reached. The polymerase then stalls at this particular position and now the 5' → 3' polymerase activity of the T4 DNA polymerase takes over its 3' exonuclease activity (Fig. 2.15). Thus, long well-defined single-stranded DNA overhangs are produced at the ends of the plasmid as well as the gene of interest. Further, the annealing happens by simply incubating the complementary overhangs-containing vector and insert together. Due to the long length of these overhangs, the annealing reaction between the template DNA and PCR-generated insert becomes highly specific and the recombinant product is quite stable for subsequent transformation without any prior need for ligation. The assembled DNA construct, however, remains nicked at the junction site of the individual pieces. This issue gets resolved inside the transformed bacterial cells, wherein the bacterial ligases quite efficiently repair the nicked sites of the assembled product during replication cycle.

2.6.2 Protocol for LIC

The LIC cloning method involves the following major steps:

1. Preparation of Vector DNA.

- (a) For linearization of the empty vectors used for LIC, typically type II restriction enzymes (e.g., BsaI) are used. These enzymes cleave the vector at a specified distinct position with respect to its recognition sequence (...5'-GGTCTC(N1)/(N5)-3'..) [23] as shown in Fig. 2.15.

For linearization of the LIC vector by BsaI digestion, the following components are added:

Reagent	Amount
10× buffer (for restriction enzyme)	5 μl
LIC vector DNA	5 μg
BsaI (10 units/μl) (to be added at the end)	2.5 μl
Nuclease-free water	Add to a volume of 50 μl

Incubate the digestion mixture at 50 °C for 1 h. The linearized vector generated upon digestion will then be separated from the reaction mixture by agarose gel

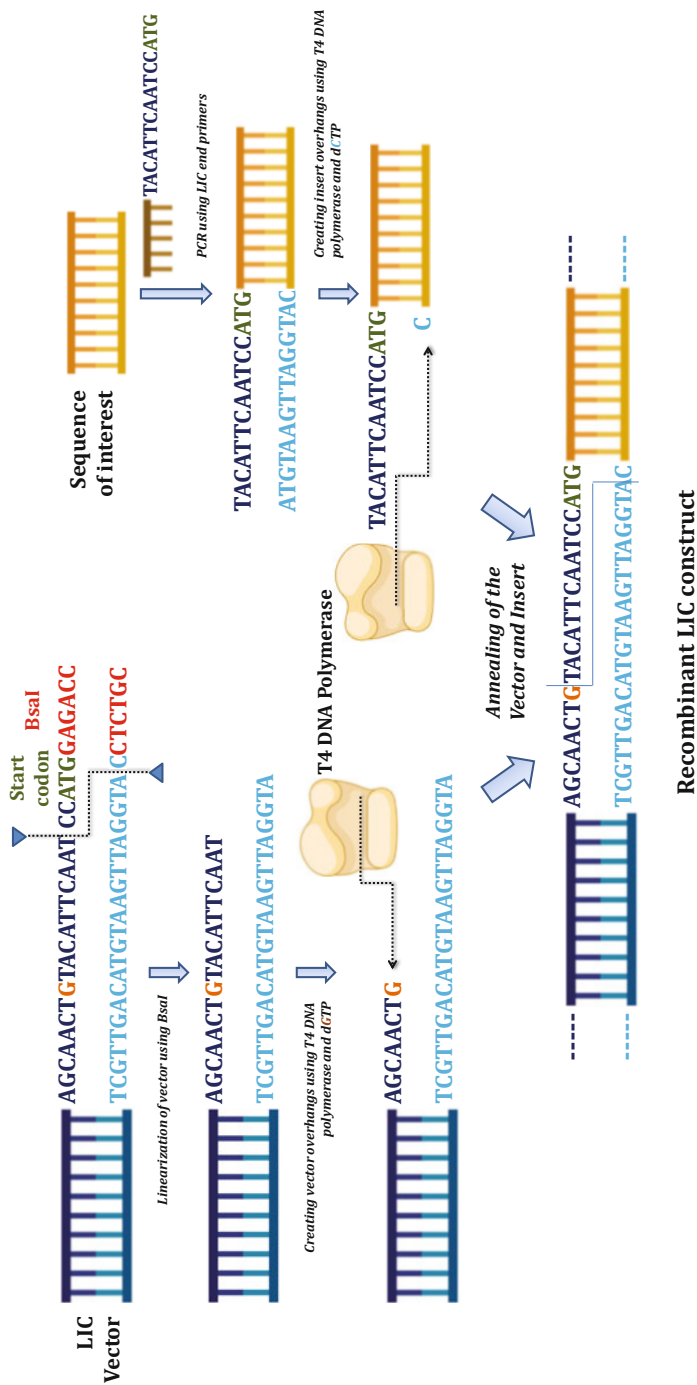


Fig. 2.15 Basic steps in LIC. The vector is linearized by using type II restriction endonucleases such as BsaI, which cleaves at a distinct site that is few nucleotides away from the recognition site. The vector and insert overhangs are generated using the T4 DNA polymerase that exhibits the chew-back (exonuclease) mechanism only till it encounters the first “G” (in case of vector) or “C” (in case of insert). The generated overhangs help in the annealing reaction to finally obtain the recombinant DNA product

electrophoresis. The digestion of the vector will remove any part of the MCS or any other portion of the vector if two BsaI sites are present. When visualized after agarose gel electrophoresis, there will be one band representing the linearized vector and the other will be the segment having two BsaI site at the ends. This is followed by extraction of the linearized vector by carefully excising the vector band only and performing gel purification using a DNA extraction kit.

