Nucleic Acids: Properties, Structures, and Functions

http://courses.chem.psu.edu/chem572
January 11, 2016

Philip Bevilacqua

Lecture 1: Introduction to Biophysical Chemistry of Nucleic Acids

Reading: BCT Chpt 1

Biological Roles of Nucleic Acids

• As recently as 1944 the chemical basis for life was entirely unknown. At an organismal level, they knew that reproduction resulted in offspring that were very similar to their parents, but they had no idea what the chemical basis for this copying of information was.

QUESTION FOR DISCUSSION: If you didn’t know anything about the chemical basis for life, what properties might you expect these molecules to have?

• A number of landmark experiments helped elucidate the molecule responsible for reproduction. I give a few here.

Experiment 1, “DNA is the fundamental unit of transforming principle of Pneumococcus Type III” 1 Avery, MacLeod, McCarty (1944)

• A,M,M found that a nonpathogenic mutant 2 of Pneumococcus, referred to as the R form for rough colonies, could be transformed 3 into the pathogenic strain, referred to as the S form for smooth colonies, simply by mixing with heat-killed wild-type 2 Pneumococcus.

• This finding set the stage for the elucidation of the chemical nature of the transforming principle. 4

• They performed an elemental chemical analysis (%C, H, N, O, P) of the transforming material, and it agreed closely with the theoretical values for DNA.

• The material had the optical, ultracentrifugal, diffusive and electrophoretic properties of DNA. 5

• There was no loss of the transforming activity upon extraction of the protein and lipid components. This observation eliminated the other major cellular components.

• No loss of transforming activity upon treatment with proteases and ribonucleases (a.k.a.

---

1 Pneumococcus is the bacterium that causes pneumonia in humans.
2 ‘Mutant’ refers to a gene that has been altered from the normally occurring, or ‘wild-type’, gene. Wild-type is the non-mutated gene.
3 The term transformation is now widely used to refer to a process in which foreign DNA is introduced into an organism. For example, genes from any organisms (such as humans) can be introduced into a vector (a circular piece of DNA) and transformed into a bacterium such as Escherichia coli, which will express, or make many copies of that gene. Don’t confuse this with cloning of humans!
4 One of the goals of this class is to relate biological phenomena to molecular (i.e. chemical) explanations. Here is a great example of this!
5 We will spend a good deal of time investigating many of these properties throughout this class.
RNases). However, treatment with DNases resulted in complete loss of the transforming activity!

The following is an excerpt from a letter from Oswald Avery to his brother Roy, written in May 1943.

For the past two years, first with MacLeod and now with Dr. McCarty, I have been trying to find out what is the chemical nature of the substance in the bacterial extract which induces this specific change. The crude extract of Type III is full of capsular polysaccharide, C(somatic) carbohydrate, nucleoproteins, free nucleic acids of both the yeast and thymus type, lipids, and other cell constituents. Try to find in the complex mixtures the active principle! Try to isolate and chemically identify the particular substance that will by itself, when brought into contact with the R cell derived from Type II, cause it to elaborate Type III capsular polysaccharide and to acquire all the aristocratic distinctions of the same specific type of cells as that from which the extract was prepared! Some job, full of headaches and heartbreaks. But at last perhaps we have it.

... if we prove to be right—and of course that is a big if—then it means that both the chemical nature of the inducing stimulus is known and the chemical structure of the substance produced is also known, the former being thymus nucleic acid, the latter Type III polysaccharide, and both are thereafter replicated in the daughter cells and after innumerable transfers without further addition of the inducing agent and the same active and specific transforming substance can be recovered far in excess of the amount originally used to induce the reaction. Sounds like a virus—may be a gene. But with mechanisms I am not now concerned. One step at a time and the first step is what is the chemical nature of the transforming principle? Some one else can work out the rest. Of course the problem bristles with implications. It touches the biochemistry of the thymus type of nucleic acids which are known to constitute the major part of chromosomes but have been thought to be alike regardless of origin and species. It touches genetics, enzyme chemistry, cell metabolism and carbohydrate synthesis. But today it takes a lot of well documented evidence to convince anyone that the sodium salt of desoxyribose nucleic acid, protein free, could possibly be endowed with such biologically active and specific properties and that is the evidence we are now trying to get. It is lots of fun to blow bubbles but it is wiser to prick them yourself before someone else tries to.

Figure 4.4
Part of a letter from Oswald Avery to his brother Roy, written in May 1943. [From R. D. Hotchkiss. In Phage and the Origins of Molecular Biology, J. Cairns, G. S. Stent, and J. D. Watson, eds. (Cold Spring Harbor Laboratory, 1966), pp. 185–186.]

