

# GENE CLONING

## 1. Isolation and purification of mRNA

mRNA Abundance

Isolation of Total RNA

Guanidinium Isothiocyanate Acid - Phenol Method

Isolation of mRNA

Oligo(dT) Method

Magnetic Oligo(dT) Bead Purification Method

Quality Control Assessment of RNA

Spectrophotometric Method

Agarose Gel Electrophoresis

## 2. Gene Cloning Strategies

Choice of Gene Cloning Strategies

Starting with cDNA Synthesis

cDNA Synthesis prior to PCR

PCR

PCR Cycles

Theoretical Number of ds cDNA copies

PCR Primer Design

PCR Optimization

RT-PCR

RACE-PCR

Primer design for first strand cDNA synthesis

Second Strand cDNA synthesis

3' or 5' RACE-PCR

Gene Specific Primers (GSP) Design

Purification of RACE-PCR Products for Full Length Cloning

Full Length Cloning of the Gene of Interest

## 3. cDNA Library Construction

cDNA Library Construction

Selection of a Cloning Vector

Plasmids

Lambda Phage

Cosmids

Ligation of the cDNA into the Cloning Vector

Restriction Digestion and Restriction Enzymes

Ligation of the cDNA into a Vector

Purification of cDNA PCR Products for Ligation into a Cloning Vector

Bacterial Transformation

## 4. cDNA Library Screening

General Screening for Transformed Bacteria

Stringent Screening for Transformed Bacteria Harboring the Gene of Interest

Different Screening Approaches

[Design of Antibodies for Screening](#)

[Design of Oligonucleotide Probes for Screening](#)

[The cDNA Library Hybridization Probe Screening Process](#)

[Protein Expression Testing of Each Clone](#)

[Cell-Free, In Vitro Transcription - Translation Protein Synthesis](#)

[Storage of a Desired Clone and cDNA](#)

[5. Development of a Stable Expression System](#)

[6. Miscellaneous](#)

[DNA Sequencing](#)

**1. Isolation and purification of mRNA**

### mRNA Abundance

Isolation of full length, undegraded, pure messenger RNA (mRNA) is important for gene cloning experiments. mRNA can be copied into double stranded DNA (known as complementary DNA or cDNA), which is then used to clone the gene of interest.

A typical mammalian cell has about 10-30 picograms of total RNA per cell. Total RNA consists of ribosomal RNA (rRNA) (~80% of total RNA), transfer RNA (tRNA) and small nuclear RNA (~10-16% of total RNA) and mRNA (~1-3% of total RNA). A viable cell usually has 15,000 active genes that encode proteins via mRNA. Transcription of all active genes in a cell results in 200,000 - 1,000,000 copies of mRNA per cell.

The relative abundance of mRNA derived from one gene is determined by the number of cellular copies of that particular mRNA. A cell typically has 10-20 genes that are transcribed in high abundance - about 12,000 copies of mRNA per gene. Another 500 genes are transcribed in medium abundance - about 500 mRNA copies per gene. The remaining genes, about 11,000, are transcribed in low abundance - about 5-15 copies of mRNA per gene. Most mammalian mRNA species are 1.5-2.0 kilobases in length, and are polyadenylated, i.e., have a series of adenine nucleotide residues (a poly(A) tail) at the 3' end.

### Isolation and Purification of mRNA for cDNA Synthesis

Isolation of full length, undegraded, pure mRNA is critically important for gene cloning experiments. mRNA is difficult to use experimentally because it is rapidly degraded if not handled with utmost care. For this reason, mRNA is transcribed in vitro into double stranded DNA (known as complementary DNA or cDNA), which is more stable and easier to use for cloning. The cDNA is used to make cDNA libraries. cDNA libraries represent the information encoded by all mRNA molecules expressed by a particular tissue or organism.

### **Isolation of Total RNA**

Isolation of total RNA results in three RNA species in the sample: mRNA, rRNA and tRNA. An additional purification step may be performed to isolate mRNA from rRNA and tRNA

#### **Guanidinium Isothiocyanate - Acid - Phenol Method**

Isolation of intact, undegraded mRNA requires scrupulous attention to analytical cleanliness. Contamination of the preparation will result in rapid enzymatic degradation by ubiquitous ribonucleases. Ribonucleases exist everywhere (or so it seems)! All solutions, glassware and plasticware must be scrupulously cleaned by special treatment procedures to eliminate exogenous ribonucleases.

The guanidinium isothiocyanate method is a one - step method developed by Chomczynski and Sacchi. This method uses guanidinium isothiocyanate to lyse cells and rapidly denature endogenous and exogenous ribonucleases. The liberated RNA, DNA and proteins are extracted and isolated by partitioning against a water - immiscible organic phase. RNA is quite hydrophilic compared to DNA and proteins. The total RNA resides in the aqueous phase, while the DNA and proteins partition into the interphase (the boundary of the immiscible aqueous and organic layers) and organic phenol phase.

Cell cultures are grown under conditions that promote the greatest abundance of the mRNA of interest. The culture media is removed and a guanidinium isothiocyanate solution is added to the cultures to lyse the cells, releasing DNA, RNA and proteins. This solution, known as a lysate, is transferred to a clean test tube for further processing (described below).

To obtain mRNA from tissues that express the gene of interest, one must quickly freeze fresh tissue in liquid nitrogen; otherwise substantial mRNA degradation will occur by cellular RNAses. The tissue is then ground into a powder using a liquid nitrogen cooled mortar and pestle. A fraction of the ground tissue is weighed and 1 gram (or less) of the pulverized tissue is added to a guanidinium isothiocyanate solution, thereby releasing DNA, RNA and proteins. This solution is transferred to a clean test tube for further processing (described below).