NOTE: It is preferred to elute the purified DNA product in nuclease-free water instead of TE (Tris-EDTA) buffer, to avoid any interference of high salts in the subsequent reactions

The concentration of the vector DNA can be determined using the absorbance at 260 nm. When measured using a spectrophotometer having 1 cm pathlength, the optical density for a 50 µg/mL solution of any dsDNA at 260 nm (OD260) equals 1.0 [24]. Thus, we can calculate the vector DNA concentration using the following equation:

$$\text{dsDNA concentration} = 50 \mu\text{g/mL} \times \text{OD260} \times \text{dilution factor}$$

The purity of the nucleic acid is estimated by calculating the OD260/OD280 ratio. The OD260/OD280 ratio of pure DNA is around 1.8, while that for pure RNA the ratio is around 2.0. Lower ratios could be because of protein (in case of genomic/plasmid DNA extraction) or phenol contamination used during gel extraction.

(b) For creating overhangs at the end of the linearized vector, it is treated with T4 DNA polymerase, and a free nucleotide (e.g., dGTP) is added in the reaction (remember to exclude all the other nucleotides from the mixture of polymerase reaction). The enzyme chews back the sequence of the vector backbone until it encounters the first G nucleotide in the sequence. As the polymerase reaction is preferred over the exonuclease reaction in the presence of the dGTP, the polymerase will add back the guanosine residue and the exonuclease activity will stall. This is the state where the equilibrium between the two reactions is reached, as shown in Fig. 2.15.

For this type of T4 DNA polymerase reaction (40 µL mixture) in the LIC protocol, the following components are added [23].

Reagent	Final concentration	Volume (µL)
10× buffer (for polymerase)	1×	4
Gel extracted vector DNA	10–50 ng/µl	20–30
dGTP (100 mM)	2.5 mM	1
DTT (100 mM)	5 mM	2
BSA (10 µg/µl)	0.25 µg/µl	1
T4 DNA polymerase (to be added at the end)	0.075 units/µl	1
Nuclease-free water	Add to a volume of 40 µl	

Mix these components and incubate the reaction at 22 °C (or room temperature) for about 30 mins. After incubation, end the reaction by heating to 75 °C for 20 mins for inactivating the polymerase. Measure the final vector concentration through absorbance (the concentration obtained should be around 10–20 ng/μl) and store at –20 °C or lower until further use.

2. Preparation of the Insert DNA.

- (a) For amplification of the insert DNA, PCR is performed using suitable forward and reverse primers that are designed complementary to the 5' and 3' ends of the gene of interest, respectively. Before proceeding to the next step of overhang generation, it is essential to remove all the free nucleotides from the PCR amplified product, as they may interfere in the exonuclease activity of the T4 polymerase in the following step.
- (b) For generating overhangs in the insert DNA, T4 DNA polymerase is used. Unlike in case of the vector, here the polymerase reaction is performed in the presence of dCTPs. Thus, the T4 DNA polymerase exhibits the exonuclease activity only till it encounters the first C (cytosine) nucleotide in the sequence.

3. Annealing of the insert and vector.

The complementary overhangs that are created in the vector (step 1) and insert (step 2) are long enough for very strong and specific, enzyme-free annealing of the two DNA fragments.

2.6.3 Advantages

1. LIC serves to be a sequence-specific, ligase-free cloning method that is simpler and time-saving.
2. It is cost-effective and works efficiently over a broad range of DNA concentrations, even when the individual DNA fragments are not present in equimolar concentrations or in a particular ratio depending on their molecular sizes.
3. It is highly sequence specific and there is no issue of self-ligating plasmid or ligation in wrong orientation, as observed in the conventional ligation protocols.
4. It does not require the usage of T4 DNA ligase but depends on the strong interaction between the long complementary overhangs of insert and plasmid, as well as on the specific bacterial DNA ligases for joining of the remaining nicks.

2.7 Choice of Host Cells

After all the labor-intensive steps of cloning have been carried out, one needs to decide a suitable host organism that would replicate this newly designed plasmid clone. The gram-negative, rod-shaped *Escherichia coli* bacteria have been the commonly used lab organism for a variety of experiments since ages. Majority of

the common and commercially available lab strains of *E. coli* used today have descended from two individual isolates, the K-12 [25] and the B strains [26]. The K-12 strain led to the common lab strains MG1655 and its derivatives DH5 α and DH10b (alternatively known as TOP10), while the host cells used for protein expression such as BL21 strain [27] and its derivatives are obtained from the *E. coli* B strain [28]. For cloning, a number of different commercial strains of *E. coli* are currently available that can be chosen based on their characteristics for selection of suitable clones. The commercial strains are marketed with specialized properties such as fast growth, routine cloning, high-throughput cloning, maximum DNA yield, cloning of unstable DNA, preparing unmethylated DNA, and much more. For making these features possible, many mutations/genetic changes are made to improve the plasmid yield and DNA quality, confer resistance to any antibiotic, and improve uptake of foreign DNA. Thus, each strain is described by its “genotype” that suggests the particular insertions and deletions that the strain carries and this helps in determining its usage for the desired cloning application. Table 2.4 provides the details of some of the popular strains derived from *E. coli* K-12 and their primary uses in the lab.

More details about the other host cells used for cloning and protein expression has been described in Chap. 4.