6 The suffix “ase” denotes an enzyme. These two enzymes degrade proteins and RNA.
7 This excerpt is from L. Stryer, Biochemistry 3rd ed.
Experiment 2. “Genetic Role for DNA” Alfred Hershey and Martha Chase (1952)\(^8\)

- The following is a scheme of a T2 bacteriophage injecting its DNA into *E. coli*. Bacteriophages are viruses that infect bacteria. Note that the virus itself does NOT enter the cell.\(^9\)

![Diagram of a T2 bacteriophage injecting its DNA into E. coli.](image)

- Phage DNA was labeled with \(^{32}\)P, the protein coat was labeled with \(^{35}\)S. These are unique labels for these molecules since there is no phosphorous in proteins and no sulfur in DNA.\(^{10}\)

- *E. coli* was separated from the phage in a blender and pelleted, with the phage staying in the supernatant.

- They found that the *E. coli* fraction became \(^{32}\)P labeled, meaning infected with labeled phage DNA.

- However, most of the \(^{35}\)S label (phage) protein remained in the supernatant.

- Led to the conclusion that “A physical separation of phage T2 into genetic and non-genetic parts is possible.”

---

\(^8\) This led to the Nobel Prize for Hershey in 1969.

\(^9\) This Figure is from Stryer Biochemistry, 3rd ed.

\(^{10}\) There are exceptions to this with modified nucleic acids and proteins, but generally this statement is true.
Experiment 3. “Chargaff’s Rule” Erwin Chargaff (1950)

• Chargaff found that the ratios of deoxy adenine to thymine, and of guanine to cytosine were nearly 1.0 in all species studied.
  
  i.e.: \( \frac{d(A)}{d(T)} = \frac{d(G)}{d(C)} = 1.0 \)

• The chemical basis for Chargaff’s rule would make sense a few years later upon Watson and Crick’s discovery of the structure of dsDNA (see next page of notes).

• Chargaff’s rule did not hold for RNA. Why?

---------------------------------------------------------------------------------------------------------

Brief Introduction to RNA

• RNA is typically composed of only one strand (i.e. it is not double-stranded like DNA). Of course portions of the RNA form double-stranded sections, but much of them do not. And even so, many of the double-stranded sections contain non-Watson-Crick base pairs.

• Major types and functions of RNA (first discovered 1950-1960): mRNA, rRNA, tRNA.\(^\text{11}\)

• Other roles for RNA (1970-present)
  1. Genetic information in viruses/retroviruses such as HIV-1
  2. Splicing and snRNPs (1977, Roberts and Sharp; 1993 Nobel Prize in Medicine)
    - Splicing is of growing interest to the RNA community. Completion of the human genome project (ca. 3 billion base pairs) led to the surprising finding that we have only ca. 35,000 genes. To achieve the diversity of function and regulation that makes us the complex beings we are epigenetics (genetic events beyond simple linear reading of the gene) are very important. Splicing can occur in different ways, i.e. exons can mix and match, which is referred to as alternative splicing, and greatly increases the genetic diversity of individuals.\(^\text{12}\)
  4. Development/translation/gene regulation and RNA interference (or RNAi). (Fire, 2001)

\(^\text{11}\) These refer to messenger RNA, ribosomal RNA, and transfer RNA. If you are unfamiliar with these, please refer to any introductory biochemistry text.

These processes are carried out by microRNAs (miRNA), ca. 22 nts in length. Named ‘Breakthrough of the Year’ by Science Magazine in December, 2002.

5. Riboswitches that bind ligands specifically and with high affinity to control genes (2003).


- The previous three experiments showed that there is a chemical basis for life, and that it must be possible to explain replication in molecular terms. The race was on!

- The structure of dsDNA has a number of important features. We will look at these in great detail further along in the course. For now, let’s just touch on some of the most important features of the most common form of DNA (called B-form DNA), which is what Watson and Crick discovered:
  - Right-handed helix
  - Major and minor grooves
  - Base pairing involving hydrogen bonding of G and C, and of A and T. Later came to be known as Watson-Crick base pairing. (see Figure below—from p32 of our text). You should commit these base pairs to memory. Know structures, numbering, and positions of grooves.

13 We will read and discuss their seminal 1953 paper.
• Was immediately ‘obvious’ how information could be transmitted from one generation to the next.

• Arguably, the single most significant scientific discover of all time. (Do you agree?)

**Nucleosides and Nucleotides**

• ATP—Critical for energy storage and conversion

![ATP structure](image)

• NAD (nicotinamide adenine dinucleotide). Important in oxoreduction reactions.

