UN2005/UN2401 '18 -- Lecture 20 -- Last Edited: 11/26/18

Handouts for today: 20B -- Cloning Vehicle
20A -- Making a Library

You may want to refer to 18B. Paper copies of handouts 18B, 20A & 20B will provided in class. (For PDF's of 18B, & 20A/B, see the handout list. Paper copies of all handouts are available after the PM lecture in the boxes on the 7th floor of Mudd.

I. Probes (A & B were discussed last time.)

A. What is a probe for?

To find the right piece of DNA -- to find the piece from a particular part of the genome. If you cut up genomic DNA, you will have many pieces of DNA, and you want to find all the ones carrying a particular sequence. For example, to detect an RFLP (such as the one shown on handout 18B), you need to look at DNA from that particular part of the genome. But if you cut up the total DNA of an organism, you get many, many pieces. How will you find the right pieces, that is, the ones carrying a particular gene or section of the DNA where the RFLP is? (How will you know if your DNA is type 1, 2, or 3?)

B. What's a probe? It's a nucleic acid that's complementary to the target sequence you are looking for, and it is usually labeled or tagged in some way -- with radioactivity, fluorescence, or something else that's relatively easy to detect. Probe may consist of RNA, single stranded DNA or double stranded DNA. (Double stranded DNA must be denatured before it will hybridize to the target DNA.)

C. How do you get a probe that is complementary to a particular gene? This is explained in detail below (section VIII). The basic methods are as follows:

1. cDNA -- make the DNA complement of the mRNA using the enzyme reverse transcriptase.

2. Oligonucleotide probes -- if you know the protein encoded in the gene, you can use the code table to work backwards to design a probe (really a mixture of probes) that will hybridize to the gene.

3. Brute force/chance -- try out random fragments of DNA until you find one that hybridizes to the area of interest. This sounds hopeless, but has been used very successfully to locate some human disease genes, as will be explained later after we cover the necessary genetics.

D. Example of use of probes -- Detection of the RFLP on handout 18B -- see below.

II. Blots -- How to avoid having to slice up a gel and test the slices piece by piece.

A. What's a Blot? -- using nucleic acid probes or antibodies to detect macromolecules that are immobilized on a solid support. For mixtures of nucleic acids or proteins, the molecules are usually separated by size first, using a gel, and then transferred ('blotted') to a sheet of paper or plastic (the solid support). The positions of the molecules on the 'blot' (the solid support) are the same as in the gel. Then a probe or antibody is used to identify the position of the 'molecule of interest.' Blots allow you to test the properties of many samples at once.

B. Types of Blots (from gels) -- Terminology & Variations -- Southerns, Northerns and Westerns.

1. Southerns -- Cut up DNA, separate DNA fragments by agarose gel electrophoresis, blot (stick DNA to sheet of paper or plastic), denature DNA in place, find desired fragment by hybridization to probe. (Add solution of labeled probe, wash off unattached (unhybridized) probe and see what spots are still 'labeled' with radioactivity, fluorescence, etc.)

2. Northerns --Separate RNAs by gel electrophoresis. (RNAs are small enough as is; don't cut them up first.) Then blot, hybridize to probe as above.

3. Westerns -- Separate proteins by SDS gel electrophoresis, blot, find desired protein using labeled antibodies specific for that protein.

C. Blots from colonies -- *in situ* hybridization = Steps 7 to 9 on handout 20A -- see below.)

D. Details for Southern Blots -- detecting DNA Bands on gels How do you find a particular fragment of DNA (the target) if you start with a mixture? How do you find the fragment containing the gene or RFLP you are looking for? See Becker, fig. 21-5 (box 18C) or Sadava fig. 15.16 (9th ed).

1. Cut DNA up, or do PCR to amplify selected pieces. (If you cut the DNA with restriction enzymes ('cleave with restriction enzymes'), either cut up the same DNA with several different enzymes, or cut up several different samples with the same enzyme.)

2. Separate pieces on gel (usually agarose gel electrophoresis). Consider: Without probe, what would pattern of bands look like? How can you find the band you want without cutting up the whole gel into slices and testing each one? A "Southern Blot" allows you to do this.