The test tube containing the guanidinium isothiocyanate solution of DNA, RNA and proteins is further processed as follows. Sodium acetate, water-saturated phenol and chloroform (or bromochloropropane) are added to the test tube. The tube is mixed and incubated on ice for 15 min. The test tube is centrifuged to separate the aqueous, interphase and organic phase layers. The aqueous phase is transferred to a clean tube and isopropanol is added to precipitate the RNA. The RNA consists of total RNA, i.e., mRNA, rRNA and tRNA. An additional purification step may be performed to isolate mRNA from rRNA and tRNA.

### **Isolation of mRNA**

Rather than isolate total RNA, some procedures are designed to isolate only mRNA. Since mRNA species are polyadenylated, (i.e., have a series of adenine nucleotide residues (a poly(A) tail) at the 3' end), one can employ a method that uses series of deoxythymidine residues to isolate the poly(A) tail via complementary base hydrogen bonding.

#### **Oligo(dT) Method**

Oligo(dT) purification of mRNA is one method to obtain high quality, undegraded mRNA for gene cloning. Oligo(dT) column chromatography is used to reliably enrich and purify polyadenylated mRNA from contaminating, non-polyadenylated rRNA and tRNA. Oligo(dT) cellulose is a chromatographic matrix where many chains of deoxythymidine (dT) residues are covalently bonded to a cellulose support. The oligo(dT) matrix is packed in small chromatographic columns, and total RNA is passed through the column. The poly(A) tails of mRNA hydrogen bond to the oligo(dT) residues in a process known as hybridization. Non-adenylated RNA species (i.e. rRNA and tRNA) do not bond to the oligo(dT) matrix, and elute off the column. The column is extensively washed to remove any trace contaminants and then mRNA is then selectively released from the oligo(dT) matrix.

mRNA hybridizes to oligo(dT) in a buffer with high-ionic strength, i.e. high LiCl salt content. The positive lithium ion acts as a counterion and electrostatically pairs up with each negative phosphate backbone of mRNA and the oligo(dT) matrix. Since the electrostatic repulsion between phosphate backbones of mRNA and the oligo(dT) is neutralized by lithium, the nucleotide bases can hybridize by hydrogen bonding. mRNA is selectively eluted from the column by a buffer of low ionic strength, i.e. low salt content. Under conditions of low ionic strength, there are few positive lithium counterions, thereby permitting the electrostatic repulsion of the phosphate backbones to overpower the hydrogen base pairing between the poly(A) mRNA tail and the oligo(dT); this results in the release of purified mRNA.

#### **Magnetic Oligo(dT) Bead Purification Method**

High quality, undegraded mRNA is required for gene cloning. Magnetic oligo(dT) beads are used to reliably enrich and purify polyadenylated mRNA from contaminating rRNA and tRNA. Many chains of deoxythymidine (dT) residues are covalently bonded to the magnetic beads. Cell cultures are grown under conditions that promote the greatest abundance of the mRNA of interest. The culture media is removed and a lysing solution is added to the cultures to break open the cells, releasing DNA, RNA and proteins. This solution, known as a lysate, is transferred to a test tube with the magnetic beads. The poly(A) tails of mRNA hydrogen bond to the oligo(dT) residues in a process known as hybridization. Non-adenylated RNA species (i.e. rRNA and tRNA), DNA and proteins do not interact with the oligo(dT) matrix. The tube is placed next to a magnet to attract and hold the beads at the side of the test tube.

The undesired solution is readily poured off without loss of any beads. The magnet is removed to release the beads from the side of the test tube. The beads are extensively washed to remove any trace contaminants. mRNA is then selectively released from the oligo(dT) beads with an elution buffer. The test tube is again placed next to a magnet to attract the beads, and the supernatant containing the mRNA is collected.

mRNA hybridizes to oligo(dT) in a buffer with high-ionic strength, i.e. high LiCl salt content. The positive lithium ion acts as a counterion and electrostatically pairs up with each negative phosphate backbone of mRNA and the oligo(dT) matrix. Since the electrostatic repulsion between phosphate backbones of mRNA and the oligo(dT) is neutralized by lithium, the nucleotide bases can hybridize by hydrogen bonding. mRNA is selectively eluted from the magnetic beads by a buffer of low ionic strength, i.e. low salt content. Under conditions of low ionic strength, there are few positive lithium counterions, thereby permitting the electrostatic repulsion of the phosphate backbones to overpower the hydrogen base pairing between the poly(A) mRNA tail and the oligo(dT); this results in the release of purified mRNA.

### **Quality Control Assessment of RNA**

The amount, purity and quality of the isolated mRNA (or total RNA) is determined through the use of two methods: spectrophotometry and agarose gel electrophoresis. Spectrophotometry aids in the quantification of the amount of isolated RNA; agarose gel electrophoresis determines the quality of the isolated RNA.

Isolation of intact, undegraded mRNA requires scrupulous attention to analytical cleanliness. Contamination of the preparation will result in rapid enzymatic degradation by ubiquitous ribonucleases (RNases). RNases exist everywhere (or so it seems) and are difficult to inactivate! All solutions, glassware and plasticware must be scrupulously cleaned by special treatment procedures to eliminate exogenous RNases.

### **Spectrophotometric Method**

The amount of mRNA is calculated from the ultraviolet absorbance of the sample at 260 nm. mRNA yields are calculated using the formula:  $\text{mRNA (ug/ml)} = \text{Abs@ 260} \times \text{dilution factor} \times 44.2$ .

mRNA purity is assessed by a ratio of the sample absorbances at 260 and 280 nm. Pure mRNA has a 260/280 ratio of 1.9-2.1. A ratio less than this suggests contamination with proteins or phenol (from the organic extraction solvent).

