UN2005/UN2401 '18 Lecture #11

An outline of this lecture, listing the main topics, is posted separately.
For the remaining lectures by Dr. M, outlines will be posted in a separate folder, and there will be a link to each outline (as it is posted) from the schedule/reading list page. No PowerPoint slides will be posted -- Dr. M will be using chalk instead of PP. Dr. M will bring handouts needed for note taking to each class. (These handouts will also be posted as PDFs if you prefer to download them on your tablet -- we will try to have them ready the night before class.) The notes for each lecture will be posted by the evening before, and the outlines should be posted earlier. Each lecture will be recorded, and the audios will be posted after the evening lecture, either that night or the following morning.

Handouts: You will need handouts 10-1 to 10-4. They are the same as the handouts from lecture #10 of last year. For links to the handouts, go to the '18 handout list.

There will also be a handout (paper only) with some articles from (& links to) the *NY Times about DNA testing and DNA in popular culture*. Extra paper copies of all handouts will be in the boxes near Dr. M's office -- between 749 and 753C Mudd.

Abbreviations,etc: a.a. = amino acids. g.m. = genetic material. References to the texts are to Sadava et al 11th ed. & Becker 9th edition. In many cases, the figure and table numbers in the previous editions are the same. References to the 10th ed. of Purves/Sadava or 8th ed. of Becker are in parentheses **if** they are different. Note that the global edition of Becker has different pagination. (Subtract 24 from the recommended pages if you have the *global* edition.) For older editions of the texts, use the indexes, or see reading lists from earlier years.

I. Introduction. Why are we about to discuss structure & function of nucleic acids (esp. DNA)? We need to look at "the story so far."

A. The big question in this class is "How does 1 *E. coli* make 2?"

B. Where we are -- how far have we gotten toward the answer?

We've seen (most of) how the cell gets energy (catabolism) and how it uses that energy to make amino acids, sugars, etc. (anabolism). ~~This topic (metabolism) will be finished up next time.~~ So we can see how 1 cell growing on minimal medium could make enough (small) stuff → 2. We also see that reactions of metabolism (both catabolism & anabolism) all depend on having the right proteins -- you need enzymes to enable the reactions to proceed at a significant rate so cell can synthesize the small stuff and get energy .

C. So now what? What is the next question?

It's "how will we make enzymes (or, to be more general, proteins)?" Where will the information come from? This comes down to, "how will we get the info to hook amino acids together in the right order?" since primary structure determines the rest.

You can make complex lipids and polysaccharides by linking the monomers or parts together, using enzymes and energy the same way as you do when making small stuff. But how will you make proteins? You can't use enzymes to make enzymes (by using an enzyme to choose the next amino acid in line) since this is an infinite regression problem -- you would need an enzyme to make each

peptide bond, but each peptide has at least 50 (usually at least several 100) bonds, so it would take at least 50 enzymes to make one peptide (enzyme). But each of those 50 enzymes is a peptide and you would need 50 enzymes to make each of those, and so on.

This argument implies we need something to provide the information to line the amino acids up. Once we do that, a small number of enzymes, used over and over, can hook up the a.a. The substance that lines up the amino acids is usually called a template. If the template is a protein, who lines it up, and so on? -- that's the infinite regression problem referred to above.

D. Why template must be passed on -- why it must have 2 jobs, not just one.

We know also that when 1 (*E. coli*) → 2 → 4 etc., the descendants will make the same proteins as the original. Therefore, the template must be duplicated, and each daughter cell must get a copy. So it seems cells need a template = stuff with 2 jobs:

- *Job 1. determines a.a. sequence* = directs synthesis of proteins -- lines up the amino acids in the correct order.

- *Job 2. determines its own sequence* = directs synthesis of copies of self (so copy can be passed on and next generation can make proteins too.) Must be passed on faithfully but with occasional mistakes to allow for mutants and evolution.

E. Why is the template called "genetic material?"

- *1. Biochemistry Rationale.* Biochemists figured out there must be a template (made of something other than protein) with the 2 properties described above.

- *2. Genetics Rationale.* Geneticists had already reasoned that something they called "the genetic material" must exist to explain how traits of an organism are determined and how variations in the traits are inherited from generation to generation. The g.m. must be inherited and must control the construction/running of an organism.

