Handouts: 18 Molecular Cloning; How to make a library. 18B = Introns & Splicing (You also need 17).

Reminder: Selected problems are indicated in **red bold**. We suggest doing these problems as soon as possible to consolidate your understanding of the material covered in each lecture. (However, you need to go over the relevant section of the notes before you try the problems.)
Important Reminder: These are NOT the only problems you should do -- you should do ALL of them. For a complete list of problems corresponding to each lecture, see the problem set info page.

I. Molecular (DNA) Cloning -- how do you isolate a gene and why bother?

   A. What is DNA cloning? Means making a chimeric (recombinant) plasmid or virus and growing it up. For an overview of how to make a chimeric plasmid, see Becker 20-27 [20-28] or handout 17A (upper right). For an overview of the whole process see Becker 20-26 (20-27) or handout 18.

 B. Why bother?

1. To get a lot of copies of a particular gene/DNA.

a. To look at gene structure. Allows you to examine structure of gene and/or its regulatory sequences. There are some big surprises from this, which are discussed below.

b. Why is cloning necessary nowadays if you have PCR? Can't do PCR if don't know flanking sequences & can't make primers -- have to clone the first time.

2. To get a lot of gene product (protein, not DNA). Set up conditions where gene will be transcribed and translated. This allows you to

a. Measure gene expression (how much product a gene makes) and its regulation. Some examples of this next term.

b. Produce useful amounts of a product that is otherwise unobtainable in large quantities. For example, protein hormones such as growth hormone or erythropoietin (EPO). See Sadava 16.17 (16.14) for info on a similar example (TPA).

3. For gene therapy -- to restore function to a defective cell. Want to add gene so it will remain and supply a missing product as needed.

   C. Why you need a " cloning vehicle" or "vector." (See "fragment vs. plasmid" on handout 16B.) Need to add your gene of interest to a plasmid or modified virus so your gene can be replicated (using the origin of replication of the plasmid or virus). If you use a virus, it is usually modified, so that the DNA you are cloning replaces some of the viral genes.

   D. Basics of how to make a recombinant plasmid. You put the 'vector' and the fragment you wish to clone together using sticky ends (See handout 17A, top right.)

   E. How to actually make a library -- a collection of chimeric plamids -- (Becker 20-26 [20-27] or Sadava 16.11 (16.7)
Numbers 1-6 match steps on handout 18.

1. Cut up all DNA. Cut up genomic DNA from say, humans, in pieces with some enzyme; cut cloning vehicle with same enzyme. (Generate matching sticky ends as on handout 17A.)

2. Make recombinant plasmids (or viruses.) Mix fragments and cloning vehicle (plasmid or virus), let sticky ends match up and add enzyme to ligate (Becker fig. 20-27b [20-28b]). This generates a collection of plasmids (or viruses) with inserts. Each plasmid = one "book." The entire collection of plasmids/books = "a library." Some plasmids may lack an insert, as indicated in middle case on handout. (How to screen them out? See below.)

3. Get recombinant (chimeric) plasmids or viruses into cells. Transform bacteria with recombinant plasmids (or infect with recombinant viruses). Ideally, each bacterium gets one recombinant plasmid or virus = gets one insert. In cases shown on handout, cell (a) has a plasmid with an insert, cell (b) has a plasmid with no insert, and cell (c) has no plasmid. You want to screen out all cases except (a).

4. Grow up the "library." Grow up the cells that received plasmids; eliminate those that got no plasmid (or no virus) -- case (c).

a. How? Set up conditions so that only cells with plasmids will grow -- step 4. Handout shows a Petri dish with 6 colonies = 6 clones of bacteria with plasmids. You know all these clones contain plasmids because only cells with plasmids will grow under these conditions.

b. What makes it a "library?" Each colony or plaque = clone = descendants of one cell with the same recombinant plasmid (or virus). Each cell (or virus) in a clone has a copy of the same 'book' in the library = carries same added sequence (same insert). The entire collection of clones with different inserts = collection of 'books' = a library.

5. Screen out colonies that got plasmids, but plasmids without inserts. (case b.) To understand how to do steps 4-5, you need to know more about the cloning vehicle & replica plating -- see E below.