### **Agarose Gel Electrophoresis**

Agarose gel electrophoresis determines the purity and quality of mRNA. In agarose gel electrophoresis, formaldehyde is used to denature any contaminating ribonucleases. The mRNA sample is electrophoretically separated according to size as it migrates through the pores of a formaldehyde agarose gel. High quality mRNA

appears as a smeared band over a 5 - 20 kb range, reflecting the variety of different sizes of mRNA. One also observes intact ribosomal RNA bands (28s and 18s subunits) with a good sample. Degraded mRNA is no longer an intact polynucleotide chain, but consists of a substantially smaller nucleotide sequences that are sufficiently small to migrate to the bottom and out of the gel. The ribosomal RNA bands are not present in a degraded sample, since the 28s and 18s subunits are degraded into substantially smaller nucleotide sequences too.

## 2. Gene Cloning Strategies

### Choice of Gene Cloning Strategies

There are many different gene cloning strategies. The choice of a gene cloning strategy is another important consideration in the experimental design. mRNA is used as a starting point for gene cloning because active genes express mRNA transcripts. However, mRNA is difficult to work with owing to its high susceptibility to RNase - mediated degradation. It is much easier to reverse transcribe mRNA into cDNA for gene cloning. cDNA synthesis from mRNA requires the use of a primer, and often the gene cloning strategy determines the type of primer to use.

Presented are three different gene cloning strategies to isolate the full length gene sequence:

1. Traditional cDNA library construction (using cDNA directly)
2. RT-PCR amplification of all cDNAs, followed cDNA library construction
3. RACE-PCR amplification of the selected cDNA of the desired gene sequence

For the first choice, a cDNA library (or gene bank) is constructed. A cDNA library is a collection of recombinant double-stranded, complementary DNA molecules (ds cDNA) that are inserted into a host (such as bacteria or viruses) that enables the propagation of the recombinant molecules. Libraries are screened with probes to identify and select the DNA sequence of interest. Searching through a library for a particular cDNA is sometimes laborious, since there are hundreds or thousands of bacteria (or viruses) each harboring a different cloned DNA.

RT-PCR and RACE-PCR are cloning methods that take advantage of the polymerase chain reaction (PCR). PCR is a method to produce many copies of a selected ds cDNA. (As an analogy, think of PCR as photocopier - one paper can be exactly duplicated many times.) PCR requires primers for the amplification of the target cDNA. One must have some knowledge, or make an educated guess of the target gene sequence so that unique primers can be designed for the gene sequence of interest.

### Gene Cloning Strategy: Starting with cDNA Synthesis

Cloning of a gene is often done using mRNA as a starting point because active genes express mRNA transcripts. However, mRNA is difficult to work with owing to its high susceptibility to degradation by ubiquitous RNases. It is much easier to reverse

transcribe mRNA into cDNA (complementary or copy DNA) for gene cloning. The single stranded cDNA is converted to double stranded cDNA (ds cDNA) which is the starting material for any cloning strategy. The ds cDNA sample will contain copies of all the genes that were expressed at the time the mRNA was isolated from the cells.

The basic approach is to select primers that hybridize (or anneal) to the mRNA. The primers give the enzyme, reverse transcriptase, a starting point to use mRNA as a template for the synthesis of the first strand cDNA. The resulting cDNA-RNA hybrid is treated with the enzyme, RNase H, to selectively cleave the mRNA into short fragments. These RNA pieces remain annealed to the DNA and serve as primers for DNA polymerase I which uses the first strand cDNA as a template for the new, second DNA strand. The second strand is not one continuous chain but exists as hybridized pieces to the first strand. The enzyme, DNA ligase, is added to covalently join (ligate) the second strand pieces into one continuous length.

### **cDNA Synthesis prior to PCR - First Strand Synthesis**

mRNA must be converted to double-stranded cDNA prior to PCR amplification. The first step is to reverse transcribe mRNA into complementary DNA in a process described as reverse transcription (RT) for first-strand synthesis. This yields mRNA-cDNA hybrids. The next step is known as second strand synthesis, where a DNA strand complementary to the first strand is made. The result of these two steps are double stranded cDNA clones of the starting mRNA. The double stranded cDNA clones are then amplified by PCR.

First strand cDNA synthesis requires an enzyme (reverse transcriptase (RT)), free nucleotides, oligonucleotide primers and a template (mRNA). The added primers hybridize specifically to the 3' end of the template. The enzyme will catalyze the addition of complementary nucleotides to the 3' end of the primer through ligation of the 5' phosphate ester ( $\text{PO}_4$  group) of the free nucleotide to the 3' ribose hydroxyl (OH) group of the growing cDNA chain. The result is an mRNA-cDNA hybrid.

For second strand cDNA synthesis, RNase H is added to nick the hybridized mRNA. The hybridized mRNA pieces serve as primers for the enzyme, DNA polymerase that synthesizes cDNA pieces complementary to the first strand. DNA ligase is added to covalently bond (i.e., ligate) the cDNA fragments into a full length, second strand.

Critical to cDNA synthesis and subsequent PCR is the choice of primers. It is easy to design a primer to hybridize to the 3' poly(A) tail end of eukaryotic mRNA - the primer is typically oligo(dT) either by itself or ligated to a unique oligonucleotide sequence. However, when cloning a new gene, the coding sequence of the 5' end of mRNA is often unknown, and this sometimes precludes the design of an appropriate primer for PCR. In this case, one strategy will covalently bond (ligate) a known RNA oligonucleotide sequence, (called a primer-adapter), to the 5' phosphate end of mRNA using RNA ligase. This primer-adapter serves as the known 5' primer for PCR.

However, mature, full length mRNA is protected by a 5' "methyl cap" which first must be enzymatically cleaved off. This exposes the 5' phosphate group of the last mRNA nucleotide for enzymatic ligation with the primer-adaptor.

