Methods of cloning

Alessandro Bertero[1], Stephanie Brown[1] and Ludovic Vallier[1,2]

[1]Wellcome Trust - Medical Research Council Cambridge Stem Cell Institute, Anne McLaren Laboratory for Regenerative Medicine and Department of Surgery, University of Cambridge, UK. [2] Wellcome Trust Sanger Institute, Hinxton UK.

Correspondence to: ab974@cam.ac.uk (A.B); lv225@cam.ac.uk (L.V).

Abstract

Molecular cloning is the collection of experimental procedures required to isolate and expand a specific fragment of DNA into a host organism in order to create a large number of identical copies. On top of allowing the study of a single DNA sequence of interest, molecular cloning is a powerful technique that permits the generation of complex combinations of DNA fragments for the most disparate downstream applications. As such, this process is key to most modern biomedical basic research studies and translational applications. This chapter will present an overview of modern molecular cloning procedures and their potential applications, and provide the basic concepts and protocols required to perform a traditional cloning experiment based on restriction enzyme digestion and ligation of two DNA fragments.

Keywords

Molecular cloning, recombinant DNA, restriction enzyme, ligation, transformation.

Objectives

- Introducing the concept of molecular cloning.
- Presenting the various molecular cloning techniques and their relative strengths and weaknesses.
- Describing step-by-step the methods involved in an experiment of "traditional" molecular cloning.
- Providing a broad outline of the applications of molecular cloning to address biomedical research questions.
- Exemplifying the use of molecular cloning to address a specific research question.

Introduction

Molecular cloning is the set of experimental techniques used to generate a population of organisms carrying the same molecule of recombinant DNA (see Glossary). This is first assembled *in vitro* and then transferred to a host organism that can direct its replication in co-ordination with its growth. This is usually achieved in an easy-to-grow, non-pathogenic laboratory bacterial strain of Escherichia coli. A single modified E. coli cell carrying the desired recombinant DNA can easily be grown in an exponential fashion to generate virtually unlimited identical copies of this DNA. As such, molecular cloning can be seen as an "*in vivo* PCR reaction", in which a desired piece of DNA can be isolated and expanded. However, molecular cloning allows more flexibility, better fidelity, higher yields, and lower costs than a PCR.

The development of molecular cloning techniques started with the discovery of bacterial enzymes known as "restriction endonucleases", which cleave DNA molecules at specific positions that are defined by their sequence. These restriction endonucleases allow researchers to break up large DNA fragments into smaller pieces that are then joined with other DNA molecules (vectors) using an enzyme called DNA ligase. The most commonly used vectors are known as plasmids, which are small circular DNA molecules physically distinct from the chromosomal DNA and capable of independent replication.

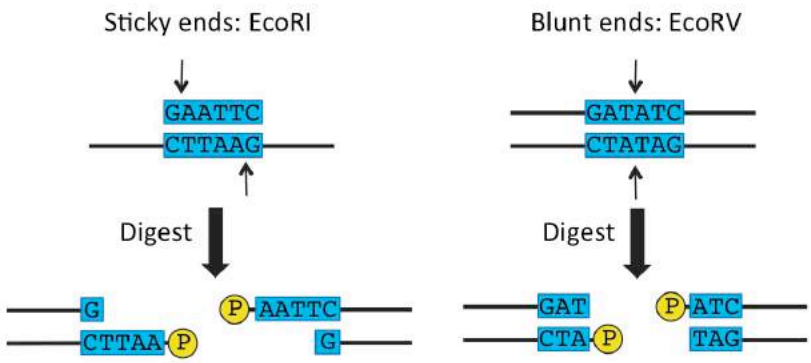

Figure 1. Type of DNA ends generated by restriction enzymes.
Representative examples of restriction enzymes generating sticky or blunt ends. The arrows indicate the cut sites. Phosphate groups attached to the 5' ends after restriction digestion are indicated in yellow.

Restriction endonucleases generate either "sticky ends", in which the DNA fragment has a single-stranded overhang (either on the 3' or 5' ends, see Glossary), or "blunt ends", in which no overhang is present (Figure 1). Both of these types of ends can be joined together (ligated), and each has its own associated advantages and disadvantages. For a sticky end fragment ligation to be successful, the two overhangs to be joined must have complementary Watson-Crick base pairing. However, this is not a requirement of a blunt end ligation, making it much more flexible. On the other hand, blunt end ligation is much less efficient than sticky end ligation due to a lack of binding stability of the two fragments. Importantly, sticky-ends can be enzymatically converted into blunt ends (either by "filling in" missing nucleotides, or by removing the overhangs), and vice versa (by using 5'-3' or 3'-5' exonucleases to create new overhangs). These cut-and-paste approaches are still widely used today and are commonly referred to as "traditional" (or conventional) cloning.