6. Find the right clone. This is the only hard part. Often you want to find 1 out of more than 50,000 possibilities! Step 6 shows one way to find the right clone -- by replica plating; steps 7-9 show another way -- by colony hybridization. See II & III below for more details.

E. Details to know about plasmid/virus = cloning vehicle. See handout 17B (left half) or Becker fig 20-27a [20-28a] or Sadava 16.9 (16.5) -- What properties should the vector have? Should allow you to tell cases a, b, & c apart.

1. Why you want 1 recognition site for each restriction enzyme per plasmid. You want 1 insertion site on each plasmid (for each dif. enzyme). There may be many different sequences recognized by dif. restriction enzymes, but only one site for each.

2. Why sites for many dif. enzymes -- so you can cut up DNAs in different ways and insert any of the fragments in the same plasmid.

3. Why drug resistant -- need a selectable "marker" on the plasmid -- gene that confers growth only to cells that got the plasmid. Then cells without the added "selectable marker" (that is, without a plasmid -- case c) won't grow under some condition. For example, the "selectable marker" can be a gene that confers resistance to the antibiotic tetracycline. (Tet-R) This allows you to select **for** cells that got a plasmid and **against** cells that got no plasmid at all. Cells without plasmids will be sensitive to the drug and won't grow in the presence of antibiotic. Only cells that have received a plasmid with the Tet-R gene will be drug resistant and will grow in the presence of antibiotic.

4. Why usually double drug resistant or has second selectable marker (See Sadava 16.10 (16.6) or

handout 17B) -- so you can detect cells that received a plasmid **without** an insert. (case b) The selectable marker (= active gene) confers some property that is easily detectable, for example, growth in the presence of an antibiotic or formation of blue colonies.

 If there is an insert in the 'selectable marker', that gene will be inactivated, and the cells will not have the corresponding property -- they will not grow, or not turn blue, etc.

 Conversely, if there is no DNA fragment inserted in the plasmid, the selectable marker gene will continue to function, and the cells will have the appropriate property -- they will continue to grow (plus antibiotic) or turn blue, etc.

Examples of selectable markers:

 a. Drug resistance. In the plasmid shown on the handout (pBR322), the second selectable marker is the gene for ampicillin (Amp) resistance. If the plasmid has no insert, it confers Amp resistance. But if the plasmid has an insert, Amp resistance is lost. You want to identify the cells that got a plasmid, but are NOT Amp resistant.

 b. LacZ. Another common selectable marker is the lacZ gene. Normal cells take up and cleave an analog of lactose; the product turns the colonies blue. Cells with an inactivated lacZ gene do not split the analog and remain colorless. (See Becker fig. 20-27 [20-28] & accompanying text).

 5. How do you check for an inactive version of the selectable marker? How do you identify cells that **don't** grow because a selectable marker **isn't** working? You need a way to identify colonies that do NOT grow under some particular conditions (and still recover viable cells). Use replica plating to selective media. (Step 5 on handout 18.)

 a. How do you make replicas? Using sterile velvet or the equivalent you transfer bacteria from the original Petri dish to a new dish. (You gently press the velvet on the old dish, and then peel it off and press it down gently on a new dish. The new dish contains bacteria arranged in the same pattern as on the original. This process can be repeated to make several replicas on different dishes.) Click here for pictures of the apparatus

 b. What good are replicas? In new Petri dishes, can have conditions that select for (allow growth of) cells with different properties. (Add Amp, or Tet, or leave out tyr, etc.) Can test for ability to grow under some set of conditions or ability NOT to grow. In step 5, you want to detect cells that **don't** grow because they have an inactive version of the gene for Amp resistance. If replica of clone does NOT grow on selective media (with Amp) then that clone has a gene with an insert. If replica of clone DOES grow, that clone got a plasmid without an insert (and you don't want that clone). After you see what grows (or doesn't) with Amp present, you can go back to the original plate and recover the right clones -- the ones that had inserts. Click here for typical results of replica plating.

Step 5 on handout 18 is replica plating to a Petri dish with ampicillin; step 6 is replicating to a Petri dish with some selective medium -- for example, without some required growth factor. (Additional replicas are possible.) Both pictures are included to show how you identify colonies that are Amp-resistant (have no insert) or able to made a particular growth factor, respectively.