- *3. Overall:* It seems clear the geneticist's g.m. and the biochemist's template must be the same thing. In other words: g.m. → proteins → job (often catalysis) → trait. Examples:

  g.m. → protein; protein is an enzyme

  What's the job? Enzyme catalyzes, say,
  (a) fructose → glucose, or
  (b) colorless material → colored material

  What's the resulting trait? Trait = visible or measurable property that results from having the enzyme. Examples:
  (a) ability to grow on fructose as carbon and energy source (in a), or
  (b) color -- or lack of it (in b).

F. What is the g.m. made of?

- *1. Historical Background:* There are only 2 serious candidates for g.m., DNA and protein, since chromosomes are made of these and there is genetic evidence they contain g.m.

- *2. Why not DNA?* DNA was not favored for g.m. (at time these experiments first done) because structure thought unequal to job. See data of Chargaff on Handout 10-4.

- *3. Why protein (or not)?* Protein was favored on general "it does it all, why not this?" grounds, but there's the infinite regression problem explained above..

II. Transformation: How did scientists figure out the chemical nature of the genetic material (that it's DNA)?

A. General Approach = Molecular Biology

Methods of biochemistry are insufficient to figure out what the g.m. is made of and/or how it works. So are the methods of genetics. However a combination approach works. The combo is known as molecular biology -- it uses the methods of both biochemistry and genetics to answer the questions of mutual interest to both fields. Molecular biology is defined by the methods it uses (both kinds -- biochemical & genetic) and the question(s) it is concerned with -- "How does the g.m. do its 2 jobs?". Next section of UN2005/UN2401 will be molecular biology; after that we'll do some formal genetics.

B. Basic Experiment = Transformation

1. What is Transformation? Test or assay for the genetic material. What chemical substance(s) -- DNA, protein, RNA, etc. -- have the ability to change properties/traits of an organism and have the change be inherited?

2. Historical Background: Why do you need to combine genetics and biochem approaches?

a. Why genetics? Need genetics to get a hereditary variant.

b. Why biochemistry? Need biochemistry to do chemical fractionation of cell extract to see what is the active 'transforming principle' that converts variant to normal (or *vice versa*)

3. Result: Only DNA transforms. (You are not responsible for the details of the actual experiment, but if you are curious, see Sadava, fig. 13-1 & 13-2 or Becker fig. 16-2 (18-2), or this excerpt from the notes of Lecture #10 of 2012.) {Q&A} *-- these questions are about the details of actual experiment, which involved transforming a PS- (unable to make polysaccharide, and avirulent) to PS+ (able to make polysaccharide and virulent).*

Note: Additional Questions and Answers can be found on Piazza, and in a searchable PDF at https://courseworks2.columbia.edu/courses/63615/files/3278194/download

4. Significance. So what?

a. Structure: Results of transformation led to closer look at DNA structure -- How does structure of DNA enable it to do its 2 jobs? Need to look carefully at structure to see how it could be g.m.

b. Practical Applications: The discovery of DNA transformation led (ultimately) to genetic engineering. Only a few organisms take up DNA "naturally" from surrounding medium and they may use this as way to exchange genes in nature. But virtually any organism can be

manipulated in a lab to pick up any DNA under appropriate experimental conditions, and this opens the way to genetic engineering, gene therapy, etc.

c. Significance of a universal 'operating system': DNA from one species will work in almost any other species. This implies that 'all living things work the same' in the sense that all organisms have a common method for using the info in DNA.

(1). Evolutionary implications: This implies common descent.

(2). Model Organisms: Since 'all living things work the same' at the molecular level, you can study virtually any organism to figure out how the basic processes work. Therefore researchers have focused on so called 'model organisms' such as *E. coli* or yeast or *Drosophila* (fruit flies) because they are small, fast growing, easy to manipulate, etc.

d. Use of this Approach: The original transformation experiment is a classic example of a molecular biology experiment -- combination of genetic and biochemical approaches needed to get the answer. We'll see many examples of the use and power of the mol. biol. approach (combination of biochem & genetics).

Next Section of the course: We'll look at DNA structure first; then Genetic Engineering later.

III. Nucleotides & DNA Primary structure -- Let's take a closer look.

A. Nucleotides

1. Why start with nucleotides? hydrolysis of proteins → aa; hydrolysis of DNA (or RNA) → nucleotides. Nucleotide = monomers of nucleic acids.