2.8 Transformation

Once the process of cloning is successful, one needs to propagate the recombinant DNA molecule into suitable host systems such as bacteria, so as to obtain ample amount of the cloned DNA required for further studies. Transformation is the method by which exogenous DNA is transferred into the host cell. The idea of inducing the bacterial cells to take up the external DNA molecule and replicate as its own has revolutionized various aspects of molecular genetics [36]. Transformation refers to the uptake of DNA into bacterial, yeast, or plant cells, whereas transfection is typically used in mammalian cells. Prior to performing any transformation method, the host cells are made competent (able to take up exogenous DNA) with the help of different methods [37]. The concept of competence and the different methods used to prepare competent cells are described in detail in Chap. 4. Classically, the procedure for introducing a DNA construct into a host cell by transformation is either by chemical method or electroporation or by particle bombardment. Chemical transformation involves treating cells with divalent cations such as calcium chloride (CaCl₂) or rubidium chloride (RbCl), which makes the bacterial cell wall more permeable to DNA. Heat shock is used to create temporary pores in the cell membrane, allowing exogenous DNA to be transferred into the cell. A mild electrical pulse is used in electroporation to make the bacterial cell temporarily permeable. Particle bombardment is generally employed for the transformation of plant cells where the DNA construct coated gold or tungsten particles are forced into the cell physically by using gene gun. Herein, we discuss the method of chemical transformation used for bacterial cells.

Table 2.4 Popular strains of *E. coli* used in cloning

Strain	Description	Resistance	Usage
MG1655	Wild type K-12 strain	–	First published sequence of an <i>E. coli</i> K-12 strain.
DH5alpha	Derived from <i>E. coli</i> K12	–	Most commonly used for general routine cloning and storage of common plasmids, used for blue/white screening [29]
HB101	Hybrid of <i>E. coli</i> K12 and <i>E. coli</i> B strains, with maximum part (98%) of K12	Streptomycin	Common lab strain for cloning and storage of pBR322 and many other plasmids [30]
DB3.1	HB101 derivative, strain with a specific mutation in the DNA gyrase	Streptomycin	Used for propagating plasmids expressing the <i>ccdB</i> gene (lethal gene that targets DNA gyrase) [31, 32]
MC1061	Parent of DH10B/TOP10 and other derived strains	Streptomycin	Common lab cloning and storage strain, leucine auxotroph, galactose, lactose, arabinose nonutilizing strain [33]
TOP10	MC1061 derivative	Streptomycin	General cloning and storage, blue/white screening, leucine Auxotroph [33]
DH10B	MC1061 derivative	Streptomycin	General cloning and storage, blue/white screening, leucine Auxotroph [33]
JM109	Partly restriction-deficient		Good strain for cloning repetitive DNA, can be used for plasmid maintenance and blue/white screening [34]
JM110	Allows for methylation sensitive restriction enzymes to cut the plasmid after preparation.	Streptomycin	For storing plasmids that should not be <i>Dam</i> (DNA adenine methylase mutation (GATC)) or <i>Dcm</i> (DNA cytosine methylase mutation (CCWGG)) methylated [35]
NEB Stable	Strain from NEB		Used for cloning into and storage of lentiviral and retroviral vectors; cloning of repeated sequences with the potential to recombine [35]
XL1 blue	Nalidixic acid resistant (antibiotic that inhibit a subunit of DNA gyrase and topoisomerase IV)	Tetracycline	Blue/white screening and routine cloning.
XL10 gold		Tetracycline and Chloramphenicol	Hte phenotype provides high competency required for cloning and propagation of large plasmids, ligated DNA, and libraries [34]

2.8.1 Protocol for Transformation

1. Remove cells from $-80\text{ }^{\circ}\text{C}$ freezer and thaw on ice.
Add 1–5 μl containing 1 pg –100 ng of plasmid DNA to the cell mixture. The concentration of DNA used depends on the competency of the cells used; more competent the cells are, lesser is the DNA used. Carefully flick the tube 4–5 times to mix the cells and DNA. Do not vortex.
2. Place the mixture on ice for 30 min. Do not mix.
3. Heat shock at exactly $42\text{ }^{\circ}\text{C}$ for precisely 60–90 s. Do not mix.
4. Post heat shock, immediately place the tube on ice for 3 min. Do not mix.
5. Add 700 μl of room temperature growth media into the transformation mixture.
Allow the cells to grow by incubating the tube at $37\text{ }^{\circ}\text{C}$ for 40–60 min.
6. Harvest the cells by centrifuging at about 5000 rpm for 5 min.
7. Resuspend the cell pellet in about 50–100 μl of the same media. Immediately spread plate onto a selection medium plate and incubate overnight (14–16 h) at $37\text{ }^{\circ}\text{C}$.

NOTE: The choice of antibiotics, other than ampicillin, may require some outgrowth before plating on selective media. Colonies develop more quickly at temperatures over $37\text{ }^{\circ}\text{C}$, although certain constructs at high temperatures may be unstable.

Details about the other methods of transformation and the troubleshooting involved have been provided in Chap. 4.

2.9 Colony Screening

After the steps of ligation and transformation, one needs to identify the colonies that have been successfully transformed with the recombinant DNA product. Antibiotic selection is one of the crude ways of identifying the plasmids that may carry the gene of interest. However, at times, self-ligated plasmids may also produce false positive results for the cloning, as the antibiotic resistance gene is present in the plasmid and not the gene of interest (insert). Hence, one needs to utilize more specific methods for screening the bacterial colonies transformed with the end product of the cloning reaction as described below:

2.9.1 Blue-White Colony Screening

2.9.1.1 Background

Blue-white colony screening method has been a classic way to detect the colonies that contain plasmid with an insert. It is an effective molecular biology tool that is widely used as a primary step in screening the final recombinant bacteria obtained from the cloning experiments. It is a negative selection system wherein bacterial lactose metabolism is used to indicate successful cloning.