3. Blot DNA from gel to paper or plastic (usually transfer to a sheet of paper or nitrocellulose.) DNA sticks to the support.

4. Denature DNA (in situ = in place) -- DNA is stuck to the support but will be able to hybridize to added probe

5. Add probe (often labeled cDNA or RNA). cDNA = complementary or copy DNA made using mRNA as template (& reverse transcriptase).

6A. Allow DNA and probe to hybridize in situ. Note: You can use stringent or nonstringent conditions, depending on whether you want a only a perfect match or are looking for approximate matches too.

6B. Wash off any unhybridized (labeled) probe.

7. Detect Bands – places were (labeled) is bound. Method depends on how probe was labeled – with radioactivity, fluorescence, an enzyme, etc. *For example:* Put blot in dark next to film if label is radioactive. Developed film is called an 'autoradiograph' = shows position of hybridized, labeled, probe.

8. Use position and intensity of bands (& number of bands) to characterize your DNA.
a. Calculate/measure size of targets (not size of probe) from position of bands.
b. Deduce relative number of pieces of each size from relative intensities of bands.

D. Example of use of Blots -- how to detect the RFLP's on handout 18B. Examples will illustrate the following features:

1. Need not purify the DNA of interest first -- you test a mix and locate where in the mix your DNA of interest is. You separate everything first by size (on gels), and then find the position of the piece you are interested in/want. Or you grow up multiple colonies containing different DNA sequences, and find the colony containing the DNA of interest. You don't have to know in advance where the sequence you want will be. Sounds like it won't work, but it does!

An analogy: Looking for a needle in a haystack. You spread out the hay in your haystack and glue to a support. Then cover it with magnetic particles and shake off the ones that don't stick. Where there are particles, your needle is underneath. When you do blots, you spread out your haystack (containing DNA samples) first, sort pieces by size and then add the 'probe.' Note that the final position of your probe (where it hybridizes or sticks) depends on the size or position of the target (the needle), not the size of the probe (the magnetic particles).

2. Can test many samples at once -- can use multiple wells and/or repeat hybridization to same blot using diff. probes.

3. Probe need not be same length as fragment

a. Probe can be shorter or longer than target sequence or fragment you are looking for. Probe and target need not be the same length, but there must be overlap, so some region of target hybridizes to probe. Sample must "capture" probe. (Think Velcro.)

b. The probe and target must have complementary sequences of 10-20 bases to form a stable hybrid. A restriction site alone is not long enough to form a stable hybrid to a probe.

Questions: (1) What if probe is complementary to a short part of the 5KB piece (from sample 3 on 18B.) How many bands will you get from each sample? How can you tell the cases apart?

(2). What if the probe is 4KB and overlaps the 5.5 and 5.0 sections of case 3? How many bands will you get from each sample? How can you tell the cases apart?

(3). Review from last time: What if the probe is the same length as the entire section? How many bands will you get from each sample? 0, 1, 2, or 3?

To review blots & probes, try problem 13-8, A-C.

III. 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 fig. 21-A2 (20-26) or handout 18A. For an overview of the whole process of molecular cloning, see Becker fig. 21-A1 (20-25) or handout 20A.

B. Why bother? Some possibilities:

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

a. To look at gene structure. Allows you to examine structure/sequence of gene and/or its regulatory sequences.

b. Why is cloning necessary nowadays if you have PCR? Can't do PCR if don't know flanking sequences & can't make primers -- may 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 do the following:

a. Measure gene expression (how much product a gene makes) and its regulation. Used in prokaryotes (as in analysis of operons) and eukaryotes (next term) to study regulation of genes.
b. Produce useful amounts of a product that is otherwise unobtainable in large quantities. For example, human protein hormones such as growth hormone (GH) or erythropoietin (EPO). See Sadava Investigating Life: Producing TPA (18.10) 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. See Sadava 15.18 (15.20).

a. Type of Vehicle: Most gene therapy uses modified viruses (not plasmids)*.
b. Targeting: The surface proteins of the modified virus attach to proteins on the surface of specific cells and therefore can target the added gene(s) to specific cell types.

*Note: Gene therapy involves adding genes to eukaryotic cells. In this semester we will stick mainly to how you add genes to bacteria, using plasmids.