### **PCR**

The polymerase chain reaction (PCR) is an experimental method to synthesize specific DNA sequence(s) starting with a known DNA sequence. PCR involves a chain reaction synthesis - in each cycle, each synthesized cDNA strand serves as a template for further DNA synthesis in subsequent cycles, as long as the primers and reactants are available and not consumed. PCR requires a double-stranded cDNA template, two different, single stranded DNA primers, DNA polymerase and excess deoxyribonucleoside triphosphates (dNTPs). Single stranded primers, complementary to the 3' ends of the first and second cDNA strands, are added in many fold excess of the DNA to be amplified. PCR is initiated by addition of primers along with enzymes and cofactors.

### **PCR Cycles**

One PCR amplification cycle consists of three phases: Denaturation, annealing (or hybridization) and extension. During denaturation, the temperature is briefly increased (90-95 degrees C for 30 sec) to separate (or melt) the double stranded cDNA. During the annealing step, the temperature is briefly decreased (55-72 degrees C for 30 sec) to permit proper hybridization of each primer to its complementary strand of template cDNA. During extension, the temperature is raised to 70 -72 degrees, so that DNA polymerase adds dNTPs to the primer's 3' end.

The second phase annealing temperature is dependent on the guanine - cytosine (GC) base content and length of the primers. If the GC content is less than 50%, annealing is done at 55 degrees; if greater than 50%, annealing is done at 60 degrees. If the primer length is 17-19 oligonucleotides (mers), annealing is done at 55 degrees; 19-21 mers, 60 degrees; 21-24 mers, 65 degrees; greater than 24 mers, 72 degrees. The time for extension is dependent on the length of amplified, target DNA: Allow 1 minute per kilobase.

The efficiency of one amplification cycle depends on not only the sequence and length of template cDNA, but also on the sequence and percent complementarity of the primers. Temperatures for each phase should be carefully optimized, which is done by performing test PCR runs. The yield of amplified product depends on the number of amplification cycles, which is typically 20-40. The number of cycles depends on the efficiency of the amplification cycle and the amount of starting cDNA template. PCR cycles > 40 can reduce the accuracy of DNA polymerase to add the correct dNTPs, amplify inaccurate sequences and consume all the primers and dNTPs.

### **Theoretical Number of ds cDNA copies**

PCR is a chain reaction - in each cycle, each synthesized cDNA strand serves as a template for further DNA synthesis in subsequent cycles, as long as the primers and reactants are available and not consumed. In routine PCR, each strand of a ds cDNA sequence is amplified exponentially with each cycle. The final result of a PCR reaction is that by the end of  $n$  cycles, the reaction contains a theoretical maximum ds cDNA copies of 2 raised to the  $n$ th power ( $2^n$ ), where  $n$  is the number of cycles. This is also true for RT-PCR or RACE-PCR, where the number of RT-PCR, 3' or 5' RACE-PCR products increase exponentially too.

In RACE-PCR, however, two types of products are generated: partial length, ds cDNA copies and full length, complementary copies of the starting 5'-3' cDNA. The theoretical maximum number of partial length ds cDNA copies is 2 raised to the  $n$ th power ( $2^n$ ), where  $n$  is the number of cycles. However, the theoretical maximum number of full length complementary copies of the starting 5'-3' cDNA is one copy produced per cycle, resulting in a linear amplification. This means that the theoretical maximum copies of the starting, full length 5'-3' cDNA is  $n$ , where  $n$  is the number of cycles.

### **PCR Primer Design**

PCR requires cDNA templates; two different, single stranded, oligonucleotide DNA primers; DNA polymerase and excess deoxyribonucleoside triphosphates (dNTPs). The oligonucleotide primers hybridize to the 1<sup>st</sup> and 2<sup>nd</sup> cDNA strands. DNA polymerase will synthesize new cDNA strands from the primers in the direction of 5' to 3' of the growing strand.

Oligonucleotide primers are complementary to each 3' end of the first and second cDNA strands. Primers are usually custom synthesized by a vendor according to the sequence specified by the scientist. By convention, vendor orders for primer sequences are specified from 5' to 3'.

In general, there are several points to consider when designing primers for PCR. The length of both primers (sense and antisense) should be 18-24 base pairs (bp). Longer primers do not confer added specificity, and in fact, may hybridize with mismatching. The desired GC composition of the primers is 45-55%, so as to promote strong base pairing through the formation of 3 hydrogen bonds per G-C hybridization (versus 2 hydrogen bonds per A-T pairing). The primer composition should avoid self base pairing, resulting in hairpin loops, and avoid hybridization to another primer, resulting in primer-dimers.

These general guidelines for primer design for PCR also hold true for RT-PCR or RACE-PCR.

### **PCR Optimization**

PCR is a powerful, but not necessarily easy method. There are many PCR conditions to optimize. And importantly, one must pay careful attention for potential contamination, otherwise contaminating DNA will be amplified as well. The general guidelines for optimizing and conducting a PCR also hold true for RT-PCR and RACE-PCR

In addition to primer design and the quantity / quality of the template cDNA, many other factors can influence the PCR reaction, such as  $MgCl_2$  concentration, the DNA polymerase enzyme, time and temperature of each phase of the amplification cycle, and the number of amplification cycles.

### **RT-PCR**

Reverse transcription - polymerase chain reaction (RT-PCR) produces copies of all the genes that were expressed at the time the mRNA was isolated from the cells. RT-PCR involves the following steps:

1. All mRNA transcripts are reverse transcribed into cDNAs for gene cloning.
2. Single stranded cDNAs are converted to double stranded cDNAs (ds cDNAs).
3. The PCR method will make multiple copies (i.e. amplify) of all ds cDNA transcripts.

The experimental approach is to design oligonucleotide primers for the RT and PCR steps. First, a known primer-adaptor is ligated (covalently bound) to the demethylated, uncapped 5' end of each mRNA. (The primer-adaptor provides a 5' primer for PCR after the RT step.) To start the RT step, one adds a known 3' primer for hybridization to the mRNA templates. Usually this primer contains a poly-T sequence for hybridization to the poly(A) tail of mRNA. This 3' primer gives the enzyme, reverse transcriptase, a starting point to use each mRNA as a template for the synthesis of the first strand cDNA. Then a reverse transcriptase enzyme, nucleotides and buffers are added to the test tube for incubation at 42 degrees C for 60 min.