 6. If vehicle (often a virus) is used for gene therapy, there are other considerations. How big a piece you can insert, how immunogenetic vehicle is, how to target added gene to right cells, whether DNA will insert into genome or not, etc.

Try Problem 13-4, parts A-D. If you want more practice on cloning, try 13-7, parts A & B, and 13-9 A-C.

II. How to find the Right Clone -- the one with the insert you want

A. The set up   Suppose you make a DNA library -- you carry out the procedure outlined on handout 18 steps 1-5. You now have many clones -- many colonies of bacteria. Each colony or clone contains a plasmid with a different insert. (You have already eliminated colonies of bacteria that got no plasmid or got a plasmid without an insert, using the procedures discussed above.)

B. The question(s): How do you find the clone that has 'the gene of interest' -- the one you are looking for? If the gene is there, how do you find out if it is working? (You may want to be sure protein is made, or it may be sufficient just to show the DNA is present.)

An example: Let's suppose you made a human DNA library and are looking for the human gene that codes for the enzyme phenylalanine hydroxylase (PAH). This enzyme catalyzes the conversion of phenylalanine to tyrosine, and is defective in people with PKU (phenylketonuria).

C. What could you look for? You need to find a cell with some selectable property because it has (or is using) the gene of interest (human gene for PAH, in this example). What will that property be? You can check at any step

<p style="text-align:center;">DNA (genotype) → RNA → protein → job → phenotype.</p>

1. DNA (can be detected by hybridization) -- minimal requirements for detection -- DNA has to be there, but doesn't have to "do anything." Need not be transcribed, translated etc. You do need a probe (something to hybridize to DNA) to detect the DNA. See next section on colony hybridization.

2. RNA from DNA (detected by hybridization with probe as above). To make RNA requires correct transcription signals but detection of RNA has advantages of amplification -- cell can make many copies of RNA from the one and only copy of DNA. RNA is usually easier to find than DNA (more copies to hybridize to probe).

3. Protein (detected using antibodies) -- Detection relies on structure, not function. Requires proper processing of mRNA (not yet discussed) and correct translation signals. (But protein need not work properly to be detected by antibodies.)

4. Job (measure function of protein -- usually enzymatic activity). Have to supply substrate. Requires all of above (transcription, translation, etc.) plus proper folding and/or modification sometimes. Note function of a protein is sometimes hard to measure if protein is not an enzyme, or requires binding to membrane, etc. in order to work

5. Phenotype (usually measure growth under certain conditions). If you start with a drug-sensitive or PAH- bacterium (tyr-) and add a plasmid, is bacterium now drug resistant or tyr+ etc? Requires that protein be made and function well inside a cell, not just in a test tube. Added DNA must supply a new function or replace (complement) a missing or defective function. Note this one is "selectable" in terms of growth/no growth in many cases. (Replica plating on different selective media can be used to find out if clones are drug resistant, tyr+ etc. Step 6 on handout 18.)

 Important: properties near the top of list require the least in terms of function, but properties near bottom are easiest to measure and/or most significant.

III. Colony Hybridization

A. What is the point of this procedure? You have many colonies, and you want to find all the ones carrying a particular fragment of DNA without testing the colonies one at a time. This method uses hybridization to a probe to detect cells (as vs bands on a gel) with the 'right' RNA or DNA.

B. How you use a Probe to find the 'right clone or colony' -- steps 7-9 on Handout 18 or Becker fig. 20-29 [20-30].

Step 7. Transfer bacterial colonies to nylon or nitrocellulose filter. (Same procedure as replica plating, but colonies are transferred to a thin sheet of nylon etc., instead of to surface of a Petri dish.)

Step 8. Treat cells (colonies) to lyse bacteria, release and denature DNA. DNA remains attached to filter, but is now single stranded and available to hybridize to probe. Note you did not have to purify DNA or isolate it from individual colonies. DNA was released from each colony & denatured *in situ* (in place).

Note: To lyse bacteria means to break them open (by any means). In this case lysis is caused by experimental manipulations, not by a virus.

Step 9. Add labeled probe in solution; let hybridize, wash off unattached probe. (If probe is complementary to the immobilized DNA on the filter, the labeled probe will hybridize and be trapped on the filter. If probe is not complementary to DNA, it will not hybridize, and it will be washed off.)