C. Why you need a " cloning vehicle" or "vector." (See "fragment vs. plasmid" on handout 17A.) 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.

D. Basics of how to make a recombinant plasmid. (Recap of handout 18A).

1. Role of Sticky Ends. You put the 'vector' and the fragment you wish to clone together using sticky ends (See handout 18A.) This connection consists of weak H bonds only.

2. Role of Ligase. DNA Ligase is needed to join the vector and the fragment so they are covalently connected into a chimeric (recombinant) plasmid.

E. How to actually make a library -- a collection of chimeric plamids -- (Becker fig. 21-A1 (20-25) or Sadava 18.5 Numbers 1-6 match steps on handout 20A.

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 18A.)

2. Make recombinant plasmids (or viruses.) Mix fragments and many copies of the cloning vehicle (plasmid or virus), let sticky ends match up and add enzyme to ligate (Becker fig. 21-A2 (20-26b). 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. Three cases are shown on the handout. You want to screen out all cases except (a).
a. cell (a) has a plasmid with an insert
b. cell (b) has a plasmid with no insert
c. cell (c) has no plasmid.

4. Grow up the "library." Grow up the cells that received plasmids; eliminate those that got no plasmid (or no virus) -- case (c) and those that got a plasmid without an insert (case b). This takes two steps. (Details below).

a. 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.

b. How will you eliminate cases b & c?

(1). 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. (Using the right cloning vehicle makes this possible -- see F below.)

(2). Screen out colonies that got plasmids, but plasmids without inserts -- step 5. To understand how to do this, you need to know more about the cloning vehicle & replica plating -- see F below.

5. Find the right clone (after you have made a library) 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. Details of this part below.

F. Details to know about plasmid/virus = cloning vehicle. See handout 20B or Becker fig 21A-2 (20-26a) or Sadava 18.3. -- 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 plasmid is drug resistant

a. The principle: You need a selectable "marker" on the plasmid -- a 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. This allows you to select **for** cells that got a plasmid and **against** cells that got no plasmid at all.

b. An example: Suppose the "selectable marker" is a gene that confers resistance to the antibiotic tetracycline. (Tet-R) 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 plasmid is usually double drug resistant or has second selectable marker -- So you can detect cells that received a plasmid **without** an insert. (case b). See Sadava 18.3) or handout 19B.

a. What sort of 'marker' needed? The second 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.

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

c. What if there is *no* insert in the 'selectable marker'? 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.

d. Examples of selectable markers:

(1). 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 in the Amp-R gene, it confers Amp resistance. But if the plasmid has an insert in the Amp-R gene, Amp resistance is lost. You want to identify the cells that got a plasmid, but are NOT Amp resistant.

(2). 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. 21A-2 (20-26) & accompanying text).

e. How this works in our example -- Suppose plasmid used is one on handout, & you use Pst I to cut the plasmid. How will you tell which transformants got plasmids with inserts? What goes in the table? (If we don't do this in class, be sure to do it at home or in recitation.)

	Grows?	
Type of Transformant (see 20A)	**+ Amp**	**+Tet**
(a) -- has plasmid w/ insert		
(b) -- has plasmid w/o insert		
(c) -- has no plasmid		

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). Solution? Use replica plating to selective media. (Step 5 on handout 20A)

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 different 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. For a diagram of replica plating, go to https://en.wikipedia.org/wiki/Replica_plating#/media/File:Replica-dia-w.svg

for sample results of a real experiment, see fig. 2 of http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1221944/pdf/11415452.pdf

(Please notify Dr. M if any of these links don't work.)

Step 5 on handout 20A is replica plating to a Petri dish with ampicillin; step 6 is replicating to a Petri dish with some selective medium. (Additional replicas are possible.) For an example of the use of selective medium, see below. Both pictures are included on 20A to show how you identify colonies that are Amp-resistant (have no insert) or able to make a particular compound required for growth (such as an amino acid, vitamin, etc.).

6. If vehicle (often a virus) is to be 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. (See below.)

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.

IV. 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 20A 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: DNA (genotype) → RNA → protein → job → phenotype.

Below is a list of properties that you might check. 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.