Following first strand synthesis, each mRNA transcript remains hybridized to its cDNA strand. To obtain the second cDNA strand, primers for second strand synthesis are generated by the enzyme, RNase H, which nicks each mRNA transcript into fragments. Another enzyme, DNA polymerase I, uses the mRNA primers to synthesize the second strand using all first cDNA strands as templates.

PCR is performed as described in the PCR section. The general guidelines for primer design for PCR also hold true for RT-PCR or RACE-PCR. The general guidelines for optimizing and performing a PCR also hold true for RT-PCR and RACE-PCR as well.

### **RACE-PCR**

Rapid Amplification of cDNA Ends - PCR (RACE - PCR), unlike RT-PCR, permits cloning of 5' or 3' ends of a specific, target cDNA, and not all cDNAs. Prior protein purification and sequencing permits the use of RACE-PCR to selectively amplify the cDNA sequence of the desired protein, unlike RT-PCR, which amplifies all cDNA sequences. The steps to perform RACE-PCR are:

1. All mRNA transcripts are reverse transcribed (RT) into cDNAs.
2. Then using primers specific for the target cDNA, separate PCR reactions are run to individually clone each end of the target cDNA.
3. The PCR products are sequenced, new primers are designed, and the entire target cDNA is cloned.

RACE-PCR requires a knowledge of the target protein's amino acid sequence, which will enable the design of primers specific for the target gene. The advantage of RACE-PCR over RT-PCR or PCR gene cloning strategies is that if one successfully copies the correct, full - length gene sequence, there won't need to construct and screen a cDNA library, thereby saving a lot of time and labor.

### **Primer design for first strand cDNA synthesis**

The experimental approach is to design specific oligonucleotide primers for the RT and PCR steps. First, a known primer-adaptor is ligated (covalently bound) to the demethylated, uncapped 5' end of each mRNA. (The primer-adaptor provides a 5' primer for RACE-PCR after the RT step.) To start the RT step, one adds a known 3' primer for hybridization to the mRNA templates. Usually this primer contains a poly-T sequence for hybridization to the poly(A) tail of mRNA. This 3' primer gives the enzyme, reverse transcriptase, a starting point to use each mRNA as a template for the synthesis of the first strand cDNA. Then a reverse transcriptase enzyme, nucleotides and buffers are added to the test tube for incubation at 42 degrees C for 60 min.

### **Second Strand cDNA synthesis**

Following first strand synthesis, each mRNA transcript remains hybridized to its cDNA strand. To obtain the second cDNA strand, primers for second strand synthesis are generated by the enzyme, RNase H, which nicks each mRNA transcript into fragments. Another enzyme, DNA polymerase I, uses the mRNA primers to synthesize the second strand from all first cDNA strands.

### **3' or 5' RACE-PCR**

RACE-PCR permits selective amplification of a target cDNA from the cDNA collection. Following ds cDNA synthesis, separate RACE-PCR reactions are run to individually clone each end of the target cDNA: 3' RACE-PCR for the 3' end and 5' RACE-PCR for the 5' end.

3' or 5' RACE-PCR requires a cDNA template, two different, oligonucleotide primers, DNA polymerase and excess deoxyribonucleoside triphosphates (dNTPs). One primer, the anchor primer, is designed to hybridize to either the 5' or 3' end of the first strand cDNA. This primer is easy to design because the sequence of each end (5' primer-adaptor or 3' RT-primer) is known. The second oligonucleotide primer for RACE-PCR is more difficult to design. This primer, called a gene specific primer

(GSP), is designed based on a knowledge the internal amino acid sequence of the target protein. The design of gene specific primer (GSP) should be based on the amino acid sequence of a unique internal peptide of the protein of interest. Selection of a non-unique peptide may lead to the PCR amplification of many cDNAs of undesirable proteins.

3' RACE-PCR captures the target protein cDNA sequence from the 3' end of the second cDNA strand. Oligonucleotide primers are designed to the anchor-primer at the 3' end of second strand and to the GSP of the first cDNA strand. 5' RACE-PCR will capture the target protein cDNA sequence information from the 5' end of the second cDNA strand, using oligonucleotide primers designed to the anchor-primer for the first cDNA strand and a GSP to the second cDNA strand.

### **Gene Specific Primers (GSP) Design**

The GSPs should be unique to the target protein and different for the 3' RACE-PCR and 5' RACE PCR. The GSP is designed from a knowledge of an internal amino acid sequence of the target protein.

The design of the GSP for 3' RACE-PCR is not difficult. The cDNA code for the primer is derived from a unique peptide sequence of the protein of interest.

The design of the GSP for 5'RACE-PCR is trickier than for the 3' RACE-PCR. The 5' RACE GSP primer must be complementary to the second strand, so the primer must be complementary to the mRNA sequence. In other words, the GSP is an antisense cDNA primer because its oligonucleotide coding sequence is complementary to the mRNA sequence, and as such, can hybridize to the mRNA. To design this GSP, one must first write the mRNA (or cDNA) code of the peptide sequence, then re-write that code into its complementary, base - paired code. This is because the mRNA codons for the peptide amino acid sequence are complementary to the first strand; that is, the mRNA code actually represents the code of the second strand. The GSP primer must be complementary to the second strand, so it must be complementary to the mRNA sequence.

### **Purification of RACE-PCR Products for Full Length Cloning**

After RACE-PCR, amplified cDNA products may not be suitable for cloning or sequencing unless the products are purified from all the full length, cDNAs (from your starting sample) and components of the reaction mix (primers, mineral oil, salts, nucleotides, and DNA polymerases). Purification and recovery of the desired RACE-PCR products varies from approximately 80% to >95%, depending on the purification method, DNA size, and amount.

