

Chapter 2

Recombinant Enzyme: Cloning and Expression

Azura Amid and Norhidayah Hassan

Abstract This chapter will discuss the general strategies of cloning recombinant enzymes from identifying a suitable DNA template to choosing the correct host expression system. However, there is no standard operating procedure for cloning a recombinant enzyme. The best strategies should be based on the researcher's creativity and experiences.

Keywords Commercialization grant · cDNA · Cloning · Genomic · Host · NA-CBZ-L-lysine p-nitrophenyl ester (LNPE) · Polymerase chain reaction · Vector

2.1 Cloning

There are general and basic steps in DNA cloning. First, the desired DNA of interest is 'cut' from the source organism using restriction enzymes. Then, the DNA is 'pasted' into a vector and the ends of the DNA are ligated with the vector. The vector is then introduced into a host cell, which is often a bacterium or yeast and the cells are grown by a process called fermentation. The host cells will copy the vector DNA along with their own DNA and create multiple copies of the inserted DNA. Lastly, the vector DNA is isolated (or separated) from the host cell DNA and purified. The DNA that has been 'cut' and 'pasted' from an organism into a vector is called recombinant DNA, and DNA cloning is also known as recombinant DNA technology.

There are many cloning methods that can be chosen according to the objective of the research. cDNA cloning and genomic cloning are used for most research studies. Table 2.1 lists the advantages and disadvantages of each cloning method.

A. Amid (✉)

Biomolecular and Bioprocess Engineering Research Unit, Department of Biotechnology Engineering, Faculty of Engineering, International Islamic University Malaysia, P.O. Box 10, 50728 Kuala Lumpur, Malaysia
e-mail: azuraamid@iiu.edu.my

N. Hassan

Kolej University Insaniah, Kuala Ketil, Kedah, Malaysia
e-mail: dayang_2712@yahoo.com

© Springer International Publishing Switzerland 2015

A. Amid (ed.), *Recombinant Enzymes—From Basic Science to Commercialization*,
DOI 10.1007/978-3-319-12397-4_2

Table 2.1 The comparison between cDNA cloning and genomic cloning

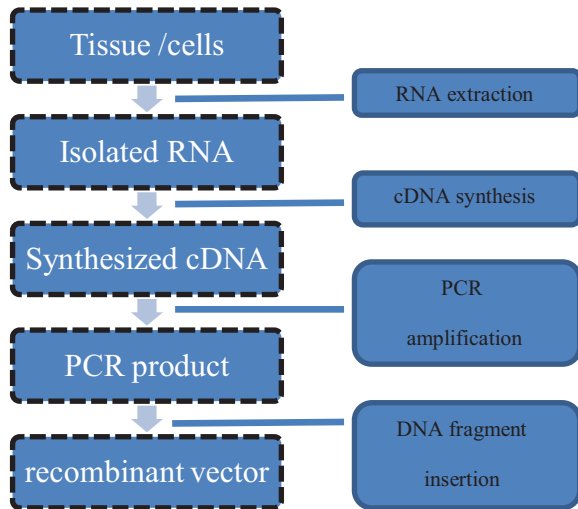
Cloning strategy	Advantages	Disadvantages
cDNA cloning	• The entire coding region can be obtained in one vector and transformed into bacteria for expression • Can obtain the exact coding region because the intron sequences are unnecessary • Proteins of interest can be produced in large quantities, greatly simplifying the task of protein purification	• Any regulatory information related to the introns may be lost • A source of mRNA from a tissue that expresses the desired produce must be found and mRNA is difficult to work with because it easily degrades
Genomic cloning	• To study a gene in its natural host, the whole gene is needed and it is easily obtained • The intron sequence may contain enhancers that enhance the transcription of genes in a gene cluster • Can use any DNA source from any tissue. This is because aside from cells of the immune system, all tissues contain the same DNA	• The gene sequences are extremely long and it will have to be pieced together in several clones • If the aim is to obtain a protein product, the gene cloned contains introns and therefore cannot be expressed in bacteria

For commercialization purposes, the enzymes produced must be in an active form so that they can be expressed by the new host. Therefore, cDNA cloning is more suitable for this objective.