This technique relies on the enzymatic activity of β -galactosidase, a tetrameric enzyme encoded by the *lacZ* α gene present in the well-characterized bacterial *lacZ* operon of *E. coli*. When lactose or its functional analog IPTG (isopropyl β -D-1-thiogalactopyranoside) is present in the cellular environment, it triggers the *lacZ* operon either by inducing the operon (lactose) or by inactivating the *lac* operon repressor (IPTG). Activation of the *lac* operon results in the generation of a fully functional β -galactosidase enzyme that metabolizes lactose into glucose and galactose. However, if the *lacZ* gene is been mutated or a part of it is deleted, β -galactosidase is not produced and the substrate remains intact [38].

Most plasmid vectors used for cloning purposes contain a short segment of *lacZ* α gene that codes for the first 146 amino acids of β -galactosidase, while the *E. coli* host strains used contains deletion mutation of the same segment, called *lacZ* Δ M15, which results into a nonfunctional β -galactosidase enzyme. Hence, when plasmid vector containing the *lacZ* α gene segment is taken up by such *E. coli* host strains, α -complementation occurs. Herein, the lack of *lacZ* α gene segment in the mutant bacterial cells is complemented by the α -peptide section present in the plasmid, resulting in the production of a functional enzyme. The plasmid vectors specifically used for cloning purposes also contain a multiple cloning site (MCS) present within the *lacZ* sequence (Fig. 2.16). Therefore, when an insert DNA is ligated into the plasmid vector, it disrupts the *lacZ* α gene segment, alpha complementation cannot occur upon transformation, and a functional β -galactosidase does not form. If the gene of interest is not inserted into the vector or is inserted at a different location other than MCS, the *lacZ* gene in the plasmid vector remains intact and this promotes the α -complementation process producing a functional enzyme.

For visualization of the β -galactosidase enzyme activity, a chromogenic dye-linked substrate called x-gal (5-bromo-4-chloro-3-indolyl- β -D-galacto-pyrano-side) is added to the agar plate, along with the inducer IPTG. Production of β -galactosidase results into break down of x-gal into galactose and an insoluble blue pigment (5,5'-dibromo-4,4'-dichloro-indigo) [39]. Thus, as mentioned earlier, if the plasmid vector contains the insert, β -galactosidase is not produced and the resultant colonies are of whitish-cream color of standard *E. coli*. If cloning reaction is unsuccessful, the α -peptide remains intact and functional β -galactosidase enzyme is produced. X-gal in the medium is hydrolyzed by these non-recombinant cells to form 5-bromo-4-chloro-indoxyl, which spontaneously dimerizes to produce the blue pigment. Thus, the colonies formed by non-recombinant cells appear blue, which can be well distinguished from the recombinant ones that appear white.

2.9.1.2 Protocol for Blue-White Colony Screening

After performing the steps of ligation and transformation, the cells are plated onto media containing suitable amount of chromogenic substrate and IPTG. Different chromogenic substrates such as X-GlcA, X-Gal, and S-Gal are commercially available and the methods for addition of such products into the media differs; some are spread directly on LB (Luria-Bertani) agar plates along with IPTG, while others are incorporated into the medium before autoclaving and then agar plates are prepared [39, 40].

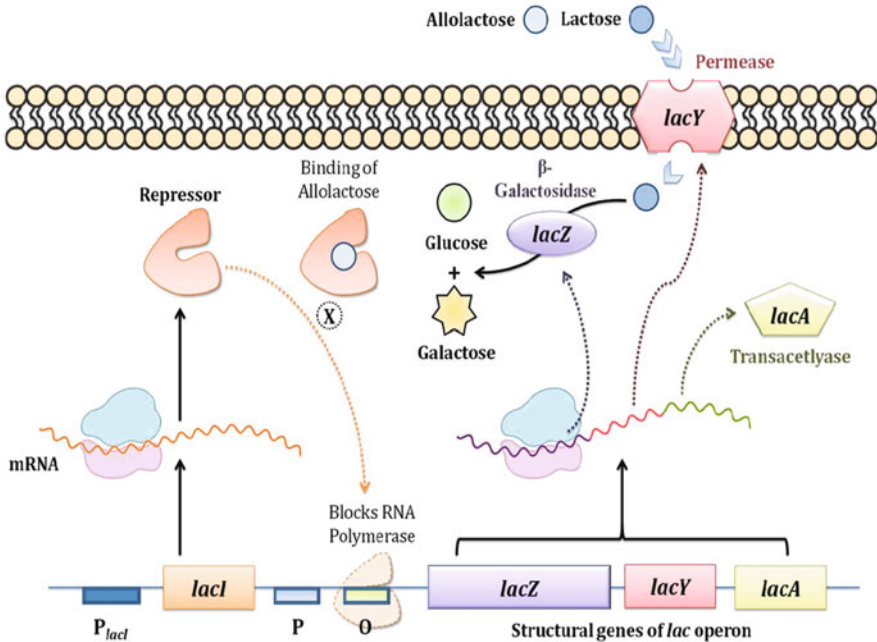


Fig. 2.16 The *lac* operon in *E. coli*. The three structural genes *lacZ*, *lacY*, and *lacA* are under the control of a single promoter (P) and are together transcribed as a continuous mRNA. The repressor protein is produced constitutively by the *lacI* gene (through the upstream promoter P_{lacI}) and its function is to bind to the operator region (O). In the presence of inducers such as lactose (or its functional analog, Allolactose), repressor binding is prevented and the structural genes are transcribed. The gene product of *lacZ* is β -galactosidase—enzyme that cleaves lactose into glucose and galactose. *lacY* encodes permease—a membrane channel protein that allows the transport of lactose into the cell at an increased rate, and *lacA* encodes transacetylase that acetylates galactosides other than lactose and prevents their cleavage by β -galactosidase

NOTE: The chromogenic substrates are light and temperature sensitive. They need to be prepared in the form of stock solutions and added to media only after autoclaving or according to the manufactures' protocol. If spread on top of LB plates, it should be evenly distributed and sufficient drying time should be provided before use.