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 for how to test multiple colonies at once.

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 20A.)

V. 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 20A or Becker fig. 20-29 [8th ed].

Basic Idea: You immobilize the target/test DNA (denatured) on a solid sheet and add labeled probe to see where it sticks. Similar to a Southern blot of a gel.

Step 7. Transfer bacterial colonies to nylon or nitrocellulose filter. Same procedure as replica plating, to start. In this procedure, a replica of the colonies is transferred to a solid support (as in blotting) instead of to surface of a Petri dish (as in the usual replica plating.) Once the cells are replicated to a sheet of paper, or nitrocellulose, etc., the DNA from the cells is released, and then it is denatured and hybridized (as in blotting) It's like blotting since you are transferring multiple samples to a solid support, but colonies (not bands in a gel) are transferred to a thin sheet of nylon, and then their DNA is released and analyzed.

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.

VI. What do you do with a cloned Eukaryotic gene, once you have identified it? Some possibilities. (Most of VI will be covered elsewhere, but is summarized here for reference.)

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 or insert it in a different vector, as in cases B & C below. What you find out if you examine eukaryotic gene structure is the next section.

B. Use bacterial cells carrying the cloned gene as a factory to make the corresponding protein. See Sadava 18.9 (& TPA story) 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 18.9.

2. Orientation. Two fragments with equivalent sticky ends can be joined in two different ways -- see handout 18A. (Q can pair with X or Z; Y can pair with Z or X.) Therefore a fragment may be inserted into a vector in two possible orientations. This will affect which strand is transcribed from the promoter of the expression vector. Gene to be cloned must be inserted so that it will be transcribed from the proper strand, and resulting mRNA can be translated.

3. You may have to clone a DNA copy of the mRNA, not the actual DNA (gene). This is because eukaryotic DNA has introns (intervening sequences in the middle of genes). More on this below and next term.

C. Use clone for gene therapy.

1. Type of Vehicle: Most gene therapy uses modified viruses (not plasmids). 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.

2. Virus Modification: Viruses used for gene therapy are usually modified so that the DNA you are cloning replaces some of the viral genes. (Therefore the virus cannot reproduce and is less likely to cause harm.)

3. You need a eukaryotic expression vector. When you add a cloned gene to a eukaryotic cell, you want the added DNA to be "expressed" -- to produce a normal protein to replace the function of a defective eukaryotic protein. Therefore you need a eukaryotic expression vector -- one with eukaryotic promoters, etc.

4. 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.)

VII. Eukaryotic Gene Structure -- What do you see if you clone a eukaryotic gene?

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, and what will you find? (For the real result, see Sadava 14.6 (14.7)) .See Becker fig. 18-24 (21-19) for a similar picture from a different gene.) A simplification is described below.

1. Make a library of plasmids with human DNA inserts. Sadava 18.5 (18.6)

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. (More on cDNA below.)

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. (Why the same enzyme?) 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 (or cDNA) and the template strand of the genomic DNA, what should the hybrid look like? Expect mRNA and DNA will match up with extra DNA on the ends.

- Note (FYI): experiment was not actually done this way -- RNA was hybridized to ds DNA, (not ss DNA). In this case, the RNA displaced the 'sense' strand when it hybridized. For a real picture, see Sadava 14.6 (14.7). For a picture from a different gene, see Becker fig. 18-24 (21-19).

5. What you actually get.

- If you make an mRNA-template strand hybrid, from the beta globin gene, you get at least one loop of ss DNA in the middle of the gene that doesn't match the mRNA.

- To check your understanding, see diagram of gene shown below with 2 introns. If you make an mRNA-template strand hybrid from this gene, how many loops of ss DNA do you expect to find in the middle of the gene that don't match the mRNA?

- Note: (FYI) In the real experiments, the investigators got so-called 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. (See Becker or Sadava.)

6. What does the hybrid structure imply?

- The DNA has 'extra' stretches in the middle of genes that don't show up in the mRNA. See Becker fig. 18-24 (21-19) or Sadava 14.7 (14.8).

- 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). Prokaryotes do not have the splicing enzymes needed to remove introns.

- 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. Sections A and C are called 'exons' and section B is called an 'intron' or intervening sequence.