2.1.1 *cDNA Cloning*

cDNA (short for complementary DNA or copy DNA) is a DNA copy of an RNA, and the starting material is usually mRNA. The first cDNA was cloned almost two decades ago (Watson and Demmer, n.d). The construction and screening of cDNA libraries has become one of the fundamental procedures in molecular biology. cDNA cloning allows researchers to obtain a sequence of DNA that directs the production of a specific protein and any procedure that optimizes cloning will be beneficial. The development of high efficiency cloning systems provided by commercial companies (e.g., Invitrogen, Promega and Clontech) has allowed the construction of a cDNA library to become a straightforward procedure. The principle behind cDNA synthesis is that an mRNA population isolated from a specific developmental stage should contain mRNAs specific for any protein expressed during that stage. Thus, if the mRNA can be isolated, then the gene can be studied. mRNA cannot be cloned directly, but a DNA copy of the mRNA can be cloned. This conversion is accomplished by the action of reverse transcriptase and DNA polymerase. The reverse transcriptase makes a single-stranded DNA copy of the mRNA. DNA polymerases cannot initiate synthesis de novo and depends on the presence of a primer. Many

Fig. 2.1 Simplified cDNA cloning strategy



mRNAs have a poly-A tail at the 3' end so oligo-dT is frequently used to prime DNA synthesis. The production of a second strand of DNA involves exposure of the DNA/RNA hybrid to a combination of RNAase-H and DNA polymerase. RNAase-H has the ability to cause single-stranded nicks in the RNA, and DNA polymerase can then use these single-stranded nicks to initiate second strand DNA synthesis. Finally, the double stranded product is introduced into an appropriate plasmid or lambda vector. The flow diagram of these mechanisms is shown in Fig. 2.1.

Another strategy for cDNA synthesis employs ribonuclease H, which recognizes the RNA component of a DNA-RNA hybrid and cleaves the RNA at a number of non-specific sites to produce short oligoribonucleotides attached to the cDNA. These fragments serve as primers for the polymerase to synthesize the second strand cDNA. DNA Polymerase I is used because the 5'-3' exonuclease activity is needed to remove RNA from the front of the enzyme. The newly synthesized strands of cDNA are joined in a vector by ligation using T4 ligase (+ ATP).

A cDNA library is a set of clones representing many mRNAs in a given cell type at a given time. Such libraries can contain tens of thousands of different clones. In general, this means that it is representative of all of the mRNAs present in a particular tissue. The technique to synthesize the cDNA is the same as described previously. However, probing or screening of the library is required to determine which plaques contain the gene of interest. A probe is a piece of DNA or RNA that is used to detect specific nucleic acid sequences by hybridization (binding of two nucleic acid chains by base pairing). Once a particular DNA fragment is identified, it can be isolated and amplified to determine its sequence. If the partial sequence of a gene is known and its entire sequence must be determined, then the probe should contain the known sequence of the gene of interest. A detailed review on cDNA cloning was provided by Harbers [1].

In addition to constructing the library to make clones, there are other more direct approaches. PCR cloning can be used for cloning both the genome and cDNA, and this approach avoids constructing a library.

There are two strategies commonly used in PCR cloning. The first is to treat the PCR products and blunt-end the PCR fragments prior to cloning. It is not easy to clone PCR products by simple blunt-end ligation into blunt-ended plasmid vectors. This difficulty arises because some thermostable DNA polymerases, including Taq DNA Polymerase, add a single nucleotide base extension to the 3'-end of blunt DNA in a template-independent fashion. The most common base added is adenine, and this produces an 'A overhang'. However, this strategy often results in low cloning efficiencies. The second strategy is to add restriction enzyme recognition sites to the ends of PCR primers. The PCR product is then digested and cloned into the desired vector. The amplicons generated with Taq DNA Polymerase typically have A overhangs, and this leads to the method referred to as T-vector cloning. In essence, the plasmid cloning vector is engineered to contain 3'-T overhangs that match the 3'-A overhang of the amplicon. The A-tailed amplicon is directly ligated to the T-tailed plasmid vector; therefore, there is no need for further enzymatic treatment of the amplicon other than the action of DNA Ligase.

The inclusion of control reactions is essential for monitoring the success of PCR reactions. Wherever possible, a positive control should be included to ensure the PCR conditions used can successfully amplify the target sequence. PCR is extremely sensitive and requires only a few copies of target template, so a negative control containing no template DNA should always be included to ensure that the solutions used for PCR have not become contaminated with the template DNA.

Polymerase Chain Reaction (PCR) has revolutionized cDNA cloning, and in some circumstances, it is not necessary to construct a library for the isolation of a particular cDNA clone. If some sequence information is available, then it is possible to clone the cDNA by PCR. It is worth an attempt because this approach is much faster and simpler than building and screening a cDNA library. This method is suitable if the target is one particular cDNA clone containing one DNA copy of one mRNA and the sequence is accessible. If this method fails to provide the clone, then a conventional library can be constructed.

There are several factors to consider when devising a cloning strategy and choosing between constructing a cDNA library or cloning by reverse transcription coupled PCR (RT-PCR). The factors are the abundance of the particular mRNA in the source organism, the length of mRNA and the screening factor (Watson and Demmer, n.d). Additionally, the vector used for cloning is also important because it will determine the nature of the recombinant enzyme produced. To produce a recombinant enzyme, an expression vector is chosen. There are many expression vectors available in the market, and each has advantages (Fig. 2.2).