In both cases, appropriate concentration of selected antibiotic is added to the medium and the following steps are followed:

1. Spread approximately 10–100 μ l of the transformed *E. coli* cells onto the LB agar plates (containing IPTG and chromogenic substrate) using a flame-sterilized glass spreader.

NOTE: Besides the recombinant product, transform empty plasmid vector (without insert) and spread plate. This plate serves as a good control indicating the quality of IPTG and chromogenic substrate.

2. Incubate the plates overnight at 37 °C for 24–48 h depending on the type of cells used.

3. After incubation, blue and white colonies appear on the agar surface. Select the recombinant cells in the white colonies to culture for DNA isolation and sequencing.

NOTE: Presence of only white colonies is also not a reliable result. It is important to provide enough incubation for any intact β -galactosidase to be expressed and process the substrate into blue pigment (16–20 h).

2.9.1.3 Limitations

1. This technique is only a visual screening method and not a selection technique. Hence, it should be used in combination with other selection methods.
2. Due to the incorporation of some mutations in the *lacZ* gene of the vector while cell maintenance, the gene may sometimes be nonfunctional. Thus, the resulting colony may appear white, but will not be recombinant.
3. Blue-white colony screening only indicates the presence of an insert, which may not necessarily be the insert of interest. Disruption of the α -peptide DNA by any cloning artifact will also lead to false positive white colonies.
4. False negative cases are rare. However, if a small fragment is inserted in-frame, read-through can occur and lead to a functional β -galactosidase enzyme, giving rise to a blue colony.

2.9.2 Other Screening Methods

2.9.2.1 Positive Selection System

This method follows similar principle as that of the blue-white screening. Here, the positive selection vectors encode a lethal gene, such as any restriction enzyme that digests the host genomic DNA. Inserting the cloning fragment inserted into the center of this gene (present in the MCS) disrupts the lethality, and thus only the recombinant clones survive. Antibiotic selection can also be used in combination with this method to ensure that positive colonies do contain the plasmid with the lethal gene.

2.9.2.2 Screening by Plasmid Miniprep and RE (Restriction Enzyme) Digests

This method involves isolating the recombinant plasmid DNA from the clones and checking the presence as well as orientation of the insert by restriction enzyme digestion. The colonies obtained after transformation of the ligated product are inoculated in LB media supplemented with suitable antibiotics, and grown overnight. Plasmid DNA isolation is carried out from these miniprep cultures using the protocol, as described in Fig. 2.17. For high-copy plasmids, one can obtain 4–10 μ g plasmid DNA per purification (1–5 ml). For low-copy plasmids, one needs to grow more amount of culture (10 mL) and can obtain 1–3 μ g plasmid DNA per purification [19].

After isolating the plasmid DNA from the expected recombinant clones, the purified plasmid DNA is digested using restriction enzymes. Before using this