• cAMP/GMP—second messengers in variety of cellular processes.

**Some Major Structural Discoveries with Nucleic Acids in the Last ca. 15 years**

• Z-DNA (Alex Rich), or left handed DNA. The sequence CGCGCG is an example.

• Naturally Bent DNA. This is important in gene expression and protein binding. A track of adenines (A’s), referred to as an ‘A-tract’, gives rise to this.

• Detailed structures of DNA dodecamer (Dickerson). Watson and Crick’s structure was low resolution and was only a fiber diffraction study. The later structures were crystal structures and of much higher resolution. This led to the appreciation that the fine structure of DNA is highly variable, explaining how it can be recognized differently by proteins, ions and drugs.

• Importance of supercoiling to exert long-range control of DNA. Formation of histones and chromatin is important for packaging of DNA. The structure of heterochromatin during cell division appears to be controlled by tiny RNA molecules.


• Roles of ions in water in greatly influencing the structure and interactions of DNA.

• Determination of many non-Watson-Crick base pairings, and specific ribose-base-phosphate interactions present in folded RNA molecules
• Structures of the ribosomes (LSU and SSU), and with various antibiotics bound to them.

• Human genome sequenced (about year 2000), and many other genomes sequenced.

**Biophysical Method Development in the Last 15 years**

• Gel electrophoresis to analyze size and shape of nucleic acids and interactions with proteins. These methods allow one to estimate bend angles in nucleic acid structure.

• Ability to prepare synthetic nucleic acids, or isolate natural ones in adequate quantities for physical studies

• Multidimensional NMR to determine structure in solution

• Data on a wide variety of nucleic acids

• Combinatorial methods by in vitro selection (SELEX)

• Computer analysis methods that point out similarities between functionally related molecules

• Thermodynamic data for predicting secondary structure, loops, bulges...

**Obtaining Nucleic Acids for Physical Chemical Studies (see also course web page)**

Large DNA:

• Viruses (2-110 million MW)
• Calf Thymus
• Plasmids (E. coli) 2-6 kb
• PCR Cloning

Shorter DNA

• Needed for bending (few hundred bp), NMR (best on <30-40 nts), thermodynamic studies (8-12 nt)
• Sheering of long DNA (pipeting, vortexing; about 5 million MW)
• Vigorous Sheering (hypodermic needle) (about 1 million MW, or 1500 bp)
• Sonication for hours (50-100 bp distribution)
• Chromatin digestion 150bp
• Chemical synthesis (DNA to 100 nt/RNA to 30 nt commercially available via solid-phase synthesis).
• PCR (50bp-kb; Mullis, 1993 Nobel Prize in Chemistry)
**Methods for Obtaining RNA for Physical Studies**

1. Purchase from Sigma (e.g. tRNA, poly (A,U)...)  


   
   a.) Transcribe off of hemiduplex  
   b.) Transcribe off of duplex (PCR fragment)  
   c.) Transcribe off of a digested plasmid (runoff transcript).

4. Semisynthetic approaches. These bring together the best features of methods 2 and 3 and allow one to introduce single atom changes in very large RNAs that could not be synthesized completely.

Moore and Sharp methodology (1991)

![Diagram](attachment:diagram.png)

**Solving Biophysical Problems**

- To study a certain molecule, **ASK:**
  1. Is it available? Native? Pure? Homogeneous?
  2. Is it the right stuff? Complete? Will there be significant differences *in vivo* vs *in vitro*?  
  3. Can you get enough of it?

---

14 *in vivo* refers to experiment in living cells and *in vitro* refers to experiments in test tubes. *in silico* is a humorous term often used to refer to theoretical experiments using computers.
4. Are purifications consistent and reproducible?

<table>
<thead>
<tr>
<th>Technique</th>
<th>Number of moles needed(^{15})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Molecule Experiments</td>
<td>tiny amounts of material (10^{-23}) mol</td>
</tr>
<tr>
<td>Radioactivity</td>
<td>(10^{-12}) (pmol)</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>(10^{-9}) (nmol)</td>
</tr>
<tr>
<td>Absorption</td>
<td>(10^{-7}) (0.1 µmol)</td>
</tr>
<tr>
<td>NMR</td>
<td>(10^{-6}) (µmol)</td>
</tr>
<tr>
<td>X-Ray Diffraction</td>
<td>(10^{-5}) (10 µmol)</td>
</tr>
<tr>
<td>Neutron Diffraction</td>
<td>Lots</td>
</tr>
</tbody>
</table>