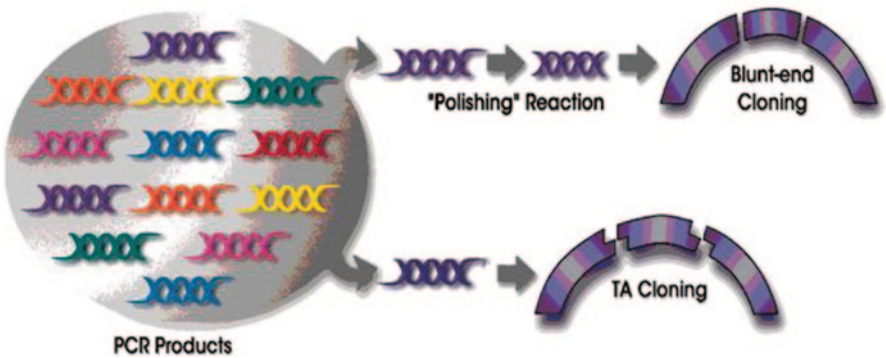


Fig. 2.2 Strategies of PCR cloning; blunt-end and TA cloning (www.invitrogen.com)

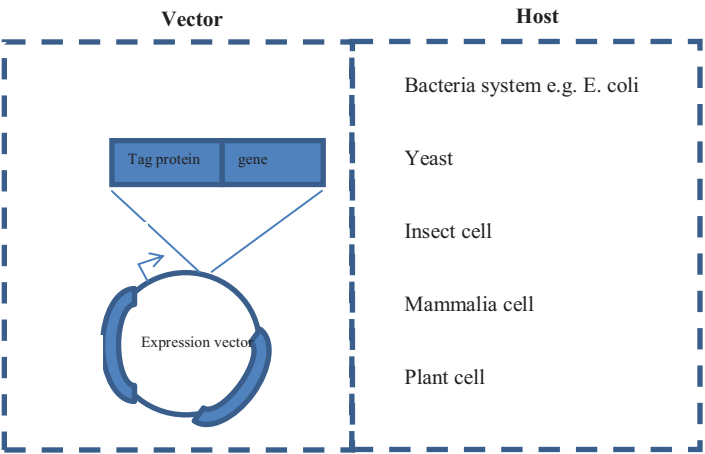


Fig. 2.3 Outline of the recombinant protein cloning

2.2 Vector for Cloning

The major activity of cloning will involve the selection of the suitable vector and expression host, as discussed thoroughly by Overton [2]. Figure 2.3 is a modification of the figure in the Overton study [2]. The focus of this subsection is to discuss vectors that are suitable for expression in *E.coli* hosts because there are too many vectors available depending on the host type. A researcher will initiate the cloning experiment using *E.coli* as a host, so it is worth using a vector that can be expressed in *E. coli*. Durani and co-workers discussed the expression vectors in detail and specifically noted ligation-free, traceless and tag-switching plasmids [3]. Additionally, Rosano and Ceccarelli [4] discussed suitable strategies to choose the correct vector. There is no specific vector that is given priority in the selection of expression vectors. However, there are several key factors that should be considered to

ensure active recombinant enzyme production. First, the selected vector must have high plasmid stability throughout fermentation. This is facilitated by the availability of antibiotic resistance genes. Examples of antibiotics that are frequently used are ampicillin, kanamycin, chloramphenicol and tetracycline. Second, the origin of replication should ensure the production of a high copy number of plasmid and this is necessary for large scale protein purification [3]. Next, there should be a cloning region or multiple cloning site in the vector where the insert will be ligated to the vector. There are five main methods of ligating into vectors, and they include the following: (1) ligation-based cloning; (2) ligation-independent cloning; (3) recombinogenic cloning (Gateway); (4) SLiCE cloning; and (5) TOPO cloning. As mentioned earlier in this chapter, the insertion method is also unique to the gene of interest, and there is no standard operating procedure for ligating the gene into the vector. This step always depends on the researcher's creativity as discussed by Hartley [5].

Other factors that need to be addressed are the transcription promoters and fusion tags. According to Durani [3], there are 3 main factors to consider when choosing a good promoter, and these factors include the following: (1) the promoter must be strong and able to express the recombinant protein in the cell; (2) the promoter must be able to express the product without adverse effects on the growth of the host cell; and (3) the promoter should be easily inducible either by a chemical or other method such as thermal induction. Recently [5, 6], auto-induction media was introduced. This system requires little user intervention during fermentation from the inoculation stage until cell harvesting. Blommel [6] claimed that this strategy is based on *E. coli* metabolic control, which automatically performs a shift from growth to recombinant protein expression and omits biomass monitoring for the correct time of inducer addition. These advantages are ideal for the large scale production of recombinant enzymes.

