

C2005/F2401 '08 Lecture #10

© Copyright 2008 Deborah Mowshowitz and Lawrence Chasin Department of Biological Sciences Columbia University New York, NY. Last update 10/05/2008 12:11 PM

Handouts: 10-1: [Structure of Nucleotides](#), 10-2: [Shorthand](#), 10-3: [Base Pairing, Structure of Chain, etc.](#), 10-4: [Bases & Tautomers](#), and a handout with some articles from the *NY Times* (not on the web). Extra copies of all handouts will be available after each lecture in the boxes outside Dr. M's office -- near 744 Mudd.

Note: The [outline](#) for each lecture is posted separately; there is a link to each outline from the schedule/reading list page. No PowerPoint slides are posted -- Dr. M will be using chalk instead of PP.

a.a. = amino acids. g.m. = genetic material. References to the texts are to Sadava et al 8th ed. & Becker 7th edition. In many cases, the figure and table numbers in the previous editions are the same. References to the 7th ed. of Purves/Sadava or 6th ed. of Becker are in parentheses **if** they are different. 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? 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 how the cell gets energy and how it uses that energy to make amino acids, sugars, etc. so 1 cell can have enough (small) stuff → 2, and we also see that metabolism (both catabolism & anabolism) all depend on having the right proteins -- you need enzymes to make 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)?" This comes down to, "how will we 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. (See notes from last time.) 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 need an enzyme to make each peptide bond, but each peptide has at least 50 (usually at least several 100) bonds, so it takes at least 50 enzymes to make one peptide (enzyme). But each of those 50 enzymes is a peptide and you need 50 enzymes to make each of those and so on.

So we need something (besides an protein) to line the amino acids up and then 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 and in the end of the previous lecture.

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 descendents 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:

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

Biochemists figured out there must be a template (made of something other than protein) with the 2 properties described above. 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. It seems clear the geneticist's g.m. and the biochemist's template must be the same thing.

F. How did geneticists or biochemists figure out what the g.m. is made of and how it does its 2 jobs? (We all know it's DNA, but how did we arrive at that conclusion?)

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) and the question(s) it is concerned with ("How does the g.m. do its 2 jobs?"). Next section of C2005/F2401 will be molecular biology; after that we'll do some formal genetics.

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

A. Assay problem

How will you recognize the G.M. when you have it? Want to break up cell into parts and test (assay) parts. What will assay be? Assay discovered by accident as follows:

B. Set up

1. Bacterial Cell types

Cell type 1 → polysaccharide; Cell type 2 doesn't. Call them PS+ and PS-; PS+ = able to make polysaccharide. Easy to tell the difference by growth under certain conditions as follows:

Type	PS?	Capsule?	Colonies on Petri Dish	Virulent? (Cause disease?) = able to replicate in mouse?
1	PS+	Yes	Smooth	yes; can kill mouse
2	PS-	No*	Rough*	no

*no outer PS layer to form capsule and to make colonies smooth

Relationship of g.m., enzymes etc. Gene = section of the g.m. specifying a particular enzyme:



2. The phenomenon -- Transformation

Adding dead PS+ converts (transforms) descendants of PS- into PS+. See Sadava 11.1 or Becker fig. 18-2. Why must you actually be transferring g.m. ? Because property of "PS+ness" is inherited.

3. How use this to set up assay on fractionated extract.

Extract of PS+ = solution containing all water-soluble molecules from PS+ cells. Extract works just as well as whole dead cells to transform PS- to PS+.

Can fractionate (separate) extracts of cells into the various macromolecules -- DNA, RNA, protein, polysacch., etc. Then you can test each fraction to see if it can transform. (Sadava 11.2, 8th ed. only)

Note: 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. DNA was not favored for g.m. (at time these experiments first done) because structure thought unequal to job and protein favored on general "it does it all, why not this?" grounds, but there's the infinite regression problem.

4. Results & Significance; only DNA works (transforms). Works for many different traits, not just PS+ or PS-. So what?

(a). This 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). This (existence of 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 way to genetic engineering, gene therapy, etc. See Sadava 11.5 (not in 7th ed.)

(c). This is a classic example of a molecular biology experiment -- combination of genetic and biochemical approaches needed to get the answer. Genetic approach = assay; requires hereditary variant. Biochemical approach = chemical fractionation.

- Next Section of course: We'll look at DNA structure first; then Genetic Engineering later. [\(Q&A\)](#)

We'll see many examples of the mol. biol. approach (combination of biochem & genetics).

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 → nucleotides. Nucleotide = monomer of nucleic acids.

2. *Structure of Nucleotides* -- [see handout 10-1](#), Sadava 3.23 (3.24 in 7th, 3.16 in 6th), Becker fig. 3-15.

Note similarities and differences between nucleotides of RNA and DNA. Some differences are necessary to distinguish the two; the cell has to be able to tell the master copy (DNA) from the disposable working copy (RNA). Differences appear to be functional, not merely arbitrary: because of differences, DNA is less easily degraded and more easily repaired. (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 3.24 (3.25 in 7th, 3.17 in 6th); 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 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?

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 Box 3A, pp. 60-61(58-59))

- Molecule = Long thin rigid rod
- G=C, A=T (see data of Chargaff on handout 10-4)
- Purines = pyrimidines = 50%
- 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 11.6, (11.4). W & C think it fits a helix. See Sadava 11-8 (11-4).
- 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. 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.

- Because $G + A = C + T$ and $G + C + A + T = 100\%$ **or**
- 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 11.9 (11.7) 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) printed in books -- 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'

The two strands of DNA are actually anti-parallel, which is why 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, as indicated by the X ray data. (Of course the sequence of base pairs is different if you invert the molecule.) So each double stranded molecule has two chains running in opposite directions, and this makes for complications when we get to replication.

Anti-parallel lineup of the 2 strands is shown on [Handout 10-3](#) . [Also shown in Becker fig. 3-18 or 18-4, or Sadava 11.9 (11.7)]. 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. Fits with X ray data, but has serious implications when you consider the fine points of DNA replication.

E. What does the final structure look like? The famous double helix. [DNA Molecule - Two Views](#) (Consult your text and/or handouts for additional pictures and details. Alternatively, go to Google, select images, and enter 'DNA'.) See Sadava, fig. 3-26 (3-27 in 7th, 3-29 in 6th) & 11.8 (11.6) or Becker fig. 3-19 or 18-4(a).

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.)

Next Time: Does DNA really replicate as indicated? How can you tell? The story continues next time with a few more details on DNA structure and the ins and outs of DNA replication.

See recitation problems #6 (next week) for additional problems on DNA structure and replication.