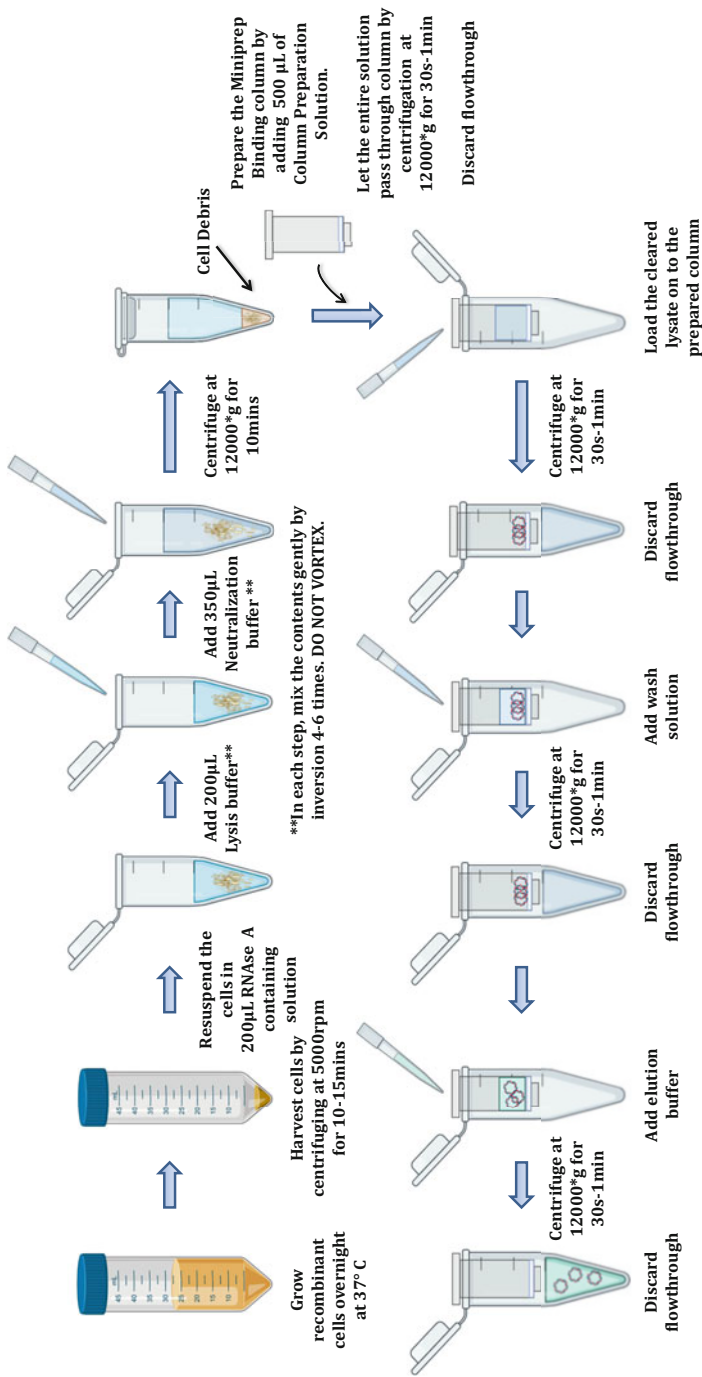


Fig. 2.17 Basic steps in Plasmid isolation. Miniprep cultures are grown overnight and the recombinant clones are harvested. Plasmid DNA extraction is then performed using a modified alkaline-SDS lysis method, followed by adsorption of the plasmid DNA on to silica matrix columns in the presence of high salts. Contaminants are then removed by a spin-wash step. The bound DNA is finally eluted in nuclease-free water or TE (Tris-EDTA) buffer

method for screening, one needs to perform restriction site mapping to identify restriction enzymes that can be particularly used to release the insert from the recombinant plasmid. Once the DNA is purified, about 0.5–1 μg of plasmid can be used for restriction enzyme digestion. The digested product can then be run on agarose gel to verify whether the vector backbone and insert are of the expected sizes.

2.9.2.3 Colony PCR

Colony screening with PCR is the most rapid and cost-effective screening test that helps determine the presence of insert DNA. It involves lysing the bacteria and amplifying a portion of the plasmid with specific primers. The most important step of this method is designing primers and determining their combinations to be used for PCR. There are three approaches for primer design depending on the requirement: (1) insert-specific primers, (2) backbone-specific primers, and (3) orientation-specific primers (Fig. 2.18) [41].



Fig. 2.18 Types of primers used in colony PCR. Designing specific primers is the most important step in colony PCR. Orange-colored box indicates the insert and the arrows indicate the primers. Depending on the need of the experiment, the specificity of the primers can be designed and the results may vary. Different combination of these primers can be used to detect the presence and orientation of the insert in a recombinant clone

In most cases, PCR can be performed either by using vector-specific primer or insert-specific primer or both. If one is doubtful about maintaining the orientation of the insert during cloning steps (especially for blunt end cloning), they can use orientation-specific primers for colony screening.

Once the primers are obtained, a standard PCR reaction (primers, dNTPs, and polymerase) is set up using a portion of the overnight grown culture of the transformed cells. Briefly the protocol comprises the following steps:

1. Lyse the cells to release the plasmid DNA by briefly boiling the sample and using the supernatant, or by directly adding the sample to the PCR master mix. The initial heating step of the PCR reaction helps in the lysis.
2. For amplification of the desired plasmid region, a standard Taq polymerase is sufficient.
3. Run the obtained PCR product with the specific controls on a 1% agarose gel to analyze the exact product size and the success of cloning.

This method allows screening of several colonies at a time and eliminates the need to first purify the plasmid DNA required for using as a template for PCR.

2.9.2.4 Sanger Sequencing

The final step in most of the cloning strategies is to verify whether the sequence of the insert, insert orientation, and the sequences of the junctions between the plasmid and insert DNA are correct. This can be achieved by sequencing the plasmid DNA using Sanger sequencing (also called as chain termination or cycle sequencing) [42, 43]. In addition to the reagents used in a standard PCR, four fluorescently labelled dideoxynucleotide triphosphates (ddNTPs: ddATP, ddGTP, ddCTP, and ddTTP) are also added in low ratio. Random addition of these distinctively labelled ddNTPs terminates the synthesis reaction. Due to the absence of hydroxyl group in ddNTPs, polymerase fails to form the phosphodiester bond with the next nucleotide and the reaction terminates. The basic steps involved in Sanger's sequencing are briefly described in Fig. 2.19. This method acts as a confirmatory step that allows one to identify any mutations that have been inadvertently incorporated in the cloning product. The sequencing is performed manually, or more commonly, in an automated fashion using a sequencing machine.

The sequencing data obtained is in the form of a color-coded electropherogram or chromatogram, which shows the fluorescent peak of each nucleotide along the length of the template DNA [44]. This can then be converted to nucleotide sequence by the computer. One can then use alignment tools like ClustalW to check the correct sequence of the insert or for the presence of any mutation in the insert.