### **Full Length Cloning of the Gene of Interest**

A knowledge of the cDNA sequence of the 3' and 5' ends permits primer design for PCR cloning of the full length cDNA target sequence.

DNA sequencing of each 5' and 3' end of the cDNA PCR products will permit the design of primers to PCR out the entire target gene DNA coding sequence. Purified 5' and 3' RACE PCR products are sequenced by an automated DNA sequencer machine. See DNA sequencing in Section 6.

### 3. cDNA Library Construction

#### cDNA Library Construction

One constructs a cDNA library for selection of the cDNA of interest from all other cDNAs. This requires ligation (that is, insertion) of the cDNA PCR products into a cloning vector. These recombinant cDNA-vector constructs are then introduced into bacteria. The bacteria propagate the cDNA-vector constructs during bacterial replication. The recombinant bacteria are grown on agar plates for screening and selection for the cDNA of interest.

#### Selection of a Cloning Vector

There are different cloning vectors from which to choose: plasmids, lambda phages and cosmids.

#### Plasmids

Plasmid vectors are circular, double stranded DNA molecules that self-replicate many identical plasmid copies. Plasmids are good vectors for cDNAs, and there are many different plasmid vectors available for a variety of cloning strategies. cDNAs with a size range of 100-10,000 base pair (bp) are readily ligated into a plasmid vector. However, there is a low efficiency of recombinant plasmid incorporation into bacteria.

All plasmid vectors have three features: a replicator (ori site), which is the signal to start DNA replication; a selectable marker, which is a gene that confers antibiotic resistance to bacteria containing the plasmid; and a cloning site where cDNA can be inserted without altering other plasmid properties. Often the cloning site is placed in the middle of the lac-z gene, which codes for functional beta-galactosidase, an enzyme that metabolizes X-galactose (X-gal). If the cDNA is successfully ligated into a plasmid that is subsequently taken up by bacteria, then the lac-z gene is disrupted and the bacteria will not have the ability to metabolize X-gal. Bacterial colonies harboring a vector that lacks the cDNA insert will appear blue since they can metabolize X-gal. (The blue color is due to the metabolite of X-gal.) Conversely, bacterial colonies harboring the recombinant vector will appear white since they are unable to metabolize X-gal, owing to the DNA insert causing disruption of the lac-Z gene.

There are different types of commercially available plasmid cloning vectors. Selection of a vector depends on its size, the number of times a plasmid can replicate to produce multiple copies, the cloning site containing specific restriction enzyme

sites for cDNA insertion, and the ability to select and/or screen for plasmids containing the desired cDNA insert.

If the goal is to clone a gene that ultimately results in a functional protein, one might decide to select an expression vector, that is, a plasmid that contains a promoter sequence to direct bacterial translation of the cDNA insert to mRNA for protein synthesis.

### **Lambda Phages**

Lambda phages are viruses (bacteriophages) that can infect bacteria and replicate new lambda phages. cDNA with a size range of 5,000 - 20,000 bp are ligated to phage DNA, which is then packaged into infectious phage head viral particles. Lambda phage vectors incorporate the cDNA and infect bacteria with much greater efficiencies than plasmid vectors. In addition, lambda phage libraries are easier to work with and preserve than plasmid libraries. However, there is a low efficiency (<10%) of packaging the lambda-cDNA construct into phage heads.

### **Cosmids**

Cosmids are cloning vectors that are a hybrid cross of a plasmid and lambda phage. Cosmids are used to clone very large cDNA or genomic DNA with a size range of 35,000 - 50,000 bp.

### **Ligation of the cDNA into a Cloning Vector**

Ligation reactions produce recombinant DNA by covalently bonding (i.e. ligating) the cDNA insert to the vector DNA. Restriction endonucleases are used to cut the circular plasmid vector. The cut plasmid vector is linearized for ligation with the cDNA through the enzymatic action of DNA ligase.

### **Purification of cDNA PCR Products for Ligation into a Cloning Vector**

Amplified cDNA products may not be suitable for cloning or sequencing unless the products are purified from the components of the reaction mix (primers, mineral oil, salts, nucleotides, and DNA polymerases). Purification and recovery of the desired PCR products varies from approximately 80% to >95%, depending on the purification method, DNA size, and amount.

There are at least three methods of PCR product purification: precipitation, silica spin column chromatography and agarose gel electrophoresis. In precipitation, the PCR products precipitate out of solution when a high amount of salt and isopropanol are added. The smaller molecules of the reaction mix will remain in solution. In silica spin column chromatography the PCR products bind to the silica chromatographic matrix. Then a wash buffer is added to the spin column, which is then centrifuged to elute off unwanted reactants. An elution buffer is added to the spin column, which is then centrifuged to elute and collect the desired PCR products. In agarose gel electrophoresis, aliquots of the PCR product sample are

electrophoresed and the band with the correct bp size is cut from the agarose gel. The PCR product contained in agarose gel fragment is then extracted from the gel. However, this method is not the simplest and quickest purification method.

### **Restriction Digestion and Restriction Enzymes**

Restriction digestion is the process of cutting DNA molecules into smaller pieces with enzymes called restriction endonucleases (known as restriction enzymes). These enzymes recognize and cut specific sequences in the DNA molecule (for example *GATATC*) wherever that sequence occurs.

Restriction endonucleases are enzymes that cut DNA within a small, specific, 4-8 base pair recognition sequence. These restriction enzymes cleave DNA to produce blunt-ended or staggered (sticky) ends. The sticky ends of two DNA fragments cut by a restriction enzyme can base pair and be ligated (covalently bonded) together with great efficiency.