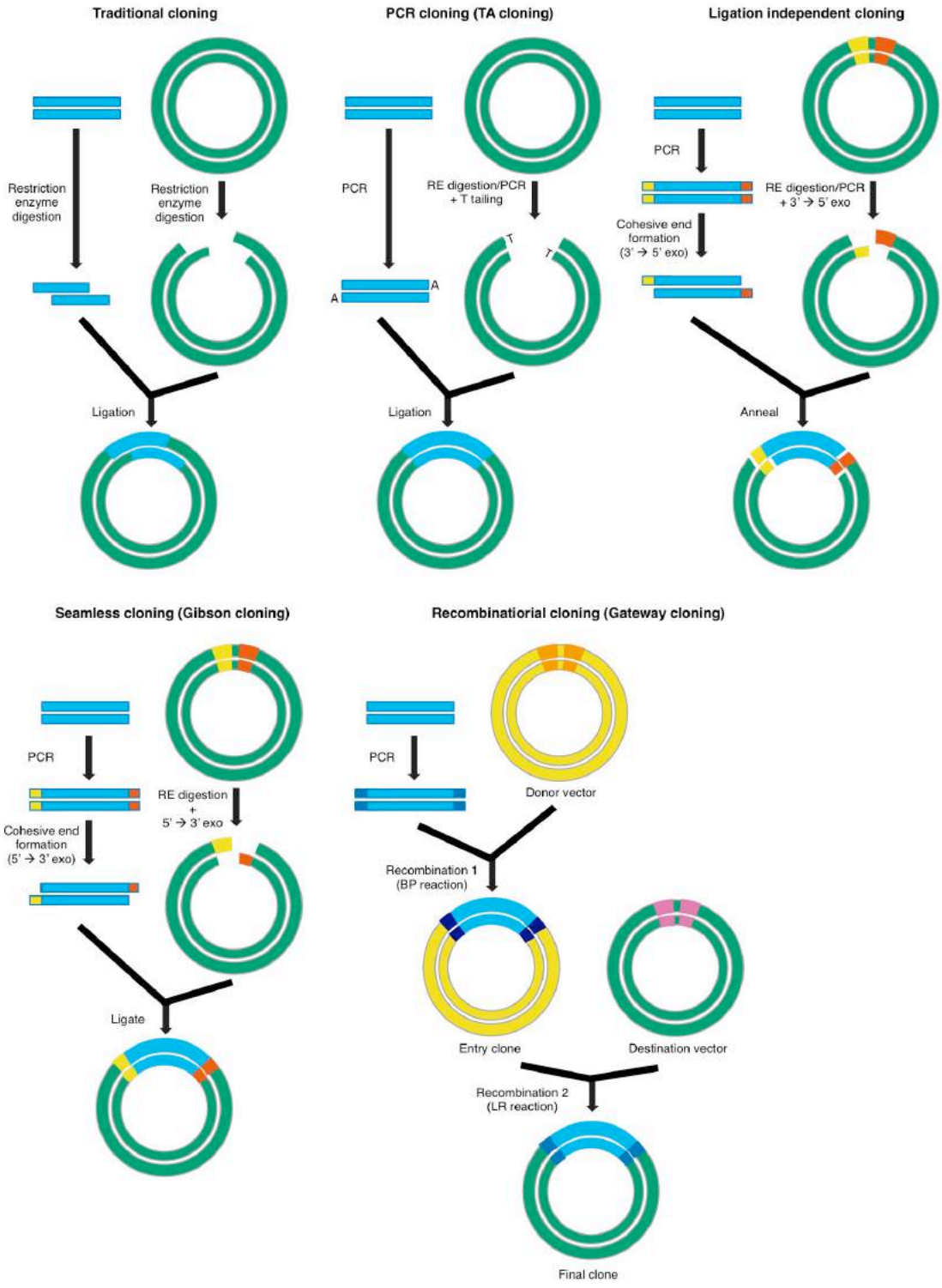

Figure 2. Overview of the main molecular cloning methodologies.
Schematic examples of cloning approaches. Refer to the text for detailed explanations.

Apart from this strategy, many other molecular cloning techniques have been developed, and nowadays there are many alternatives to perform a molecular cloning

experiment (Figure 2), each presenting distinct advantages and disadvantages (Table 1):

- PCR cloning involves the direct ligation of a PCR-generated DNA fragment without using restriction enzymes to cut the insert. One of the most commonly used PCR cloning method takes advantage of an adenine (A) residue that is added by the *Taq* polymerase at the 3' ends of the DNA fragments during the amplification process. These "A-tailed" products can be directly ligated with "T-tailed" vectors.

- Ligation independent cloning (LIC) is usually carried out by adding short sequences of DNA to the fragment to be cloned that are homologous to the destination vector (this is easily accomplished by using modified primers during the PCR amplification). Complementary cohesive ends between the vector and insert are then formed by using enzymes with 3' to 5' exonuclease activity (which chews back 3' ends to create 5' overhangs), and the resulting two molecules are then mixed together and annealed. The resulting plasmid has four single-stranded DNA nicks that are efficiently repaired by the host organism. Importantly, the resulting product does not contain any new restriction enzyme sites, nor other unwanted sequences, and is therefore "scar-free".

- Seamless cloning is a group of techniques that allow sequence-independent and scar-free insertion of one or more DNA fragments into a vector. The most well known of these methods is the Gibson Assembly Method, in which up to

ten fragments can be easily combined. Similarly to LIC, this relies on the addition of regions of homology at each end of the fragments to be cloned. Then, the combined action of an exonuclease (which chews back 5' ends to create 3' compatible overhangs), a DNA polymerase (which fills in gaps in the annealed fragments), and a DNA ligase (which seals the nicks in the assembled DNA) allows the generation of the recombinant DNA.

- Recombinatorial cloning uses site-specific DNA recombinases,enzymes capable of "swapping" pieces of DNA between two molecules containing the appropriate sequences (known as recombination sites). The most widely used system in this category is the Gateway cloning System (Life Technologies/Invitrogen), which relies on two proprietary enzyme mixes ("BP Clonase" and" LR Clonase") to swap a DNA fragment across various recombination sites. First the appropriate recombination sites are inserted by PCR on either side of the insert to be cloned, and then this is recombined with a Donor vector to create an Entry clone. This Entry clone is recombined again with a Destination vector (the required final vector) to make the final construct. Importantly, a large collection of Entry clones is already available on the market to facilitate Gateway cloning.

Regardless of the technique used to generate the recombinant DNA, molecular cloning is arguably the cornerstone of most biomedical sciences research labs. Indeed, the ability to isolate and expand a specific fragment of DNA that can be then introduced into a secondary host is often the first crucial step in both basic and translational scientific studies. The rest of the chapter will focus only on the

methodologies associated with traditional cut-and-paste molecular cloning. This method has many advantages over the more recently derived techniques (like the very low costs and simplicity of execution), and the concepts and technical skills that the reader will acquire by learning traditional cloning will be instrumental to then expand his/her study into other molecular cloning techniques.

Table 1. Comparison of the main molecular cloning methodologies.

Cloning method	Cost	Sequence dependency	Throughput	Assembly of multiple fragments	Directional cloning	Need for dedicated vectors	Examples of commercially available products
Traditional cloning (restriction enzyme-based)	Low	Yes (restriction enzyme sites)	Low to mid (can be increased by using ligation adapters)	Difficult for more than 2 fragments	Possible	No	-
PCR cloning	Medium (vectors)	No	High	Challenging (requires special modifications)	Difficult	Yes (for certain applications)	TOPO® TA
Ligation independent cloning	Medium (reagents)	Limited (vector)	Low	Yes	Yes	No	In-Fusion®
Seamless cloning	High (reagents)	No	Low	Yes	Yes	No	Gibson assembly GeneArt®
Recombinatorial cloning	High (reagents and vectors)	No	High	Challenging (requires special modifications)	Yes	Yes	Gateway® Echo Cloning™ Creator™

In principle

A traditional molecular cloning experiment can be divided into nine steps (Figure 3):

1) Selection of the host organism

2) Selection of cloning vector

3) Preparation of the vector

4) Preparation of the insert

5) Generation of the recombinant DNA

6) Introduction of the recombinant DNA into the host organism

7) Selection of the clones of organisms containing the vectors

8) Screening for clones with the desired recombinant DNA molecules

9) Expansion and isolation of the recombinant DNA

Importantly, before performing a cloning experiment it is always recommended to perform an *in silico* simulation of the procedure using dedicated software for DNA sequence manipulation (several free and commercial options are available). This same software is also useful to align DNA sequences and create publication-quality plasmid maps.