Picture at end -- shows spot where probe hybridized. You detect location of probe by its radioactivity or other label. Now you can go back to the Petri dish shown at end of step 4, and isolate the colony containing the nucleic acid of interest.

==C. How do you get a probe that is complementary to a particular gene?==

1. cDNA (works well for abundant proteins). Use reverse transcriptase to make DNA copy of mRNA. (Use additional enzymes if you want ds cDNA. See Becker fig. 20-30 [20-31].) DNA copy of mRNA is called "cDNA." (cDNA = complementary DNA or copy DNA.) cDNA is commonly made and used as a probe for a highly expressed genes -- ones that make a lot of protein and therefore a lot of mRNA.

> Question 1: Why use cDNA instead of mRNA itself? DNA is more resistant to degradation and easier to work with than RNA. See Sadava 16.12 (16.8).

> Question 2: If you have the cDNA, why do you need the gene itself? The gene or genomic DNA is usually **not** the same as the cDNA, because of introns, as explained below.

2. Oligonucleotide probes (good if know amino acid sequence) -- Sequences up to 50-100 nucleotides long can be synthesized by chemical means. An oligonucleotide of 15-20 nucleotides is usually sufficient to act as a probe. (That is, the sequence hybridizes to the gene you want and not to other genes.) You actually need a mixture of oligonucleotides -- you need more than one because the code is degenerate. Can't predict exact DNA/mRNA sequence from amino acid sequence. For example, if amino acids are lys - asp - met etc, DNA would be AAG/A CAC/T ATG etc. Don't know if base #3 is G or A; # 6 is C or T etc. So you make a mixture of oligonucleotides -- some have G at position 3, and some have A etc. (During chemical synthesis use a 50:50 mixture of G and A at position #3.) You'll get a mix of oligonucleotides, and one of the combinations will be the right one to hybridize to your gene.

3. Brute force/chance -- cut up DNA and get random collection of pieces; then use some procedure to figure out which is right piece. This sounds insane, but has been used very successfully to locate some human disease genes, as will be explained later after we cover the necessary genetics.

IV. What do you do with a cloned gene, once you have identified it? Some possibilities.

==A. Isolate the DNA from the clone.== How do you recover the cloned DNA fragment once you have identified the right clone? Say the one for PAH?

1. Lyse cells to release DNA.

2. Separate plasmid DNA from chromosomal DNA. Easy since plasmid is small and circular, unlike chromosomal DNA.

3. Treat plasmid with restriction enzyme used to make recombinant plasmid in the first place. This should release the cloned fragment. Separate fragment from rest of plasmid by electrophoresis on gels (separates DNAs by size).

4. Use the DNA. Examine the DNA (see V) or insert it in a different vector, as in cases B & C below.

B. Use bacterial cells carrying the cloned gene as a factory to make the corresponding protein. See Sadava 16.16 & 16.17 (16.13 & 16.14) for details and an example.

1. You use an "expression vector" = a plasmid with all the right signals for prokaryotic transcription and translation. See Sadava 16.16 (16.13).

2. You have to clone a DNA copy of the mRNA, not the actual DNA (gene). This is because DNA has introns, as will be explained below.

C. Use clone for gene therapy.

1. You need a eukaryotic expression vector. In this case, you want the DNA to be "expressed" in eukaryotes, so you need a different type of expression vector -- one with eukaryotic promotors, etc.

2. Most cloning vectors (or vehicles) used for gene therapy are modified viruses, not plasmids. The viruses are modified versions of human viruses. Different viruses have different features -- Some are useful at targeting specific tissues, some at integrating into host DNA, some at not provoking much of an immune response, etc.

3. Regulation. In some case the added gene is designed to have a switch -- transcription of the gene and production of the protein is turned on by addition of a drug. (This is similar in principle to having an inducible operon, but the mechanism is somewhat different, because eukaryotic and prokaryotic gene regulation work somewhat differently.