2. Structure of Nucleotides -- ***see handout 10-1*** , Sadava fig. 4-1 or Becker fig. 3-15.

- Note similarities and differences between nucleotides of RNA vs DNA.
- Some differences are necessary to distinguish the two; the cell has to be able to tell the irreplaceable master copy (DNA) from the replaceable working copy (RNA).
- Differences appear to be functional, not merely arbitrary: because of differences, DNA is less easily degraded and more easily repaired. (Why? Ribose is more reactive than deoxyribose, and nucleic acids with T are more repairable than ones with U. We will get to details of this later or see text.)

B. How are nucleotides hooked to each other? Primary structure of all nucleic acids (DNA or RNA) is the same. Nucleotides are linked using phosphodiester bonds from 3' of one sugar to 5' of next. See whole di-nucleotide (handout 10-2); clearly need shorthand way to write all this.

See problem 6-1, A-E for a review of nucleotide structure.

C. Short hand -- ***see handout 10-2*** and Becker fig. 3-16. {Q&A}

D. What does a long chain look like? ***See handout 10-3 on top right*** Sadava fig.4.2; or Becker fig. 3-17. {Q&A}. Important Features:

- Backbone (repeating sugar-phosphate.); bases stick out & phosphates are negatively charged.
- 2 ends of chain are not the same -- 5' vs 3'
  - Nucleotide on one end is not connected to another nucleotide by its 3' position. This nucleotide has a "free" 3' end. This end of chain is called the 3' end.
  - Similarly, nucleotide on other end has its 5' position "free," that is, unconnected to another nucleotide. This is the 5' end.
  - Doesn't matter if "free" position has a OH or phosphate; all that matters is that it is not linked to another nucleotide.
  - Since either end can have a phosphate or a OH at the end of the chain, the ends are called 5' and 3' ends, not phosphate and hydroxyl ends.
- 5' → 3' conventions: 5' always written on left.
- Note similarities to protein structure & conventions. Proteins also have
  - backbone structure with variable parts (R groups) that stick out
  - Unique ends so chain has direction -- Same end (amino) always written on left.
- Other imp. features: Very long, % C+G (base composition) different in dif. DNA's from dif. organisms (but same in all DNA from dif. cells of same org.)
  See handout 10-4 at bottom or Becker table 16-1 (18-1).

E. What does primary structure explain? How explains job #1 -- if chain is read 3 (or more) bases at a time, it could specify amino acid sequences of proteins. (Details later.) Need to read at least 3 bases at a time so there would be enough different combinations of bases to specify all 20 different amino acids. {Q&A}

IV. 3D DNA structure -- Why bother? ***Any of the remaining topics that are not covered today (Lecture #11) will be covered in lecture #12.*** Need it to figure out how job #2 is done. Analogy to protein 3D structure -- can't figure out how proteins/enzymes work unless you consider their 3D protein structure. Similar situation for DNA.

A. Starting info available to Watson & Crick: See Becker, fig. 16-6 (in 9th ed only: Becker Box 3A, pp. 60-61)

- Molecule = Very long, thin, rigid rod
- G=C, A=T (see data of Chargaff on handout 10-4) and/or Sadava ***section*** 13.2
  .
- Purines = pyrimidines = 50% = G + A = C + T
- Definitely >1 chain/molecule; Probably 2 chains

- Many weak bonds since heat denatures (unfolds -- causes loss of rigidity)

- Has regular repeating structure from X ray crystallography, which implies certain parameters (such as distance between repeats) about molecule, but interpretation of data debatable. See Sadava fig. 13.5 (13.6). W & C think it fits a helix. See Sadava fig. 13-6 (13-7) or Becker fig. 16-6 (not in 9th ed.)

- Symmetry. Turning molecule upside down doesn't change X-ray picture ==> 2 chains point in opposite directions (are anti-parallel) so molecule as whole (as opposed to a single chain) has no top or bottom.

B. Phosphates in or out?

1. Function implies phosphates in, bases (the part that specifies amino acid sequence) out.

2. Structure implies phosphates (ionized part) out, bases in.

C. Base Pairing -- How hold 2 chains together if phosphates out?

1. Why only purine-pyrimidine pairs allowed (to keep constant diameter)

Double stranded DNA must consist of 50% purines and 50% pyrimidines. There are two ways to see this.

a. Because G + A = C + T and G + C + A + T = 100 % **or**

b. Because each base pair must have 1 pur and 1 pyr (to give a molecule with a constant diameter)

2. Why only G-C and A-T. (To allow optimal # H bonds). See handout 10-3 (top left) or top of 10-4 for how H bonds go between pairs, Sadava ***section*** 4.1, or fig. 13.7 (13.8) or Becker fig. 3-18. Remember that only H's attached to O or N will form hydrogen bonds.

