Nucleic Acids: Properties, Structures, and Functions
http://courses.chem.psu.edu/chem572
April 11, 2016
Philip Bevilacqua

Lecture 36b: Nucleic Acid Structures from Diffraction Methods

Reading: BCT Chpt 4

Overview

- We will spend a relatively short period of time on this expansive topic. Our goal will be to be able to read, understand, and evaluate papers from the literature on nucleic acid structure using diffraction. We will use Kielkopf et al.¹ as the way to introduce and understand this material.

- We have discussed methods to get low resolution structure (including structure mapping, modification interference, NAIM, NAIS, crosslinking, mutagenesis...). However, to get detailed, atomic resolution structures we need to use x-ray diffraction or NMR.

Why do we want high resolution structure?

- The main motivation here is to make the connection between structure and function in biological chemistry. Dickerson, an important nucleic acid crystallographer, sums up the issue at hand, "In what way does sequence influence the deformation and deformability of the helix (i.e.: structure), and are either of these properties used in the recognition process (i.e.: function)."

Diffraction: Systems

- To carry out X-ray diffraction, one needs either single crystals (for highest resolution) or oriented fibers (which Rosalyn Franklin used in 1953).

- One key theme that has emerged from diffraction methods is that at the atomic level, DNA structure is polymorphic. This variation of DNA structure with sequence helps explain how sequence-specific recognition can be achieved by various agents such as proteins and drugs. (Please re-read p2 of BCT, especially the first bullet.) We will look at ways to describe this heterogeneity shortly.

- The best known examples of dsDNA structures are B, A, and Z-DNA; of course, Z-form DNA is unusual in that the duplex is left-handed. Interestingly, a form of DNA resembling Pauling’s original model with the bases on the outside and the phosphates on the inside has been found for stretched and overwound DNA.²

- The structure of RNA can also be obtained by X-ray diffraction methods. The simplest RNA structure is dsRNA, which is of the A-form. Important advances have occurred in the last five years and have led to the structures of increasingly complex and novel RNA molecules. These include structures of large, complex, globular RNA and RNA-protein complexes, also referred to as RNP’s. The major early breakthrough in RNA structure occurred with tRNA, which is about 80 nts and has a complex tertiary structure (early 1970’s). More recently, riboswitches, small ribozymes (catalytic RNAs) have been solved in various states, including pre-cleaved, cleaved, metal bound, and bound with transition state mimics. In addition, the structure of the large subunit and small subunits of the ribosome have been solved. Unfortunately, time prevents us from going into these

systems in any kind of detail, but they are fascinating problems and we are sure to see additional major breakthroughs in the coming years.

• For our purposes, we will look at the basics of crystallography by focusing on dsDNA and on Dervan’s polyamide compounds bound to dsDNA.

**Diffraction: Methods**

• **Low resolution fiber diffraction.** See Fig 4-1. Note that there are very few spots on the film.

• **High resolution fiber diffraction.** See Fig 4-2. Note the increasing number of spots on the film.

• **High resolution single crystal diffraction of dsDNA.** See Fig 4-3. Note the large number of well-defined and intense spots on the film.

• Bragg’s Law allows us to understand the basic features of fiber and crystal diffraction.

\[ n\lambda = 2d \sin\theta_b \]

• Where, \( n \) is an integer

  \( \lambda \) is the x-ray wavelength. (In Kielkopf et al. this is = 0.98 Å--see Table 1 heading.)

  \( \theta_b \) is 1/2 the scattering angle

  \( d \) is the regularly repeating features in the structure

• Note that because \( n\lambda \) is a constant,

\[ \sin\theta_b \propto 1/d \]

• High resolution means seeing minute features, which means \( d \) is small. Small \( d \) means maximum in \( \sin \theta_b \), which means \( \theta_b \) being towards 90°. This means that the highest resolution spots scatter towards the outermost part of the film. Take home message is SPOTS ON THE OUTSIDE OF THE PATTERN ARE GOOD. Note this feature in Fig 4-3.
Fiber Diffraction Methods

- Gross features of DNA structure can sometimes be obtained by inspection of fiber diffraction patterns.

Limitations to Fiber Diffraction Data

- Structure is underdetermined. (i.e.: there is not enough data to determine the structure well.)
- Dependent on model chosen and the following assumptions.$^3$
  1. All repeating units have the same conformation
  2. Connections between the repeating units is identical
  3. The backbone conformation is independent of sequence.
  4. Structure is C2 symmetrical