V. Eukaryotic Gene Structure -- What do you see if you clone a gene? (Handout 18B)

A. An example of how all the cloning procedures are used. What do you expect if you isolate the gene for the beta chain of hemoglobin and compare the gene to the corresponding mRNA (or cDNA)? How do you do it? See handout 18B for what is expected, and what is found. (Also shown in Becker fig. 21-20 [21-21] or Sadava 14.7 (14.6)

1. Make a library of plasmids with human DNA inserts. Sadava 16.11 (16.7)

2. Identify clone with beta globin chain gene. Use mRNA or cDNA as probe to identify clone with beta globin chain gene (grown up on plasmid). Using cDNA is a common way to get a probe for a highly expressed gene -- one that makes a lot of protein.

3. Isolate DNA from your clone. Lyse cells to release DNA. Separate plasmid DNA from chromosomal DNA, and digest plasmid with restriction enzyme used to make recombinant plasmid in the first place. Separate insert from rest of plasmid by gel electrophoresis.

4. Compare mRNA and genomic DNA. Genomic DNA is longer than mRNA. So you assume DNA includes sequences on the ends of the gene that are not found in the mRNA. (Extra = spacers? regulatory regions?). If you make a hybrid with mRNA and the genomic DNA, what should the hybrid look like? You can distinguish double stranded DNA (or DNA/

RNA hybrids) from single stranded DNA in the electron microscope -- double stranded regions are straighter and thicker. (See handout or texts for expected picture -- one single stranded loop.)

5. What you actually get. You get a structure with two R-loops as shown on handout 18B or Becker fig. 21-20 [21-21] or Sadava 14.7 (14.6) **R** loops = DNA loops formed because of **R**NA binding = single stranded loops of DNA formed when mRNA binds to template strand of DNA and displaces the sense strand of DNA.

6. What does picture imply? The DNA has 'extra' stretches in the middle of genes that don't show up in the mRNA. See Becker fig. 21-22 [21-23] or Sadava 14.5 (14.4).

- These extra stretches are called introns or intervening sequences.

- Most eukaryotic genes have introns. Prokaryotic genes do not (with a few very rare exceptions which we will ignore).
- An example: you can have a gene with sections A-B-C that produces an mRNA containing sections corresponding to only A and C. (Each letter represents a stretch of nucleotides. See handout 18B.) Sections A and C are called 'exons' and section 'B' is called an intron.

B. RNA Splicing process -- Overview. See bottom panel of hand out 18B, or Sadava 14.11 (14.10) or Becker fig. 21-24.

1. What happens to the extra stretches (introns or intervening sequences)? Why are they missing from the mRNA? We now know the entire gene is transcribed, including the 'extra' stretches, and then the 'extra' stretches are removed from the RNA. The removal process is called 'splicing.' In the example above, the DNA (from A to C) is transcribed and then the section corresponding to B is spliced out.

2. Terminology: The "ex" in exon stands for 'region that is **ex**pressed' meaning the region that is needed in the mRNA to make protein. The 'ex' does not stand for 'region that is excised' (or 'region that is extra'.) 'Extra' region that is excised = intron. Region that ends up in mRNA (not necessarily translated -- includes UTR's) = exon.

3. Splicing occurs in several steps, all catalyzed by the spliceosome = RNP = ribonucleoprotein particle. The spliceosome, like a ribosome, is a complex structure containing both RNA and protein. Saba Valadkhan, who was a graduate student here at Columbia in Jim Manley's lab, and a TA for this course, won the **AAAS Young Scientist Award** in 2004 for figuring out that some of the RNA's in the spliceosome are ribozymes -- catalysts made of RNA. See texts for details of spliceosome structure, and for which part does what, if you are interested.

4. All splicing occurs in the nucleus. After splicing is finished, what happens? The sections (introns) that were spliced out are degraded, and the mRNA containing only exons goes to the cytoplasm.

C. Some implications of Splicing/Processing

1. Genetic engineering implications. There are no introns or spliceosomes (catalysts for removing introns) in prokaryotes. This means that genes containing introns cannot be properly "read" in bacteria. That's why cDNA is often cloned instead of the actual gene if you want to use the DNA to make a human protein in bacteria.

2. Why is splicing of mRNA so rare in prokaryotes? Messenger RNA processing of all kinds (splicing, capping etc.) is absent or minimal in prokaryotes. Rationale? There is no separate compartment (nucleus) to hold unprocessed stuff. In bacteria, transcription and translation occur in the same compartment, and translation starts before transcription ends. This would not work if mRNA had to be extensively modified before it could be translated. (Ribosomes would attach and move down the newly made mRNA before it could be spliced.) In eukaryotes, mRNA is processed entirely in the nucleus and then shipped out to the cytoplasm for translation after all modifications are finished.