3. Tautomers. Why only G-C and A-T base pairing are allowed was not obvious to W & C at first -- why not? Wrong tautomers (enol forms) were printed in books -- see handout 10-4 for enol and keto structures of T and G. Tautomer = special type of isomer = keto vs enol form of C-O combo. If the O sticks out of a ring, is it C=O (keto form) or C-OH (enol form)? For diagrams, see handout 10-4.

To review base pairing, see problems 6-1 F & 6-2 to 6-3. If you want more practice, try 6-4 A&B and 6-5 A. (Which base will pair with I?)

D. Parallel or anti? The two chains of DNA could be parallel or anti-parallel as shown by pop bead model or as follows:

anti-parallel	**parallel**
5' AGC.....TTA3'	5' AGC.....TTA3'
3' TCG.....AAT5'	5' TCG.....AAT3'

Each chain has direction. Does the double helix? No. The two strands of DNA are actually anti-parallel, so each double-stranded molecule has two chains running in opposite directions. Therefore the two ends of the double stranded molecule are equivalent -- the overall structure or shape of the helix is the same if you turn it upside down. (Of course the sequence of base pairs is different if you invert the molecule.) This fits with the available experimental data, but has serious implications when you consider the fine points of DNA replication.

For Reference: Anti-parallel lineup of the 2 strands is shown on **Handout 10- 3 .** [Also shown in Becker fig. 3-18 or 16-7 (18-4), or Sadava fig. 4.4 or 13.19 (13.11)]. On the handout, chain on left of handout runs "down" 5' to 3'. Chain on right runs "up" 5' to 3'. The individual sugar-phosphate connections on the two chains are visibly different. The 1' on the sugar always connects to a base a little above the sugar. But look at the 5' position = the bend in the sugar-phosphate connection (on handout). The bend is just below the sugar on the right chain and just above the sugar on the left chain. Each chain has a direction but double helix doesn't.

E. What does the final structure look like? The famous double helix There are nice pictures on the Wikipedia page. Consult your text and/or handouts for additional pictures and details. Alternatively, go to Google, select images, and enter 'DNA'.) See Sadava, fig. 13.6 (13.7) or Becker fig. 3-19 or 16-7(a) [18-4(a)]. Why is this structure famous? See V.

V. So what does 3D structure explain? Implications/Consequences of 3D structure

W & C say at start of their famous paper: "This structure has novel features which are of considerable biological interest." What are they talking about? Why did they say they had discovered the secret of life? At the end of their paper they say "The genetic implications of this structure have not escaped our notice." In other words, they think it is obvious from the structure how DNA does job #2 -- replicates itself. The basic idea is two halves of the molecule (the two single strands) are complementary to each other, so if molecule comes apart (by breaking only weak bonds) into two strands, each strand can act as a template for the missing half. So 1 molecule (double stranded) can → two. (Assuming a supply of raw materials, enzymes , etc.)

'Does DNA really replicate as indicated? How can you tell?' The story continues next time after a few more details on DNA structure and the 'non-genetical' Implications of the structure.

See recitation problems #6 (to come) for additional problems on DNA structure. RP #7 will include problems on DNA replication.

Note: The double helix structure was figured out in 1953 by Watson & Crick. In 2003, 50 year later, there was a series of articles in the *NYTimes* commemorating the discovery of DNA structure. One of the articles, which explores some of the practical consequences of the discovery, is at http://www.nytimes.com/2003/02/25/science/a-revolution-at-50-how-the-arms-of-the-helixes-are-poised-to-serve.html . You can get all the articles in the series from the *NYTimes* archive.

For the latest and the greatest stories in the *NYTimes* involving DNA go to
http://www.nytimes.com/topic/subject/dna
For stories involving DNA evidence, go to http://www.nytimes.com/topic/subject/dna-evidence
Another interesting source of recent articles on DNA is http://www.livescience.com/topics/dna-genes

VI. Additional Details on DNA Structure (FYI only)

A. Base Stacking. Bases are flat and stacked on top of each other. For pictures, see Becker fig 3-19 (and/or Box 3A, 8th ed), Sadava fig. 13.6 (13.7), or white model – will be brought to class in lecture 11 or 12.

Base stacking is stabilized by Van der Waals forces and hydrophobic forces = weak interactions between flat nucleotide rings that are stacked on top of each other in the double helix. Helps hold helix together along its length, as opposed to H bonds between the bases, which hold it together across its width.