B. RNA Splicing process -- Overview. This will be covered in detail next term, as splicing occurs in eukaryotes, not in prokaryotes. If you want to know more now, see Sadava 14.9 (14.10) or Becker fig. 18-25 (21-20).

1. What happens to the extra stretches (introns or intervening sequences)? Why are they missing from the mRNA in eukaryotes?

- First, the entire gene is transcribed, including the 'extra' stretches. The transcript is called the 'primary transcript.'

- After transcription, the 'extra' stretches are removed from the primary transcript.

- The intron removal process is called 'splicing.'

- The sections (introns) that were spliced out are degraded, and the nucleosides are recycled.

- In the example above, the DNA of the entire gene (including A, B & C) is transcribed and then the middle section corresponding to B is spliced out.

2. Terminology: See C below (including diagram) to be sure you have all the terminology straight.

- 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'.) An 'extra' region that is excised = an intron.

- A region that ends up in mRNA = an exon.

- Note that exons are not necessarily translated -- Exons encode both the translated regions and the UTR's.

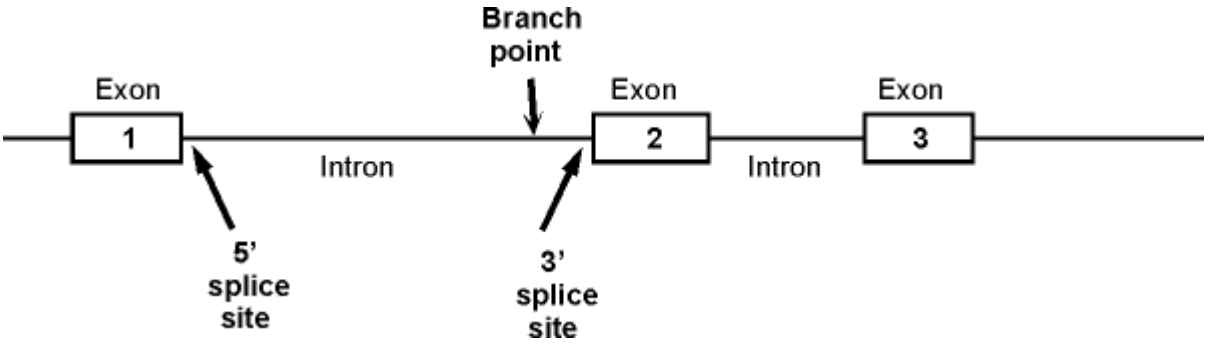
3. (FYI) Splicing involves a ribozyme. Splicing is catalyzed by the spliceosome = RNP = ribonucleoprotein particle. The spliceosome, like a ribosome, is a complex structure containing both RNA and protein. Saba Valadkhan, who was a graduate student here at Columbia in Jim Manley's lab, and a TA for this course, won the a 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. The mechanism of splicing will be covered next term.

4. Reminder: Only eukaryotes do RNA splicing.

- **Where does splicing occur?** Splicing happens in the nucleus. Then the mRNA containing only exons goes to the cytoplasm. (Reminder: prokaryotes do not have a separate nucleus and cytoplasm – they only have one compartment. Details of eukaryotic cell structure next term.)

- ***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, so transcription and translation occur in the same compartment (see Becker fig. 18-2), and translation starts before transcription ends. (See Becker fig. 18-3 (21-16). 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 or modified in any way.) In eukaryotes, mRNA is processed entirely in the nucleus and then shipped out to the cytoplasm for translation after all modifications are finished.

C. A typical picture of a gene with introns and exons . The picture below shows a section of the sense strand of the DNA that includes a gene with 3 exons and 2 introns. This picture is included to explain the standard conventions. Ignore the “branch point.”



- Cellular DNA is double stranded, but genes are often shown as in the picture above, with only the sense strand actually drawn in. When only one strand is given, it is almost always the sense strand.
- Transcription would start at the 5' (left) end of exon 1 and go to the right.
- 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.
- Terminology: The positions of splicing joints are specified from the intron's point of view -- that's why they are called the 5' splice site and the 3' splice site. These are shown for the first intron only. (See also top of fig. 18-27 (21-22) in Becker or Sadava fig. 14.9 (14.10) .)