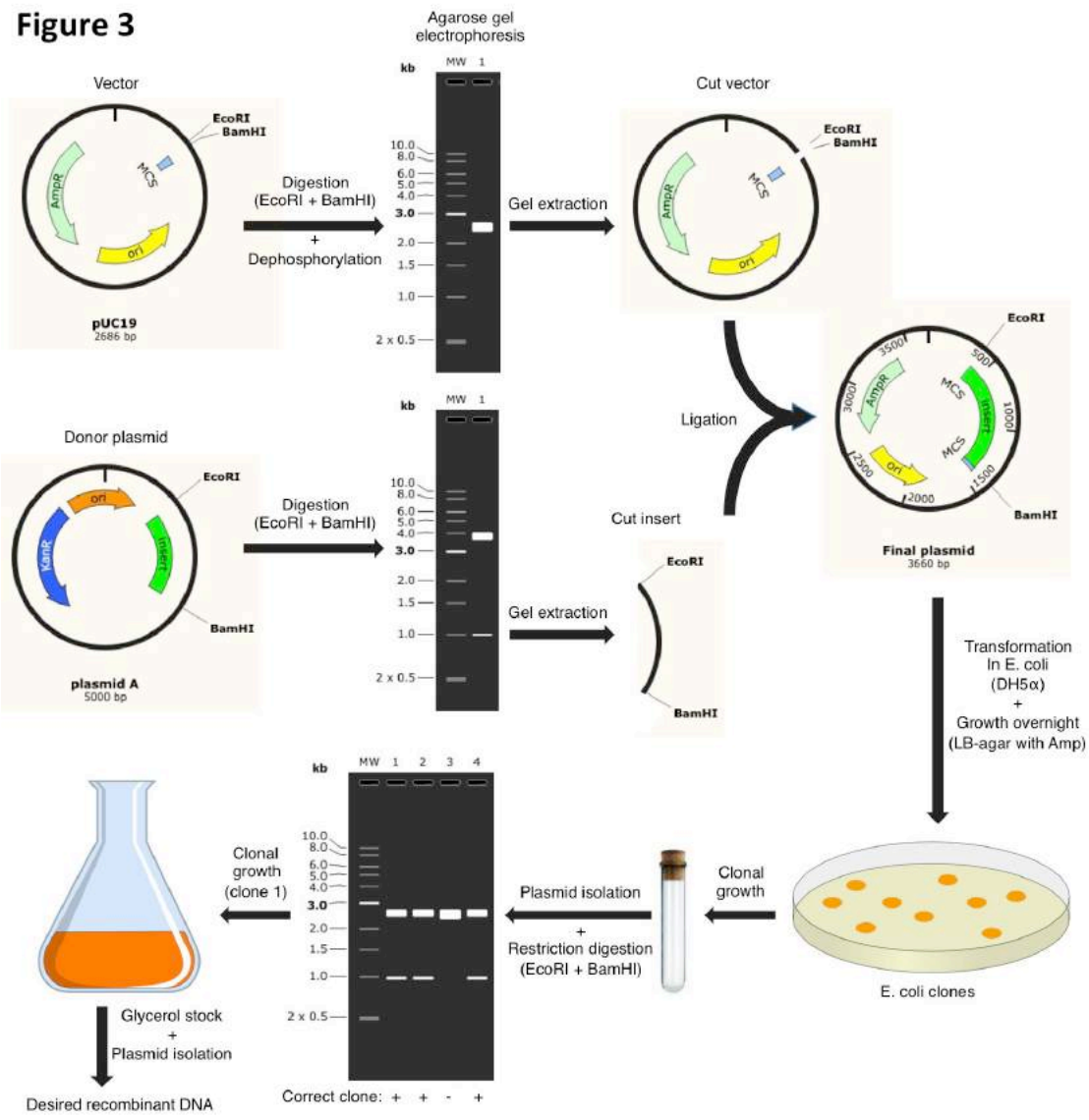

Figure 3. Overview of a traditional cloning experiment.
Schematic outline of the steps involved in a traditional cloning experiment. Refer to the text for detailed explanations.

Selection of the host organism.

As mentioned above, the most commonly used host organisms are non-pathogenic laboratory strains of E. coli. Several E. coli strains are available on the market, and these have been genetically engineered for optimal performance in certain applications and should be chosen accordingly (Table 2).

Table 2. Examples of the most common E. coli strains used for molecular cloning.

Strain	Key attribute	Optimized for:
BL21(DE3)	Expresses the T7 RNA Polymerase under the lacZ promoter (inducible by the lactose analog IPTG)	General expression of recombinant proteins
BL21(DE3) pLysE*	Lower basal expression levels of T7 RNA polymerase compared to BL21(DE3)	Expression of toxic recombinant proteins
DB3.1	Mutation in gyrA gene makes it resistant to toxin from ccdB gene	Propagation of plasmids expressing the ccdB gene (Gateway system)
DH5α	-	General cloning procedures
JM110	Lacks DNA methyltransferases	Growth of plasmids that must not be methylated
Origami2 (DE3)	Enhanced activity of enzymes that facilitate protein folding (reductases)	Expression of proteins poorly soluble
Rosetta2 (DE3)	Contains additional tRNAs for rare codons that are poorly expressed in E. Coli	Optimized expression of eukaryotic proteins by bypassing codon-bias problems
Stbl2	Lacks an enzyme involved in DNA recombination (recA)	Growth of plasmid containing with high potential to recombine (like lenti- and retroviral plasmids)
XL10 Gold	Exhibit the Hte (high transformation efficiency) phenotype	Transformation of large plasmids and preparation of DNA libraries

Selection of the cloning vector

Bacterial plasmids are by far the most commonly used cloning vectors given their simplicity of use and the fact that they are appropriate for most common cloning experiments as they can hold up to 20 kilo base pairs (kb) of foreign DNA. In its simplest form a plasmid must contain the two following DNA elements (Figure 2):

- Origin of replication (ORI).This recruits the DNA replication machinery and allows the propagation of the plasmid. Different types of ORI exist and they can affect the number of plasmid copies per bacterial cell.
- Selectable marker. This allows the selection of plasmid-containing bacteria. The most commonly used are drug-resistance genes such as those that confer ampicillin, kanamycin, and chloramphenicol resistance.