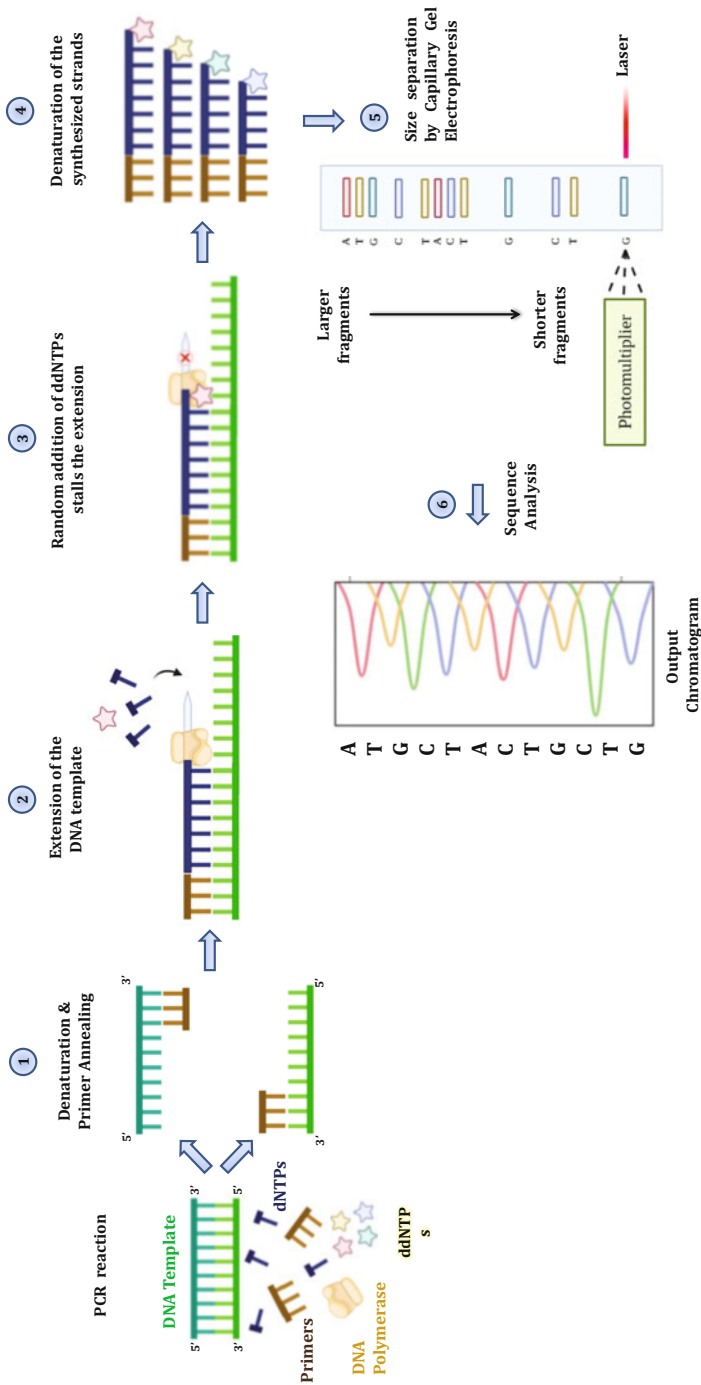


Fig. 2.19 Basic steps of automated Sanger sequencing. The dsDNA is denatured into two ssDNA and the respective primers bind at their 3' ends. Extension of the new strand occurs until a termination nucleotide (ddNTP) is randomly incorporated. The resulting DNA fragments are again denatured into ssDNA and these are further separated by gel electrophoresis for determination of the sequence

2.10 Troubleshooting for Subcloning Experiments (Table 2.5)

Table 2.5 Alternative strategies to be used for solving cloning errors

Problem	Reason	Solution
Host cells do not contain plasmid after transformation	<ol style="list-style-type: none"> 1. Less antibiotic used 2. Satellite colonies that are very small in size are chosen 	<ol style="list-style-type: none"> 1. Add adequate amount of antibiotic and use freshly made media and antibiotics 2. Selection of well-established bigger colonies is important
Host cells contain the wrong construct	<ol style="list-style-type: none"> 1. PCR amplification was incorrect 2. Recognition site of restriction enzyme used for digestion site present in the gene 3. Mutations are present in the cloned DNA 	<ol style="list-style-type: none"> 1. Standardize the protocol for PCR; purify correct PCR product 2. Check for the presence of restriction digestion site in the coding region of the gene by using certain software like Snapgene 3. Repeat sequencing of the clone; use a high-fidelity DNA polymerase in PCR reaction
No ligation product obtained	<ol style="list-style-type: none"> 1. Insufficient ligation reaction 	<ol style="list-style-type: none"> 1. Temperature used for incubation of the reaction was not optimum 2. Ligase enzyme has been inactivated due to improper storage 3. Ratios of insert and vector need to be varied 4. Contaminants such as EDTA in DNA purification buffers need to be removed

2.11 Conclusions

In this chapter, we have seen how we can clone a gene of interest into a suitable vector and produce large copies of recombinant clones. One can generate gene libraries using either the genomic DNA or cDNA, utilize different restriction enzymes to cut the gene of interest and ligate it in a compatible vector using the ligase enzyme. These recombinant clones can be further used for the sequence analysis study, understanding the functional relevance of the gene by performing protein expression as well as for developing probes that are used for studying its expression within cells. The next two chapters (Chaps. 3 and 4) describe the nature of the vectors that can be used for different cloning purposes and the techniques employed for introducing the recombinant DNA molecules into suitable host cells in detail.

Acknowledgments The authors thank Advanced Centre for Treatment, Research and Education in Cancer (ACTREC) for providing necessary infrastructure and resources for successful completion of the chapter. The authors acknowledge Ms. Chanda Baisane, Bose Lab, ACTREC for formatting the manuscript.

Problems

Multiple Choice Questions

1. In lac operon, repressor molecules bind to _____
 - (a) Enhancer
 - (b) Promoter
 - (c) Hormone Receptor
 - (d) Operator
2. In Sanger sequencing, separation of the different sized templates is done through ____
 - (a) Agarose gel electrophoresis
 - (b) Capillary gel electrophoresis
 - (c) SDS- PAGE
 - (d) Electrophoretic mobility shift assay
3. Which of the following statements is false?
 - (a) The competent cells with no recombinant plasmids are drug resistant
 - (b) Plasmid replication is independent of cell division
 - (c) The competent cells that take up recombinant plasmid molecules become drug resistant
 - (d) The competent cells are originally drug resistant

Subjective Questions

1. A plasmid vector and gene under study were digested with the same restriction enzyme. The vector and the insert were ligated using DNA ligase. The ligation product was used to transform the host E. coli. The colonies were obtained on a suitable selection medium. What possible products are expected after isolating DNA from the colonies. What are the ways to determine these different products obtained?
2. Digestion of an insert with a restriction enzyme produces blunt ends. In order to ease the process and ensure that the insert ligates with the vector, what modifications can be made to the blunt-ended insert?