**Kinds of Questions to Ask\(^{16}\)**

1. Structure
   a.) What is it?
      1\(^{st}\) Structure-DNA sequencing (Human genome project)
      2\(^{nd}\) Structure
         i.) X-Ray (Best); NMR (Best)
         ii.) Sequence Comparison (Often used)
               --Compare with known structures (threading); used for proteins.
               --Phylogeny. Compare possible secondary structures with same
               function, different sequence; used for RNA
      3\(^{rd}\) Structure
         i.) X-Ray (Best); NMR (Best)
         ii.) Modeling (Future?)

   b.) Is the structure predictable?
      Easier for nucleic acids (more later in semester)

   c.) What are the dynamics of the structure?
      i.) Look at different conditions.
      ii.) Can look at a phase diagram for macromolecule.

   d.) Is there anything familiar or unusual about the structure.\(^{17}\) **Maintain a healthy level of skepticism at all times.**
      e.g.: It is usual to find a charged amino acid on the outside of a protein. If vice versa, WHY?

      e.g.: tRNA motifs showing up on viroids and in catalytic RNAs. WHY? (no translation in either case.) May be remnants of divergent evolution, or maybe tRNA is a common solution to a compact folding motif.

   e.) Do individual residues have altered properties? e.g. unusual \(pK_a\)’s. WHY?

   f.) How is the structure attained?
      i.) Number of RNA structures possible is \(\approx 1.8^N\), where \(N\) is the number of
      nucleotides.

---

\(^{15}\) Here we are referring to an *extensive* property, which tells us how much stuff we will need. An *intensive* property does not depend on the amount of material. For example, NMR requires ca. 1 mL of a 1 mM solution; 1 mM is intensive, and 1 µmol \(\approx 10^{-4}\text{L} \times 10^{-3}\text{M}\) is extensive.

\(^{16}\) This section is largely from Cantor & Schimmel Biophysical Chemistry, vol I.

\(^{17}\) Be careful not to overgeneralize, however, since similar structures can have different functions, and similar functions may have diverse structures.
ii.) If \( N = 80, >2.6 \times 10^{20} \) structures! But age of the universe is only \( \approx 10^{10} \) years.  
(This problem is often referred to as Levinthal’s Paradox)

iii.) Structure is therefore not necessarily the equilibrium product. It may be 
kinetically trapped or determined, and therefore folding may be path dependent. 
  e.g. Insulin

2. **Function**
   a.) Can the function be predicted or rationalized.  
   Is a process plausible? *In research, the liberal use of common sense cannot be 
   overemphasized.*  
   Is it energetically or kinetically favorable?

b.) Is it necessary?  
   What subunits or sequence can be removed or deleted without changing function?  
   *in vitro? in vivo?*

   c.) What other molecules does it interact with?  
   Functional studies.  
   Structural model of substrates.  
   How does presence of other binders affect activity?  
   Competitive binding by other molecules?  
   Cooperative binding by other molecules?

   d.) What is the structure of the functional site\(^\text{18}\)?  
   What region of the macromolecule?  
   What atoms?  
   Is function accompanied by a structural change?

   e.) How is function regulated?  
   Total amount of enzyme present?  
   Fraction of enzyme total that is active? Effectors? Inhibitors?  
   Overall functional communication (transmission of binding energy throughout 
   molecule to change function)?

---

\(^{18}\) In the case of enzymes, this is referred to as the active site. In other molecules that do not do chemistry (i.e. don’t make and break chemical binds) but instead bind ligands (e.g. hemoglobin), this will be referred to as the functional site.
**Strategies to Answer these Questions**

1.) Use a smaller system as a model that still has your activity.

   e.g.: Use the following ‘Boundary Experiment’, which is a form of combinatorial chemistry.

2.) Observe only one part of the system

   e.g.: Absorption of heme in visible region of spectrum

   e.g.: Labels (but peeking may perturb!) such as fluorescence or spin labels.

3.) Compare two systems that are almost identical

   a.) Macromolecules with similar functions

   b.) Same function, different organisms

   c.) Mutated molecules

      Genetics

      Site-directed mutagenesis

      Chemical Modification (e.g.):

4.) Isolate discrete states for the molecule

   Vary temperature, pH, salt, macromolecule concentration
Major Goal of This Course: *To dissect molecular biology papers at the level of the physical chemistry of these processes.*

- **An Example:** A protein binding to a regulatory DNA sequence. What structural features cause specific binding to occur?

- How can the protein search so much intracellular DNA to find that particular sequence?

- How much energy is required to deform the DNA and protein for optimal binding?

- How can the effects of mutations on binding strength be understood?

- How does change in ionic strength affect binding?

- How does protein binding affect the structure and activity of distant DNA sequences? Binding of other proteins?