2.3 Host for the Recombinant Enzyme

Rosano and Ceccarelli [4] mentioned that before starting with other expression hosts, it is important to first try an *E. coli* strain because it offers dozen of candidate strains that might be suitable for your enzyme expression. Bacterial expression systems for heterologous protein production are attractive because of their ability to grow rapidly and at a high density, low cost and high productivity. *E. coli* facilitates protein expression because it is simple and inexpensive and a well-known genetic system. Bacteria can also be utilized with a large number of biotechnology tools. However, there are few major drawbacks encountered during the expression of heterologous proteins in *E. coli*. These problems include the formation of inclusion bodies, incorrect protein folding and degradation of the heterologous protein, especially when the heterologous enzyme is originally isolated from prokaryotes. Other expression systems, such as yeast, insect and mammalian systems, might be suitable if a researcher experiences any of these problems.

Table 2.2 Advantages and disadvantages of bacteria, yeast and mammalian host cell systems in cloning recombinant proteins

System	Type/strain	Advantages	Disadvantages	References
Bacteria	<i>E. coli</i> <i>Bacillus subtilis</i> <i>Streptomyces spp.</i>	• Fast growing • Inexpensive to culture • Easily scalable • Robust cellular structure	• Limited post-translational modification • Limited protein size and complexity • Usually require protein extraction and refolding • Traces of endotoxin in gram-negative species	[11, 12]
Yeast	7 strains <i>Saccharomyces cerevisiae</i> , <i>Pichia pastoris</i> , <i>Hansenula polymorpha</i> (<i>Pichia angusta</i>), <i>Arxula adeninivorans</i> , <i>Kluyveromyces lactis</i> , <i>Yarrowia lipolytica</i> , <i>Schizosaccharomyces pombe</i>	• Post-translational modification • Secretion • Fast growing • Inexpensive to culture • Robust cellular structure • Synthesis of complex and multi subunit proteins	• Glycosylation pattern dissimilar to humans • Complex scaling up fermentation	[11, 13]
Mammalian	Chinese ovary hamster (CHO) Murine cell lines Monkey cell lines Human cell lines	• Innate secretion mechanism • Synthesis of complex, multisubunit proteins • Greatest post-translational modification capabilities	• Slow growing • Requires expensive and complex media • Sensitive to osmotic shock and shear stress • Can harbor human pathogens	[11]

As of January 2009, there were 151 protein-based recombinant pharmaceuticals licensed by the US Food and Drug Administration and European Medicines Agency. Of the protein products, 30% were from *E. coli*, 19% were from yeast, 39% were from mammalian cell lines and 11% were from hybridoma cells [7]. Table 2.2 describes the advantages and disadvantages of each system. The final cloning strategy is still based on researcher creativity and experiences. Overton [2] suggested that although bacterial systems have many disadvantages, advances in bacterial post-translational modifications [8, 9] and protein release systems (such as periplasmic release and secretion systems) [10] will enable further improvements to be made by simplifying protein production processes and enabling even more rapid syntheses of target proteins.

References

1. Harbers M (2008) The current status of cDNA cloning. *Genomics* 91:232–242
2. Overton TW (2014) Recombinant protein production in bacterial hosts. *Drug Discov Today* 19:590–601
3. Durani V, Sullivan BJ, Magliery TJ (2012) Simplifying protein expression with ligation-free, traceless and tag-switching plasmids. *Protein Expres Purif* 85:9–17
4. Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* 5:172
5. Hartley JL (2006) Cloning technologies for protein expression and purification. *Curr Opin Biotechnol* 17:359–366
6. Blommel PG, Becker KJ, Duvnjak P, Fox BG (2007) Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnol Progr* 23:585–598
7. Ferrer-Miralles N, Domingo-Espin J, Corchero JL, Vazquez E, Villaverde A (2008) Microbial factories for recombinant pharmaceuticals. *Microb Cell Fact* 8:17
8. Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol* 22(11):1399–1408
9. Vera A, Gonzalez-Montalban N, Aris A, Villaverde A (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol Bioeng* 96:1101–1106
10. French C, KeshavarzMoore E, Ward JM (1996) Development of a simple method for the recovery of recombinant proteins from the *Escherichia coli* periplasm. *Enzyme Micro Tech* 19:332–338
11. Mohamed VP (2014) Development of effective CHO-K1 host system targeting at nutrient-regulated Insulin-like Growth Factor 1 (IGF-1) pathway. International Islamic University Malaysia, Kuala Lumpur
12. Nicholl DST (2009) An Introduction to genetic engineering. International Student Edition. Cambridge University Press, Singapore
13. Celik E, Calik P (2012) Production of recombinant proteins by yeast cells. *Biotechnol Adv* 30:1108–1118

Recombinant Enzymes - From Basic Science to
Commercialization

Amid, A. (Ed.)

2015, X, 188 p. 61 illus., Hardcover

ISBN: 978-3-319-12396-7