In addition to these two features there are several accessory elements that are found in most common plasmids (Figure 4):

- Multiple cloning site (MCS). A region engineered to contain multiple restriction enzyme sites to facilitate the cloning procedure.
- Promoter. This drives the expression of the cloned DNA. RNA Polymerase II or III promoters are respectively used for cDNA or short, non-coding RNA.
- Protein tag. This is fused to the cloned DNA to generate chimeric proteins with specific properties, for example a green fluorescent protein (GFP) to allow easy monitoring of the protein localization.
- Poly-adenylation signal. This is located after the cloned DNA and induces termination of the messenger RNA (mRNA) transcription by poly-adenylation.
- Secondary selectable marker. This allows for further selection of organisms that contain the plasmid. Drug resistance genes (such as for Neomycin and Puromycin resistance) are commonly used to select for mammalian cells

Finally, specialized vectors exist that allow the generation of recombinant viral particles that can be used to transduce mammalian cell types with the cloned DNA at high efficiency. Table 3 reports some of the most commonly used plasmids.

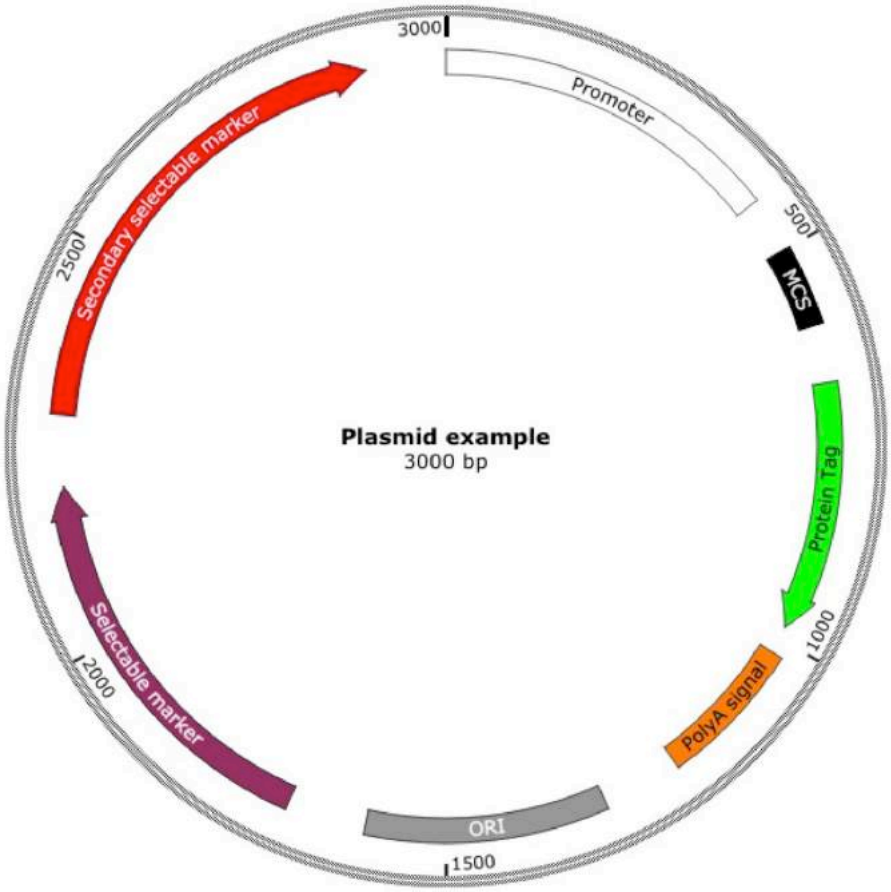

Figure 4. Schematic structure of a basic bacterial plasmid.
A representative plasmid which contains the most common features. Refer to the text for detailed explanations of each feature.

Table 3. Examples of commonly used bacterial plasmids

Aim	Key features	Plasmid
Mammalian protein expression	CMV promoter;Flag and HA tags.	pCDNA3 Flag HA
Transgenic mice generation	Neomycin resistance gene.	pBigT
Retroviral-mediated mammalian protein expression	Retroviral packaging sequences	pBABE
Lentiviral-mediated mammalian shRNA expression	Lentiviral packaging sequences; U6 promoter	pLKO.1
Recombinant protein production and purification from bacteria	Glutathione S-transferase (GST) tag	pGEX
Recombinant protein production and purification from bacteria	Maltose binding protein (MBP) tag	pMAL
Simple cloning experiment	-	pUC19

Preparation of the vector

The cloning vector can be prepared by two main methods:

- Restriction digestion. Blunt or sticky-ends can be generated depending on the restriction enzyme used (Figure 1, also see Glossary). Once digested, the vector is purified by size-selection using agarose gel electrophoresis.
- PCR. This generates blunt ends that can be further cut by restriction enzymes as described above. Compared to the previous method, amplifying the vector allows for more flexibility (as new restriction sites or other desired sequences can be introduced at the end of the vector by using modified primers containing such sequences), but the resulting vector has the potential to contain mutations introduced during the PCR and hence must be sequenced to exclude this. Before ligation, the vector must be purified.

If the prepared vector has compatible ends that have the potential to re-ligate without the presence of the desired insert the phosphate groups need to be removed from the 5' ends to prevent this from happening. Indeed, DNA ligase requires a phosphate group to be present on the 5' end of a DNA fragment in order to covalently link two fragments together. DNA ends generated by restriction enzymes retain phosphate groups on the 5' ends (Figure 1), however they can be removed *in vitro* using enzymes such as Alkaline Phosphatase. PCR products do not have phosphate groups on the 5' ends.

Preparation of the insert.