When one restriction enzyme is used to cut both plasmid and target cDNA, the ligation reaction will produce the plasmid containing the cDNA insert, as well as re-circularized vector (lacking the cDNA insert), as well as vector-vector and cDNA - cDNA concatemers. To avoid forming these undesirable products, two restriction enzymes are used to each cut a different end of the cDNA and of the vector. Proper selection of two enzymes will also insure that the cDNA insert ligates into the vector in the proper orientation for mRNA synthesis.

One must also select restriction nucleases that specifically cleave the circular plasmid for linearization and insertion of a cDNA, and also cleave the PCR cDNAs at only the primer ends. Restriction enzymes that cut the cDNAs at internal sites will produce partial cDNA fragments that are not suitable for construction of a cDNA library for the isolation of a full-length gene.

### **Ligation of the cDNA into the Vector**

A successful ligation of the linearized plasmid vector with the PCR cDNA products requires an overnight incubation with DNA ligase enzyme. An aliquot of the recombinant plasmid is then used to transform bacteria.

### **Bacterial Transformation**

Construction of a cDNA library requires bacterial incorporation of recombinant plasmids, a process known as bacterial transformation. Only bacteria that are competent, that is, they have the ability to internalize foreign DNA, are used. There are three basic methods to alter the bacterial cell wall for bacterial uptake of recombinant vectors: use of cold calcium chloride, heat-shock (30 sec incubation at 42 degrees C) or electroporation (use of a brief electrical pulse). A once the bacterial wall is altered, the bacteria and recombinant vectors are incubated so that the bacteria will take up and

internalize the vector. Following incubation, the bacteria are spread out on an agar plate containing nutrients, antibiotics and X-gal. The plates are incubated overnight to promote the growth of bacterial colonies that are derived from a single bacterium. It is at this point that one has their cDNA library!

#### 4. cDNA Library Screening

##### **General Library Screening for Transformed Bacteria**

Screening for transformed bacteria (called transformants, i.e., bacteria with internalized recombinant vectors) is done by bacterial growth on agar plates containing an antibiotic and X-gal, a substrate that can be metabolized by transformed bacteria. Transformed bacteria are antibiotic resistant due to the antibiotic resistance gene present in the vector; these bacteria will form colonies. Non-transformed bacteria that do not have the plasmid (containing the antibiotic resistance gene) will be killed by the antibiotic. Bacteria with recombinant plasmids containing the cDNA insert will not express the gene to metabolize X-gal because the cDNA insert is ligated into the middle of the X-gal gene in the vector; these colonies are white. Bacteria with self-ligated plasmids lacking the cDNA insert will express the gene for X-gal metabolism because the X-gal gene sequence is not disrupted; these colonies appear blue.

##### **Stringent Library Screening for Transformed Bacteria Harboring the Gene of Interest**

General or nonspecific screening based on antibiotic resistance and blue / white colony appearance (via expression of lacZ gene that metabolizes X-gal) helps identify colonies harboring the recombinant plasmid. However, more stringent screening is performed to identify a transformant with the recombinant vector containing the gene of interest.

Screening a recombinant library is a common approach to isolating a recombinant cDNA clone encoding a particular gene. A recombinant library consists of a large number of clones, with each clone containing a different cDNA insert in a vector.

##### **Different Screening Approaches**

There are several methods to screen a library. Common approaches use hybridization of a nucleic acid probe to a known cDNA sequence or antibody recognition of the protein expressed by the transformed bacteria. However, to screen a library using antibodies requires special experimental methods to make antibodies.

For identification of the cDNA clone encoding the protein of interest, it is easier to use synthesized oligonucleotide probes based on the protein's partial amino acid sequences (previously obtained by protein purification, isolation and sequencing).

##### **Design of Oligonucleotide Probes for Screening**

When specifying oligonucleotide probes based on a unique amino acid sequence, all possible oligonucleotide sequences are specified, synthesized and used for hybridization screening. Probes are labeled with radioactive  $^{32}\text{P}$  so that those

probes that are specifically hybridized to target cDNA can be detected by radioactive emissions.

Several considerations factor into the selection of an optimal amino acid sequence for subsequent oligonucleotide probe design. The amino acid sequence should be unique to the target protein. Since many proteins share common amino acid sequences, selection of a non-unique sequence will lead to an enormous number of false positives. One can avoid common amino acid sequences by searching computer protein sequence databases. Another consideration is that oligonucleotide probes should be 17-20 bases long, as this is the minimum length that confers hybridization uniqueness. This length requires knowledge of a minimum of six amino acids in sequence from the protein. Lastly, the degeneracy of the genetic code makes it difficult to design one or two probes. Some amino acids are coded by as many as six different codons, necessitating the synthesis of oligonucleotide probes containing all possible codon permutations. Use of a large number of oligonucleotide probes will lead to a higher false positive positives. If possible, select unique amino acid sequences where a maximal number of amino acids are encoded by one or two codons.

### **The cDNA Library Hybridization Probe Screening Process**

This screening process requires a lot of work. Often times, probe hybridization screening may lead to many false positive or false negative clones. Hence, one selects many clones that hybridize to the probe, followed by testing of each clone to see if it does or does not has the gene of interest. If a clone does have the cDNA for gene of interest, the next step is to determine if the cDNA encodes the full length or partial length of the gene. This additional testing is accomplished by protein expression methods, cDNA sequencing, or both.

The major problem with the cDNA library approach is that many false positives are detected. The oligonucleotide probes will hybridize to clones that do not encode the desired sequence or to a sequence that is part of a different gene. Only further testing of each and every positive clone will determine if the clone contains the desired cDNA sequence corresponding to the gene encoding the target protein. As a result of the high number of false "hits", one may screen as many and one million bacterial clones!