B. Super coiling.

DNA double helix is folded back on itself like a kinked telephone cord -- this is called super coiling. Corresponds to a tertiary structure (if double helix is secondary, in parallel to proteins). This means DNA must be ***un***folded extensively before it can be duplicated or read out to make proteins. We will ignore how this is done; if you are interested, some details of proteins required for folding and unfolding are in texts.

Some numbers to illustrate the need for supercoiling:

Length of *E coli* DNA = 3×10^6 base pairs (bp or BP); how long is that?

4 Angstroms (A) per bp, so length in A = 12×10^6 A

Length in meters = 10^{-10} m/A X 12×10^6 A = 12×10^{-4} m = 1.2 mm ≈ 1 pinhead

But 200,000 *E. coli* can fit on the head of a pin! Or about 600 *E. coli* lined up the long way can go across the pin. So there must be higher orders of folding to allow the DNA to fit inside *E. coli*. {Q&A}.

How DNA is folded up with proteins (histones) in eukaryotic chromosomes will be discussed next term.

VII. "Non genetical" implications of DNA structure -- Denaturation, Renaturation & Hybridization

A. Denaturation

1. Double Stranded DNA can be separated into its single strands

Since the 2 strands of DNA are held together entirely by weak bonds, you can take double stranded DNA apart into single strands by heating the DNA, or treating with chemicals that break weak bonds. This is called denaturation by analogy with proteins -- heating a protein or DNA unfolds the molecule and makes it lose its functional shape. (However, heat denaturation of DNA is reversible, as explained below, while heat denaturation of proteins is not reversible.)

2. T_m -- Melting Temperature

a. T_m -- What is it? If you heat a particular DNA up, it denatures abruptly at a specific temperature as shown in the picture below and Becker fig.

16-11 (18-8). The temperature at which the DNA denatures, or unzips or "melts" is called the T_m for melting temperature.

b. What determines T_m? The T_m is proportional to the #of weak bonds holding the 2 strands of the double helix together. In normal DNA, the T_m is proportional to the % G-C (= %G + %C) in the molecule. The more G-C the DNA has, the more H bonds there are to hold the two strands together and the more you have to heat the DNA up to denature it. This is because

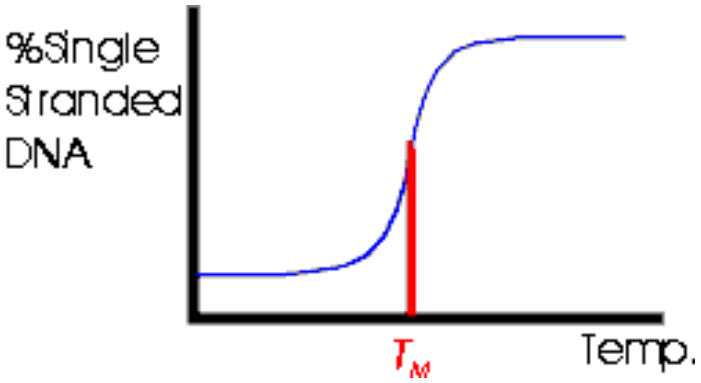
(1). Each G-C base pair is held together by 3 hydrogen bonds (not by a triple bond, but by 3 separate H bonds)

(2). Each A-T base pair is held together by only 2 H bonds (not by a double bond, but by 2 separate H bonds).

See handout 10-3, top left -- H bonds are drawn in for A-T and G-C pairs. Then look at handout 10-4, and be sure you know where the H bonds go (& why enol of T will not form H bonds to A). See also Recit. Prob. #6.

c. What use is the T_m? From T_m you can figure out the entire base composition (%A, G, C, & T) of a DNA. You heat up the DNA and measure its T_m. From the T_m (compared to that of a standard DNA) you get the % G + C in the DNA. (See Becker fig. 16-12 (18-9).) Once you have the % G + C, you can calculate the rest since G = C, A = T and G + C + A + T = 100%.

d. An example: If %G + C = 42%, What % is A? What is the entire base composition of the DNA? (What % is A, G, C, & T?)



To review denaturation, do Problem 6-4, parts A-C.