The DNA to be cloned (insert) can come from multiple sources:

- Genomic DNA. This can be purified from any organism of interest using appropriate protocols. Before cloning, it is either cut with a restriction enzyme or mechanically sheared to produce fragments with the desired size. The resulting DNA is used to prepare whole genomic DNA libraries.
- Complementary DNA (cDNA). This is double-stranded DNA obtained from cellular mRNA by means of reverse transcription. This can be used to generate cDNA libraries containing all expressed genes in a certain cell type.
- Plasmid DNA. This already contains the sequence of interest that needs to be transferred to a new vector. The sequence is first removed from this donor plasmid using appropriate restriction enzymes, then size-selected before cloning.
- PCR product. This can be obtained using appropriate primers from genomic DNA, cDNA, or plasmid. Restriction sites or other desired sequences can be added to the ends of the PCR product by using primers that include such sequences. Blunt-ended PCR products can either be used directly or cut with restriction enzymes. Size selection of the fragment of interest is usually performed.
- Synthetic DNA.This can be created *in vitro* without using any pre-existing DNA template and hence allows the highest level of flexibility in generating a desired sequence. Before cloning, these fragments are treated as for PCR products. Sequencing of the cloned product is always required as mutations are often introduced during the DNA synthesis procedure.

Regardless of its source, the insert DNA must contain ends which are compatible with the ones of the vector prepared as described above. Moreover, if a dephosphorylated

vector is to be used, it is essential for the insert to be phosphorylated on each 5' end. When the insert has been digested by restriction enzymes, the 5' ends retain these phosphate groups. However, PCR derived products or synthetic DNA fragments must be phosphorylated *in vitro* using enzymes such as T4 Polynucleotide kinase before ligation.

Generation of the recombinant DNA.

Once prepared, the vector and inserts are mixed and the ATP-dependent DNA ligation reaction is performed using enzymes such as the T4 DNA ligase.

Introduction of the recombinant DNA into the host organism

E. coli can be made competent for DNA uptake in two main ways:

- Chemical competence. This is achieved by pre-treating the cells with chemicals (often calcium chloride) under cold conditions, followed by a short pulse of heat shock, which together increases permeability of the cell membrane to the DNA. Plasmids up to 10kb can be efficiently introduced using this method, and given its simplicity this is the most commonly used technique to introduce recombinant DNA into bacteria.
- Electroporation. The cells are subjected to a brief electric shock that generates small pores in to the cell surface, thus allowing plasmid DNA to enter. This technology has a higher efficiency compared to chemical transformation and is used mostly to introduce very large plasmids.

Selection of the clones of organisms containing the vectors.

Bacteria are cultured on semi-solid agar-media petri dishes containing the appropriate selection agent (usually an antibiotic). This allows selection of the organisms that

carry this marker and therefore the desired plasmid. After overnight incubation, individual bacteria grow up into visible colonies of several million cells, which can be isolated, expanded and analyzed.

Screening for clones with the desired recombinant DNA molecules

This can be achieved by three main strategies:

- Colony PCR screening. The presence of the desired insert is identified by performing a PCR directly on a bacterial colony.
- Restriction digestion screening. The plasmid is first isolated from the bacterial clone and then subjected to a restriction digest that generates fragments of DNA of a certain size only if the desired recombinant DNA is present.
- DNA sequencing. This is performed on isolated plasmids and used to confirm the presence of the insert and its correct sequence.

Expansion and isolation of the recombinant DNA.

After screening, the correct bacterial clone carrying the desired recombinant DNA is expanded in liquid culture to amplify the plasmid. The amplified plasmid can be extracted and purified using appropriate protocols or commercial kits.

In practice

By now, it will be clear to the reader that it is not possible to have a protocol which can be universally applied to any traditional cloning experiment. In this section we will provide a "basic protocol" that can be used as a practical guide and tailored to different experimental requirements. Application of this protocol will require the use of basic molecular biology techniques that will be only briefly summarized here. More detailed protocols for these techniques can be found in other chapters of this text, in other laboratory manuals (1), or from the suppliers of commercial molecular cloning products/kits(2–5).

The aim of this general protocol is to transfer a 1kb DNA fragment from a donor plasmid (plasmid A, size of 5kb) into a destination plasmid for a different downstream application. The DNA of interest can be removed by digestion with two commonly used restriction enzymes (EcoRI and BamHI, see Figure 3).

Selection of the host organism

For this example a DH5α E. coli strain will be used.

Selection of cloning vector

For this example a pUC19 plasmid vector will be used.

Preparation of the vector

- Cut 10μg of pUC19 by mixing with EcoRI and BamHI in the appropriate digestion buffer and incubating at 37 °C for 2h.

NOTE: refer to the enzyme supplier for recommended digestion conditions.

- Run the vector on a 1% agarose gel for DNA electrophoresis and extract the size-fractionated vector using a commercially available (or an equivalent in-house protocol) gel-extraction kit.

Preparation of the insert

- Cut 10µg of plasmid A with EcoRI and BamHI as described above.
- Size-fractionate the insert from the vector backbone using a 1% agarose gel.

Generation of the recombinant DNA

- Quantify the concentration of vector and insert previously prepared by measuring the absorbance of the DNA solution at 260nm using a spectrophotometer.
- Prepare the ligation mix in 10µl by adding 50ng of vector, 50ng of insert, the T4 DNA ligase, and the T4 DNA ligase buffer. Incubate the mix for 2h at room temperature.

NOTE: Refer to the enzyme supplier for recommended ligation conditions. In this example the cut pUC has a size of 2665bp whilst the fragment is 1000bp long. To calculate an equimolar quantity of the two fragments use the formula shown below:

2665bp/50ng = 1000bp/Xng.

(1000bp*50ng)/2665 = 18.76ng.

It is recommended to use a 1:3 vector to insert molar ratio, as such use 18.76ng*3 = 56.28ng (rounded down to 50ng).

- OPTIONAL: perform a negative control ligation by adding 50ng of vector but no insert.

NOTE: this optional reaction can be useful to estimate the level of background due to contaminating uncut plasmids or ligation by-products and therefore the efficiency of the cloning experiment.

Introduction of the recombinant DNA into the host organism

- Transform 50µl of chemically competent DH5α E. Coli with 5µl of the ligation.

NOTE: refer to the supplier for the recommended transformation protocol.

Selection of the clones of organisms containing the vectors

- Plate the transformed E. coli onto Luria Bertani (LB) broth-agar plates containing 100µg/ml ampicillin and incubate overnight at 37 °C in a humidified bacterial incubator.

NOTE: in this case ampicillin is used as pUC19 carries an ampicillin resistance gene.