Note: The Learner’s Manual uses the terms 'donor' or 'acceptor site'. These refer to the 5' and 3' ends of the intron respectively. There will be no exam questions this semester about the mechanism of splicing, and none with the terms 'donor' or 'acceptor' site. (The term ‘branch point’ also occurs in the manual. You can skip any questions or parts about the ‘branch point.’ We’ll go over it next term.)

- FYI: Every intron contains a branch point. When the intron is spliced out it forms a lariat -- the 5' end of the intron becomes attached to the branch point. The branch point of first intron (only) is shown in the diagram. (The mechanism of splicing will be covered next term.)

Now that you know more about introns and exons, Try problems 13-5 and 13-6, and finish problems 13-4 (E) & 13-7 (C). To review genetic engineering and introns, exons, etc. try 13-9.

The following information will probably not be covered in class, but is included for completeness.

D. Some implications of Splicing/Processing – This section is FYI.

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. Practical consequence: That's why cDNA (DNA copy of mRNA) is often cloned instead of the actual gene if you want to use the DNA to make a human protein in bacteria.

2. Alternative Splicing (FYI) – a Preview of Eukaryotic Issues – see next page.

a. Significance: Alt. splicing allows production of multiple proteins from one eukaryotic gene. A transcript from a gene with more than one intron can be spliced in more than one way, so the gene can code for more than one protein. Therefore the number of possible proteins in an eukaryotic organism (the proteome) can exceed the number of genes (the genome). This helps explain why we are so much more complex than, say worms, although we don't have many more genes! More on this next term.

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. 18-28 (21-23).

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.

Q: Which is larger, the 'genome' or the 'proteome?' Another way to ask this: What's the ratio of the # of different polypeptides/# different polypeptide coding sequences in the DNA? Is the ratio 1, <1, or >1? Is the answer the same for eukaryotes and prokaryotes?

VIII. Probes, Revisited.

This is largely for reference – it has all the backup details for I-C.

A. Why do you need a probe? To find the right piece of DNA. (See I-A & 1-B.)

B. How do you get a probe that is complementary to a particular gene? I-C Details:

1. cDNA

a. Works well for abundant proteins. See point c.

b. How it's done: Use reverse transcriptase to make DNA copy of mRNA.

(1). DNA copy of mRNA is called "cDNA." (cDNA = complementary DNA or copy DNA.)

(2). Use additional enzymes if you want ds cDNA. See Becker fig. 21-7 (20-28).

(3). What is Reverse transcriptase? It is an enzyme found in retro viruses (but not in normal cells) that uses RNA as a template to make the complementary, single stranded DNA. For the basics on retro viruses, click here.

c. 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.

d. Questions:

(1): Why use cDNA instead of mRNA itself? See Sadava 18.5B.

(2): If you have the cDNA, why do you need the gene itself? Isn't the cDNA the same as the gene?

2. Oligonucleotide probes

a. Good if know amino acid sequence of protein encoded by gene.

b. How it's done (if you know the AA sequence but not the DNA 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 GAC/T ATG etc. Don't know if base #3 is G or A; # 6 is C or T etc.
- 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.

c. If you know the DNA sequence (which is relatively common since we have sequenced many genomes) you can order a single oligonucleotide sequence to act as a probe. .

Q: In (b) and (c), what strand would you probably order to act as a probe? A copy of the template strand? The coding/sense strand? (Either one?)

3. Brute force/chance

a. Used when you don't know what the gene codes for but know something about its location.

b. How it's done: Cut up DNA from at or near the location you have identified, and get a random collection of cloned pieces. Do this for multiple individuals. Then correlate presence of gene (or mutation) of interest and some piece of the DNA.

c. It works! This sounds hopeless, but has been used very successfully to locate some human disease genes, as will be explained later after we cover the necessary genetics.

Next Time: Mitosis and normal karyotypes; then meiosis, life cycles and nondisjunction (how abnormal karyotypes occur). This section (eukaryotic cell division and genetics; last 3 lectures) will be taught by Dr. Michelle Attner. See (Revised) Lecture Schedule for topics and readings.