B. Renaturation -- if you cool denatured DNA down, it will renature or reform a double helix. (If you cool down a protein, will it renature -- will it refold to its original shape?) As you cool the single stranded denatured DNA, the bases in the complementary single strands match up again, and the single strands reform a double helix. See Becker fig. 16-13 (18-10). This procedure can be used to:

- Determine the concentrations of various DNA's in a mixture -- by the rate of renaturation of the different components in the mixture. Sequences at higher concentration will find partners and renature more quickly. (For an example, FYI, see Becker fig. 16-23 (8-15).) Therefore more common sequences pair up faster.

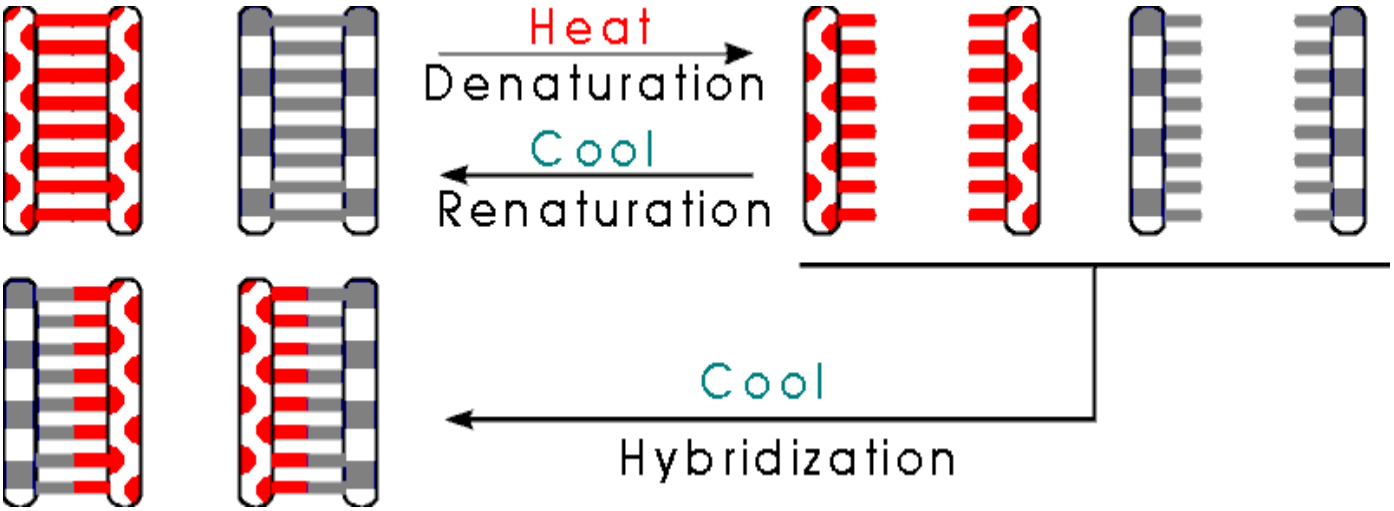
- See if single stranded DNAs from different sources will match up -- this is called hybridization; see below.

C. Hybridization -- Suppose you want to know if 2 samples of DNA (obtained separately) are the same. You can find out by denaturing each sample, mixing them, and cooling down the mixture. If the 2 DNA's are identical, then single strands (say "Watsons") from one DNA sample can match up with complementary single strands (say "Cricks") from the other sample, and you can form hybrid molecules. This is called hybridization since you match up complementary strands from 2 separate molecules to form a single 'hybrid' molecule. Variations on this procedure are used as follows:

- To test for similarities between 2 DNA's -- by the extent of base pairing in the hybrids. Hybrids can form between similar, but nonidentical DNAs under appropriate conditions. (Under other, more strict, conditions, only identical DNA's will hybridize successfully.) If the DNA's are similar enough, some, but not all, of the bases will match up and form base pairs. If a few base pairs do not match, but most do, the strands can still form enough H bonds to hold the hybrid together. (How will you measure the proportion of bases that are not paired up?)

- To detect specific DNA sequences -- by their ability to hybridize to test DNA's [called probes] that are radioactive or labeled in some other way. You denature the DNA you want to analyze (the unknown or 'DNA of interest'), and add the probe (usually single stranded). If the two DNA's form a hybrid (as detected by binding of probe) you known the "unknown" contained a sequence complementary to the probe.

Some examples of the uses of hybridization will be discussed in class as they come up.



Note: In picture above, there is only one starting DNA molecule of each type (red or black). In a real experiment, there would be many 'red' molecules and many 'black molecules'.

To review renaturation/hybridization, try problem 6-4, part D.

Next Time: *Wrap Up of anything above we don't get to. Then: How does DNA replicate?*