3. Alternative Splicing

a. Significance: Alt. splicing allows production of multiple proteins from one gene. A gene with more than one intron can be spliced in more than one way, so it can code for more than one protein. Therefore the number of possible proteins in an organism greatly exceeds the number of genes. This helps explain why we are so much more complex than, say worms, although we don't have many more genes!

b. An example: Consider a gene with 2 introns, such as the one shown below. It could be spliced twice, removing both introns, to produce a message containing exons 1, 2 & 3. Alternatively it could be spliced once, joining the end of exon #1 to the beginning of exon #3. (In this case, there is really one large intron including exon 2 that is spliced out.) This would produce a message with exons 1 & 3. These two messages would code for related but different proteins. Alternative splicing of the same transcript is known to occur in different tissues or at different times. (Some examples will be discussed next term.) See also Becker fig. 21-25 (not in 6th ed).

c. Proteomics vs Genomics. Because of alternative splicing, "proteomics," or the study of the proteins encoded in the DNA, has emerged as an area different from "genomics," or the study of the DNA sequences alone.

D. A typical picture of a gene with introns and exons (for reference). The picture below shows a section of the sense strand of the DNA that includes a gene with 3 exons and 2 introns. (The picture on the handout has 2 exons and one intron.) Conventions:

- The picture on the handout shows double stranded DNA, but genes are often shown as in the picture below, with only the sense strand actually drawn in.
- Transcription would start at the 5' (left) end of exon 1 and go to the right.
- Important features of intron: Branch point, 5' splice site (also called the donor site) and 3' splice site (also called acceptor site). These are shown for the first intron only. (See also top of fig. 21-24 in Becker or 14.11 (14.10) in Sadava.)
- Also note that the region to the left of exon 1 is NOT an intron -- it is not part of the gene. It is part of a spacer in between this gene and the previous one.
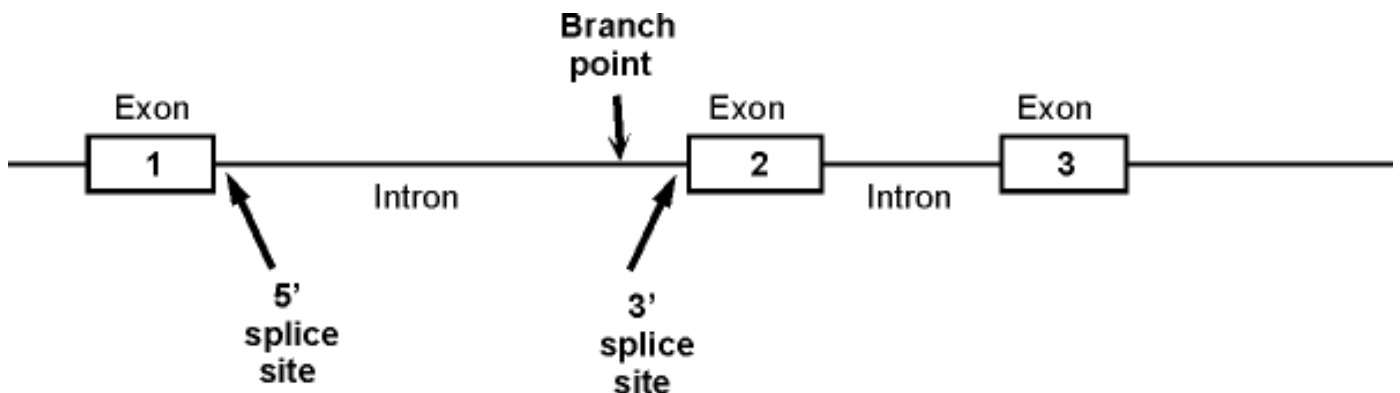
Now that you know more about introns and exons, Try problems 13-5 and 13-6, and finish problem 13-4 (E). . To review genetic engineering and introns, exons, etc. try 13-7 & 13-9 .

Next time: We'll wrap up RNA splicing. Then we'll do the mitotic chromosome cycle, and the actual steps of mitosis; then meiosis and some implications.