Screening for clones with the desired recombinant DNA molecules

- Individually pick 8 bacterial colonies using a sterile tip and inoculate them in 5ml of LB supplemented with 100µg/ml ampicillin in a microbiology tube. Grow them for 16h at 37 °C with agitation (225rpm) in a bacterial orbital shaker.

NOTE: if a negative control ligation was performed, the number of clones screened can be adjusted to take into account the level of background colonies observed.

- Transfer 1ml of the liquid culture to a fresh tube and store at 4 °C as stock and backup.
- Extract the plasmid from the remaining liquid culture using a commercial kit (or an equivalent in-house protocol).
- Digest the extracted plasmids using EcoRI and BamHI as described above.
- Run the digested plasmids on a 1% agarose gel to identify clones containing the insert.

NOTE: in this case correct clones will have two bands (the vector at 2.6kb, and the insert at 1kb), while incorrect clones from self-ligated vectors will have only one band (at 2.6kb).

Expansion and isolation of the recombinant DNA

- Inoculate 500µl of the bacterial liquid culture previously stored at 4 °C into 250ml of LB supplemented with 100µg/ml ampicillin in a microbiology vessel. Grow for 16h at 37 °C in agitation (225rpm) in a bacterial orbital shaker.
- Save 1ml of liquid culture, add 1ml of 50% glycerol, mix and store at -80 °C as long-term bacterial stock.
- Extract the plasmid from the rest of the liquid culture using a commercial kit (or an equivalent in-house protocol).
- The recombinant DNA of interest is now ready for any downstream application.

Applications

Molecular cloning is arguably one of the cornerstones of most modern biomedical basic studies and translational applications. Initially developed to study a single DNA sequence, molecular cloning techniques now allow an unprecedented ability to generate complex combinations of DNA fragments.

Study of gene function

With regards to basic research studies, on top of being an important tool to determine the sequence of a particular DNA fragment, cloning is essential to characterize the function of both genes and non-coding elements of the genome. Gene function can be investigated by cloning a cDNA into a mammalian expression vector to induce overexpression in a target organism (gain of function studies), or by cloning a specific short-hairpin RNA, a sequence capable of suppressing the expression of the gene of interest using the miRNA pathway (loss of function studies)(6). Other methods to suppress gene function include the use of programmable genome editing tools to generate knock-out cells or organisms by disrupting a gene sequence. These tools include Zinc-finger nucleases, transcription activator-like (TAL) effector nucleases (TALENs), or CRISPR/Cas9 nucleases(7), all of which must first be cloned into specific vectors. Moreover, gene function can also be assessed by introducing specific mutations through site-directed mutagenesis techniques, or by generating protein truncation mutants, both of which rely on molecular cloning procedures (8). Among other examples, all of these technologies are key to the production of transgenic animals models of human diseases that can be used to find novel therapeutic targets

and screen for potential drugs(9).

Study of genomic regulatory regions

The function of non-coding elements can also be characterized by cloning putative gene promoters, enhancers or silencers into specific vectors that allow measuring their ability to regulate gene transcription(10). This is can be done both *in vitro* and *in vivo* by measuring the activity of a reporter gene (such as luciferase, β-galactosidase, and GFP) cloned downstream of the genomic element of interest. In the case of established gene regulatory elements, similar reporter constructs can be instead used to visualize a certain cell type. Furthermore, by using tissue-specific promoters in the context of programmable gene editing technologies (such as Cre/lox-mediated genetic recombination), it is possible to create lineage tracing tools that are invaluable to modern developmental biology as they allow the study of the specification process of a cell type of interest from its early progenitors (11).

Translational applications

Molecular cloning is also a key technique in the context of biomedical translational applications. For example, it is key for the production of recombinant proteins for therapeutic or diagnostic uses. Notable examples of these are growth factors (such as insulin, erythropoietin, and growth hormone), enzymes, antibodies, and vaccines(12). These are commonly produced and then purified from E. coli, yeasts or insect cells (baculovirus) that are transformed with the appropriate expression vectors. Molecular cloning is also used to generate transgenic plants and animals with the aim of

improving their nutritional value (for example by introducing vitamins or essential nutrients), or introducing genes that code for useful pharmaceuticals (a process known as molecular farming(13)). Finally, gene therapy applications rely on complex molecular cloning experiments. In these cases a gene of interest is inserted into a modified virus with the aim of delivering it into a patient to induce a therapeutic response(14).

Scenario

In this section, we will provide a case study of how a series of simple traditional cloning experiments can lead to the generation of an extremely powerful tool to address a complex biological question.

Let's imagine we recently performed a screening experiment aimed at identifying novel regulators of liver differentiation. Among these potential targets, one transcription factor (TF X) appears very interesting as our screening suggests that high levels of TF X might be capable of inhibiting hepatocyte specification. As such, our aims are to: (1) study the effect of overexpressing TF X during liver development with regards to its specification; (2) determine the transcriptional changes induced by TF X overexpression.

To address these two biological questions, we decide to generate a plasmid that allows: (1) hepatocyte-specific overexpression of TF X to exclude potential confounding effects due to TF X overexpression in other cell types and to avoid premature expression of TF X at early stages of hepatocyte development; (2) co-expression of a EGFP reporter gene, to be able to isolate TF X-overexpressing cells by flow-cytometry to perform detailed transcriptional analysis in a pure population; (3) selection of mammalian cells carrying the plasmid, to allow the generation of a stable transgenic organism for our studies.

Therefore, we perform the following cloning procedures:

- Cloning of a hepatocyte-specific promoter (apolipoprotein A-II, APOA-II

(15)) into a vector containing a mammalian secondary selectable marker (pBigT, containing Neomycin resistance gene). We amplify the APOA-II promoter from genomic DNA, then clone this into the pBigT vector using the traditional cloning method described earlier. This generates a pBigT-APOAII plasmid

- Cloning of the TF X into the pBigT-APOAII plasmid. TF X is amplified by PCR from cDNA obtained from liver cells in which TF X is expressed. This cDNA is cloned downstream of the albumin promoter to generate pBigT-APOAII-TFx.
- Cloning of EGFP reporter gene. In order for the EGFP to be expressed from the APOA-II promoter together with TF X we take advantage of an internal ribosome entry site (IRES) sequence. This allows for translation initiation in the middle of an mRNA and generates an open reading frame containing TF X-IRES-EGFP. For the cloning, we cut an IRES-EGFP sequence from a pre-existing plasmid (pIRES-EGFP-puro) and ligate it downstream to TF X in order to generate pBigT-APOAII-TFx-IRES-EGFP. Note that the pBigT contains a human beta globin poly-adenylationsequence that terminates the transcription of the cloned DNA (Figure 5).