The cDNA library consists of agar plates that nourish transformed bacteria containing the recombinant vector. Nitrocellulose filter papers are laid on top of the bacterial colonies to create a paper replica of the agar plate. Some bacteria adhere to the paper, the rest remain on the plate. The papers are treated to lyse the adherent bacteria and the exposed recombinant plasmids are covalently bound to the paper. Labeled oligonucleotide probes are added to the nitrocellulose filters for hybridization to the immobilized recombinant plasmids. Bacterial colonies harboring the desired cloned gene are selected based on blue/white color and probe hybridization to the filter paper.

### **Protein Expression Testing of Each Clone**

The hybridization screening approach may detect false positives: the oligonucleotide probes will hybridize to clones that do not encode the desired sequence or to a sequence that is part of a different gene. Additional testing of each and every clone will determine if the clone contains the desired cDNA sequence.

If a bacterial protein expression cloning vector was used to transform the bacteria, then one might screen those clones that were identified by probe hybridization. Bacterial expression cloning vectors are used to transform bacteria for in vivo cloning and expression of recombinant proteins in *E. Coli* bacteria. Transformed bacteria normally do not express the recombinant protein under normal growth conditions. However, addition of an inducing agent will trigger bacterial mRNA polymerases to synthesize high levels of mRNA that subsequently leads to recombinant protein synthesis. The newly synthesized protein is assayed to determine if it is the full length protein of interest.

Each clone is typically tested by (1) in vitro transcription of the recombinant plasmid DNA into mRNA; (2) in vitro translation of the mRNA into protein; followed by (3) identification of the protein. The first step in this protein synthesis test requires a plasmid DNA miniprep experiment - the small scale purification of recombinant plasmid DNA from an overnight culture of transformed bacteria. For a miniprep, the bacterial culture broth is centrifuged and the resulting bacterial pellet is resuspended in lysis media to release intracellular components. Recombinant plasmid DNA is purified from contaminating bacterial proteins by anion exchange chromatography. Agarose gel electrophoresis is performed to assess the quality of the isolated recombinant plasmids.

### **Cell-Free, In Vitro Transcription - Translation Protein Synthesis**

The rapid characterization of recombinant plasmid clones is done using cell-free, in vitro transcription - translation procedures for the synthesis of proteins from cloned genes. The next step is to use the purified recombinant plasmid for in vitro transcription to mRNA and translation to protein. The circular recombinant plasmid (containing the cDNA insert), has a T7 or SP6 promoter site for RNA polymerase binding. The recombinant plasmid is added to a tube containing the appropriate polymerase, nucleotides and buffer salts for successful mRNA synthesis. In addition, the tube contains a cellular lysate (from rabbit reticulocytes or wheat germ) that provides all the components necessary for protein synthesis: tRNA, rRNA, amino acids and other factors. The tubes are incubated for 60-90 minutes to permit synthesis of mRNA transcripts and subsequent protein synthesis from the RNA transcripts.

### **Storage of a Desired Clone and cDNA**

Once a clone is successfully identified, it is important to prepare an aliquot of the transformed bacterial colony and purified plasmid for long-term storage. Bacteria are stored in glycerol at -80 degrees C. A small amount of transformed bacteria harboring

the clone of interest are grown to produce much recombinant plasmid DNA, which is then purified from the bacteria and stored at -20 degrees C.

## 5. Development of a Stable Expression System for Recombinant Protein Production

The next step is to subclone the desired cloned gene into a mammalian expression vector for mammalian cell expression. Mammalian cells glycosylate proteins; bacteria, yeast and insect cell hosts do not. Some proteins must be glycosylated for biological activity. If this is the case, the cDNA must be cut from the bacterial recombinant plasmid and subcloned into a mammalian expression vector.

Subcloning requires isolating the cDNA (encoding the gene of interest) from recombinant plasmids obtained from a transformed bacterial colony. The cDNA is religated into a mammalian expression vector, which is then used to transfect mammalian cells.

The mammalian expression vector, a plasmid, must contain a strong mammalian promoter (usually CMV) for high-level protein expression and a neomycin-resistance gene for stable selection in mammalian cells. The steps in subcloning are:

- (1) restriction enzyme digest to cut and isolate the cloned cDNA from the bacterial recombinant plasmid;
- (2) restriction enzyme digest to cut the mammalian vector;
- (3) ligation of the cloned gene into the mammalian vector;
- (4) purification of the recombinant vector to assure high purity for transfection;
- (5) transfection of mammalian cells (usually Chinese Hamster Ovary (CHO) cultured cells using a calcium phosphate transfection method);
- (6) growth and selection of transfected cells using an antibiotic (usually G418 or neomycin).

At best, only one of 100 cells will be transfected with the recombinant plasmid. Addition of an antibiotic to the cell culture media will select and identify antibiotic-resistant cells harboring the recombinant vector. Cells lacking the vector will die. The antibiotic resistant cells may transiently express protein from the cloned DNA insert. Stable and constitutive protein expression of the cloned gene occurs in a small number of cells when the recombinant DNA is appropriately integrated into the cellular genome.

The level of constitutive, recombinant protein expression depends on the promoter, the site of vector integration into a chromosome, the copy number of the vector and the type of protein.

Following cloning, subcloning and stable expression of the gene encoding the protein of interest in mammalian cells, the next task is to grow large-scale cultures and devise a protein purification process (described in the next textbook!)

## 6. Miscellaneous

**DNA Sequencing**

DNA sequencing involves the synthesis of new cDNA strands from primers, the usual deoxynucleotide triphosphates (dNTPs), and four dye-labeled dideoxynucleotides (ddNTPs) that lack a 3' hydroxyl group. When a ddNTP is incorporated into the growing DNA strand, synthesis is terminated due to the absence of a hydroxyl group and inability to form a phosphodiester bond. The final reaction products are cDNA fragments of different lengths. Each fragment has the fluorescent dideoxynucleotide (ddNTP) at the 3' end. The reaction products can be electrophoresed on a polyacrylamide gel that separates the fragments by a difference of one bp. The fluorescent labels permit automated fluorescent detection of each fragment with the incorporated ddNTP.