References

1. Dale JW, Schantz MV, Plant N. From genes to genomes: concepts and applications of DNA technology. 3rd ed. John Wiley & Sons: Chichester; 2011.
2. Russo E. Special report: the birth of biotechnology. *Nature*. 2003;421(6921):456–7.
3. Brown TA. Gene cloning and analysis an introduction. 6th ed. Chichester: John Wiley & Sons, Inc.; 2010.

4. Dale JW, von Schantz M, Plant N. From genes to genomes: concepts and applications of DNA technology. 3rd ed. Chichester: John Wiley & Sons, Inc.; 2011. p. 400.
5. Bajpai B. High capacity vectors. India: Springer; 2013.
6. Jeremy W, Dale MVS, Plant N. From genes to genomes: concepts and applications of DNA technology. Chichester: John Wiley & Sons, Inc.; 2011.
7. Sambrook J, Russell DW. Molecular cloning: a laboratory manual, vol. 2. 3rd ed. New York: Cold Spring Harbor Laboratory Press; 2001.
8. Coffin JM, Hughes SH, Varmus HE. Retroviruses. New York: Cold Spring Harbor Laboratory Press; 1997.
9. Bhagavan NV, Ha C-E. Essentials of medical biochemistry. 2nd ed. Netherland: Elsevier; 2015.
10. Eberwine J, et al. Complementary DNA synthesis in situ: methods and applications. *Methods Enzymol.* 1992;219:80–100.
11. New England Biolabs. <https://international.neb.com/applications/cloning-and-synthetic-biology/dna-ligation>.
12. A comprehensive list of all the known restriction endonucleases and their recognition sequences. REBASE: <http://rebase.neb.com/rebase/>.
13. Pray LA. Restriction enzymes. *Nat Educ.* 2008;1(1):38.
14. Primrose SB, Twyman R. Principles of gene manipulation and genomics. 7th ed. Oxford: Blackwell Scientific Publications; 2006. p. 672.
15. Lodish H, Berk A, Zipursky SL. Molecular cell biology. 4th ed. New York: W. H. Freeman; 2000.
16. Tabor S. DNA ligases. *Curr Protoc Mol Biol.* 2001;CH 3:Unit3.14.
17. Aumiller WM, et al. Chapter Five - Phase separation as a possible means of nuclear compartmentalization. *Int Rev Cell Mol Biol.* 2014;307:109–49.
18. Kleppe K, Lillehaug JR. Polynucleotide kinase. *Adv Enzymol Relat Areas Mol Biol.* 1979;48:245–75.
19. Promega. Subcloning notebook.
20. Weiss B, Thompson A, Richardson CC. Enzymatic breakage and joining of deoxyribonucleic acid. *J Biol Chem.* 1968;243(17):4543–55.
21. Zimmerman SB, Pfeiffer BH. Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or *Escherichia coli*. *Proc Natl Acad Sci U S A.* 1983;80(19):5852–6.
22. Haun RS, Serventi M, Moss J. Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *BioTechniques.* 1992;13(4):515–8.
23. <https://www.addgene.org/protocols/lic/>.
24. Carlos F, Barbas III, et al. Quantitation of DNA and RNA. *CSH Protoc.* 2007;2007:pdb.ip47.
25. Bachmann BJ. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev.* 1972;36(4):525–57.
26. Abedon ST. Selection for lysis inhibition in bacteriophage. *J Theor Biol.* 1990;146(4):501–11.
27. Daegelen P, et al. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *J Mol Biol.* 2009;394(4):634–43.
28. Samuelson JC. Recent developments in difficult protein expression: a guide to *E. coli* strains, promoters, and relevant host mutations. *Methods Mol Biol.* 2011;705:195–209.
29. Grant SG, et al. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci U S A.* 1990;87(12):4645–9.
30. Carzaniga T, et al. Autogenous regulation of *Escherichia coli* polynucleotide phosphorylase expression revisited. *J Bacteriol.* 2009;191(6):1738–48.
31. Bernard P. New ccdB positive-selection cloning vectors with kanamycin or chloramphenicol selectable markers. *Gene.* 1995;162(1):159–60.
32. Maki S, et al. Partner switching mechanisms in inactivation and rejuvenation of *Escherichia coli* DNA gyrase by F plasmid proteins LetD (CcdB) and LetA (CcdA). *J Mol Biol.* 1996;256(3):473–82.
33. Durfee T, et al. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol.* 2008;190(7):2597–606.

34. Hoseini SS, Sauer MG. Molecular cloning using polymerase chain reaction, an educational guide for cellular engineering. *J Biol Eng*. 2015;9:2. <https://doi.org/10.1186/1754-1611-9-2>.
35. Ferenc BM. Plasmids 101: common Lab E. coli strains. Watertown, MA: Addgene Blog; 2014.
36. Cohen SN, Chan ACY, Hsu L. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci U S A*. 1972;69(8):2110–4.
37. Hanahan D, Jessee J, Bloom FR. Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol*. 1991;204:63–113.
38. Shuman HA, Silhavy TJ. The art and design of genetic screens: *Escherichia coli*. *Nat Rev Genet*. 2003;4(6):419–31.
39. Green MR, Sambrook J. *Molecular cloning a laboratory manual*. New York: Cold Spring Harbor Laboratory Press; 2012.
40. Padmanabhan S, Banerjee S, Mandi N. *Screening of bacterial recombinants: strategies and preventing false positives*. London: IntechOpen; 2011.
41. Kenkel BB. Plasmids 101: Colony PCR. <https://blog.addgene.org/plasmids-101-colony-pcr>.
42. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975;94(3):441–8.
43. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463–7.
44. Dewey FE, et al. DNA sequencing: Clinical applications of new DNA sequencing technologies. *Circulation*. 2013;125(7):931–44.