Once generated, this plasmid can then be stably introduced into a host mammalian organism to finally address our biological questions. For example, human pluripotent stem cells can be genetically modified to integrate such plasmid, and then *in vitro* hepatocyte differentiation can be performed to study the role of TF X in human liver development(16). Moreover, transgenic mice carrying our plasmid can also be generated to validate the function of TF X during *in vivo* liver development(9).

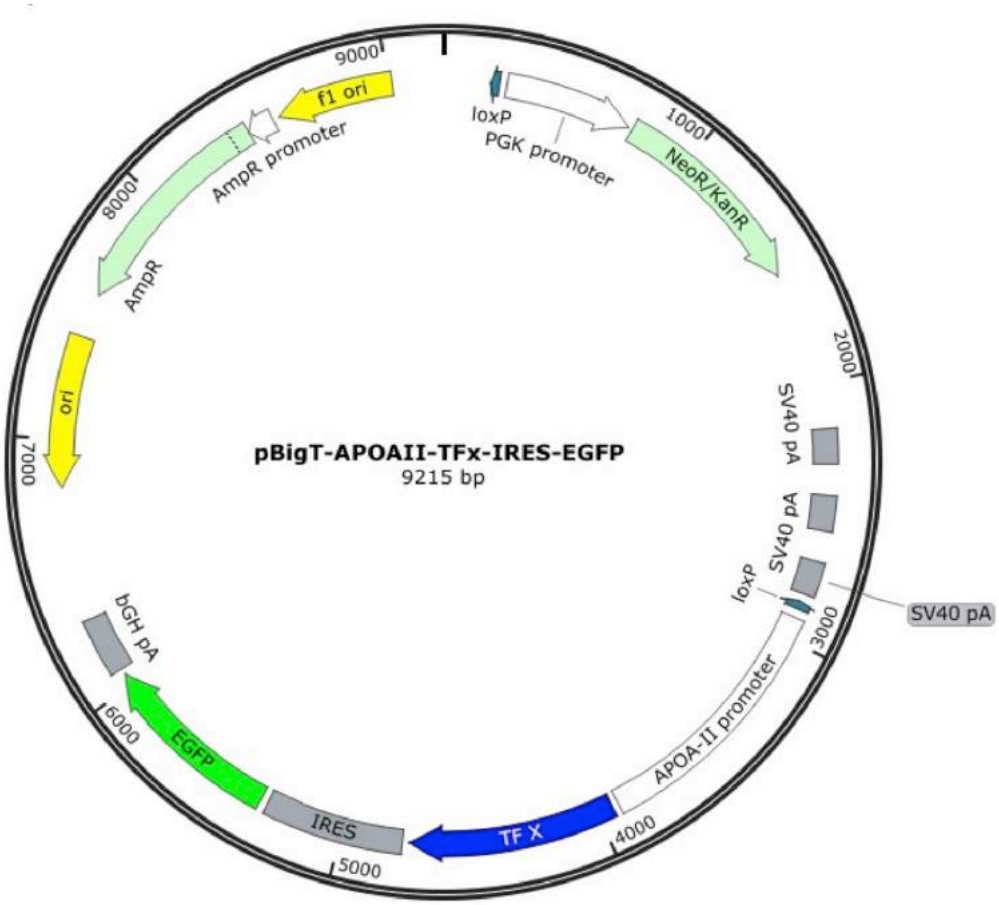

Figure 5. Example of an application of molecular cloning experiments.
Structure of the pBigT-APOAII-TFx-IRES-EGFP plasmid described in the text. The two loxP sequences can be used to eliminate the Neomycin resistance cassette using Cre-mediated recombination. The three SV40 polyadenylation sequences are used both as terminators for the Neomycin gene and as insulators between the two expression cassettes on the vector.

Key limitations

While being widely used because of its simplicity and low costs, the traditional cloning procedure presented in this chapter has some limitations (Table 1). First, it relies on the use specific restriction enzymes, which create a sequence dependence that can complicate a cloning experiment. Secondly, throughput is limited unless specific modifications (such as the use of ligation adapters) are put in place. However, this also often results in lower cloning efficiency. Finally, the assembly of multiple fragments by traditional cloning either involves multiple steps (which is time consuming), or can be very inefficient (especially when more than two inserts have to be ligated to a vector at the same time). As discussed earlier, these various limitations can be easily overcome by more expensive or more complicated molecular cloning techniques (Table 1).

Overall, molecular cloning does not suffer from major limitations. Indeed, careful planning of a cloning experiment will generally result in the predicted positive outcome. However, problems can arise with regards to the biological applications of cloning experiments. For example, overexpression of a transgene can be challenging due to the transcriptional silencing of the promoter used. This is particularly common for viral promoters such as CMV when used in certain cell types like stem cells. Moreover, the generation of complex vectors harboring multiple promoters and transgenes can result in sub-optimal expression of such elements due to so called "promoter interference" mechanisms(17). Finally, complex vectors are also subjected to size limits.

Troubleshooting

Stage	Common problem	Possible cause and potential solutions
Preparation of the vector/insert	Incomplete or no digestion.	- Digestion is inhibited by DNA methylation. Check if the enzyme used is methylation sensitive, and if so use a DNA methyltransferase-free E. coli strain to grow the plasmids. - DNA is contaminated with salt, solvents or other inhibitors of the restriction enzyme. Clean up the plasmids using phenol-chloroform extraction or commercial kits - Presence of slow-sites or supercoiled DNA result in less-efficient digestion. Increase the enzyme units used and/or the incubation time
Preparation of the vector/insert	Unexpected bands	- Enzymes show non-specific cleavage (star activity) under conditions of low ionic strength, high pH, high (>5%) glycerol concentration, high enzyme concentration and prolonged incubation times. If possible use enzymes modified to lack star activity (commercially available); otherwise, decrease glycerol content, enzyme units used in the digestion, and perform shorter incubation times - Partial digestion results in bands of unexpected size. Follow recommendations for incomplete digestion described above.
Introduction of the recombinant DNA into the host organism	No colonies or very few colonies	- Inefficient transformation. Test transformation using an uncut plasmid and calculate the transformation efficiency. If this is too low re-prepare competent cells or by new commercial ones. If the construct to be transformed is >10kb consider performing electrophoration. - Ligase is not functional. Test the ligase on a plasmid digested with a single enzyme that generates sticky ends. If no colonies are obtained, buy a new ligase and ligase buffer. Note that ligase buffer must not be frozen-thawed multiple times as it contains ATP. - Inefficient ligation conditions. Vary the vector to insert molar ratio from 1:1 to 1:20, test different temperatures (4 °C, 16°C and 25 °C) and increase the ligation time to

		overnight (16h). If this does not give better results, the vector or insert might be contaminated with ligation inhibitors: consider clean up. - The DNA is toxic to the cells (only for bacterial expression vectors). Incubate the plates at low temperature (25 °C), or use E. coli with more tight transcriptional control over the toxic DNA
Screening for clones with the desired recombinant DNA molecules	High background of colonies without the insert	- Dephosphorylation of the vector was insufficient. Increase the phosphatase units used and/or the incubation time. - Contamination of uncut plasmid. Follow recommendations for incomplete digestion described above.
Expansion and isolation of the recombinant DNA	Low plasmid yield	- Incomplete antibiotic selection. Use fresh antibiotic and incubate cells for not more than 16h to avoid degradation of the antibiotic. - Inappropriate origin of replication. For large plasmids (>15kb) using a low-copy ORI might be required. In this case, larger cultures will also be needed.

Conclusions

In this chapter, we provided an overview of the various molecular cloning techniques and described their application to modern biomedical research. Moreover, we provided a general cloning procedure to introduce the basic concepts involved in molecular cloning experiments. The reader should now have the sufficient background to directly apply these procedures in his research questions, as well as to expand his understanding of other molecular cloning techniques using the resources recommended in the chapter.

Acknowledgments

We thank Dr. Fotis Sampaziotis and Simon Rodler for their critical comments on this chapter.

Key References and Suggested Further Reading

1. Green MR, Sambrook J. Molecular Cloning: A Laboratory Manual (4th Edition). Cold Spring Harbor Laboratory Press; 2012.

2. New England Biolabs Cloning and Synthetic Biology [Internet]. Available from: https://www.neb.com/applications/cloning-and-synthetic-biology#tabselect0

3. Promega Cloning Applications Guide [Internet]. Available from: https://www.promega.co.uk/resources/product-guides-and-selectors/protocols-and-applications-guide/cloning/

4. Sigma Aldrich Nucleic Acid Purification Resources [Internet]. Available from: http://www.sigmaaldrich.com/life-science/molecular-biology/dna-and-rna-purification.html

5. QIAGEN DNA Sample Technologies [Internet]. Available from: https://www.qiagen.com/gb/products/catalog/sample-technologies/dna-sample-technologies/

6. Kappel S, Matthess Y, Kaufmann M, Strebhardt K. Silencing of mammalian genes by tetracycline-inducible shRNA expression. Nat Protoc. 2007 Jan;2(12):3257–69.

7. Kim H, Kim J-S. A guide to genome engineering with programmable nucleases. Nat Rev Genet. 2014 May;15(5):321–34.

8. Zawaira A, Pooran A, Barichievy S, Chopera D. A discussion of molecular biology methods for protein engineering. Mol Biotechnol. 2012 May;51(1):67–102.

9. Mouldy S. Transgenic Animal Models in Biomedical Research. Methods Mol Biol. New Jersey: Humana Press; 2006 Oct 15;360:163–202.

10. Jiang T, Xing B, Rao J. Recent developments of biological reporter technology for detecting gene expression. Biotechnol Genet Eng Rev. 2008 Jan;25:41–75.

11. Kretzschmar K, Watt FM. Lineage tracing. Cell. Elsevier Inc.; 2012 Jan 20;148(1-2):33–45.

12. Soler E, Houdebine L-M. Preparation of recombinant vaccines. Biotechnol Annu Rev. 2007 Jan;13:65–94.

13. Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R. Molecular farming in plants: host systems and expression technology. Trends Biotechnol. Elsevier; 2003 Dec 12;21(12):570–8.

14. Kay MA. State-of-the-art gene-based therapies: the road ahead. Nat Rev Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011 May;12(5):316–28.

15. Yang G, Si-Tayeb K, Corbineau S, Vernet R, Gayon R, Dianat N, et al. Integration-deficient lentivectors: an effective strategy to purify and differentiate human embryonic stem cell-derived hepatic progenitors. BMC Biol. 2013 Jan;11:86.

16. Sampaziotis F, Segeritz C-P, Vallier L. Potential of human induced pluripotent stem cells in studies of liver disease. Hepatology. 2014 Dec 11;

17. Shearwin KE, Callen BP, Egan JB. Transcriptional interference--a crash course. Trends Genet. 2005 Jun;21(6):339–45.

Glossary

Term	Definition
Molecular Cloning	Procedure by which a DNA fragment of interest is isolated and expanded in a host organism.
Recombinant DNA	DNA molecule that contains sequence originating from different organisms. Usually used to refer to the product of a molecular cloning experiment.
Vector	DNA molecule that is able to self-propagate in a host-organism and is consequently used as a carrier of a desired DNA fragment.
Insert	DNA molecule of interest to be cloned into a vector.
Plasmid	Common cloning vectors able to self-propagate in bacteria.
Restriciton enzyme	Endonuclease capable of cutting a DNA fragment at a specific sequence (restriction site).
3'-end	Three prime end: single-stranded DNA terminus consisting of the third carbon of deoxyribose, which is normally linked to an hydroxyl group.
5'-end	Five prime end: single-stranded DNA terminus consisting of the fifth carbon of deoxyribose, which can be linked to a phosphate group.
3'-5' and 5'-3'	Conventional terminology that defines directionality in a single strand of DNA fragment with respect to its ends.
Sticky end	An overhanging piece of single-stranded DNA at the end of a double stranded DNA fragment. The term sticky is used because a complementary piece of single-stranded DNA from another fragment can lead to semi-stable binding of the two molecules.
Blunt end	End of a double-stranded DNA fragment that does not contain overhangs.
Compatible ends	Ends of two DNA fragments that can be efficiently ligated (either blunt ends, or sticky ends with complementary overhangs).
Ligase	Enzyme capable of forming covalent bonds between two single-stranded DNA fragments located in close proximity.
Transformation	Bacterial uptake of DNA from the extracellular environment.