Single-Crystal Diffraction Methods

- Produces lots of data. In general ca. 1,000 to 10,000 spots, or internal reflections, are found in a typical dsDNA. The best structures have the most data. See Table 1 of Kielkopf et al. for their number of measured reflections.
- Note that not all reflections are useful. Important to look at the number of unique reflections. Typically want this number to be >50/bp. In Kielkopf et al., they have ca. 300/bp (see Table 1).
- It is harder to get crystals than fibers. Getting crystals is often the rate-limiting step in solving the structure, especially today with the availability of excellent software for solving structures.
- The first single crystal structure of a DNA was in 1979, and was of Z-DNA by Alexander Rich et al. It is curious that the first high resolution structure of DNA was of left-handed DNA. In 1980, Dickerson solved the structure of a dodecamer (12 bp of dsDNA).
- Crystals typically contain cations, spermine (a polyamine cation with a charge of 4+), and many water molecules. Spermine: $\text{H}_3\text{N}^+(\text{CH}_2)_3\text{N}^\prime\text{H}_2(\text{CH}_2)_4\text{N}^\prime\text{H}_2(\text{CH}_2)_3\text{NH}_3^+$
- A dehydrating alcohol is typically used (e.g. organic diols such as ethylene glycol). This is to help ‘dry out’ the DNA and force it to crystallize. Often the DNA is hung in a drop on a cover slip above a solution. This ‘hanging drop’ method works by concentrating the DNA to the point where it hopefully crystallizes before precipitating. Do be aware that the crystal is formed under conditions that are not the same as in solution.
- High solvent content (ca. 50-75%) often leads to disorder and low resolution (>2 Å).

Phase problem:

$^3$Note that what is common to these assumptions is that the structure is homogeneous. Obviously this is not true, as revealed by diffraction on single crystals (see below). These are the assumptions that Watson and Crick used when solving the structure of dsDNA, and it was not until the 1980s that the microheterogeneity of dsDNA structure was revealed.
• The intensity pattern on the diffraction is missing phase information on x-rays. This is gained by collecting data in the presence of heavy atom derivatives, which alter the phase by scattering the x-rays somewhat. This is a complex mathematical process that we will not get into, but we should be aware of ways to solve the phase problem.

• Proteins (and large RNAs and RNPs): 1.) Use isomorphous replacement with heavy atoms such as Hg, Pb, Sm, ... 2.) Use multiwavelength anomalous diffraction (MAD) phasing. This method was pioneered by Wayne Hendrickson. Synchrotron x-rays are used at two different wavelengths, and the ‘heavy’ atom is typically a medium-heavy atom such as selenium, which can be readily incorporated, and often with minimal perturbation to the structure, as selenomethionine.

• dsDNA: Use molecular replacement based on idealized models and refine until the disagreement with a model is minimized. (This was used in Kielkopf et al.—see Table 1 heading.)

Judging the quality of the structure
• The goodness of fit of a proposed structure is compared with the observed diffraction pattern and is summarized in terms of the R-factor (in %). $R_{\text{cryst}}$ should be <25% for a good structure. (Kielkopf et al. have $R_{\text{cryst}}$ of 21 to 22%—see Table 1).

$$R = \frac{\sum |F_0| - |F_c|}{\sum |F_0|}$$

where $F_0$ and $F_c$ are the observed and calculated factors, respectively (see Eq 4-2 p82 of BCT).

• You can lower the R-factor by including small molecules such as water, if justified by the data. Note that there shouldn’t be many more than 10 water molecules/bp, otherwise the data may be being artificially ‘improved.’ (Kielkopf et al. have 6.5 to 8.7 water molecules/bp—Table 1.)

• $R_{\text{free}}$ should be within 5% of $R_{\text{cryst}}$ (see Kielkopf et al. Table 1.)

• We need to know the position for each non-hydrogen atom in the molecule. Assumptions of standard bond lengths and angles reduce the number. For example, 12 bp of dsDNA needs ca. 1,500 independent, or unique, data reflections. Kielkopf et al. have about 3,000 in 10 bp (Table 1).

• The resolution should be under 3Å. Here are some rules of thumb.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Å</th>
<th>Structural Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>3-7</td>
<td>Secondary structure (but not individual atoms)</td>
</tr>
<tr>
<td>Medium</td>
<td>2-3</td>
<td>Side Chains (nucleobases and amino acids)</td>
</tr>
<tr>
<td>High</td>
<td>&lt;2</td>
<td>Individual atoms</td>
</tr>
</tbody>
</table>

• Remember that hydrogen atoms are too small to scatter electrons and are therefore not observed. The positions of hydrogens are inferred from standard p$K_a$ assignments. This is

4 These assumptions are OK depending on the case. For example, aromatic bases seldom deviate from the expected bond lengths and angles and planarity, so assumptions of standard bond lengths and angles are often used here. However, sugars pucker and dihedral angles vary depending on the environment, and assumptions should be avoided here if at all possible, and they should be determined from real experimental data.
fine unless a base happens to have a shifted $pK_a$. Shifted $pK_a$’s too are often inferred from structures. For example, if two heteroatoms, such as two ring nitrogens end up close (2.2 to 3.5 Å apart), then it is supposed that one of them is protonated to allow a hydrogen bond to form. This assumption is reasonable, but remember that it is an assumption.

- Hydrogen atoms can be seen by neutron diffraction methods. And while highly useful, neutron diffraction is less common and generally not used unless one is particularly interested in the positioning of hydrogen atoms.

**Problems with Crystallography of DNA**

- Not limited by size of nucleic acid like NMR is.
- If part of the structure is flexible, you won’t see that part of the structure.
- If water molecules or metal ions bind in an *average* of several sites, you won’t see them.
- No information is gleaned about conformational variability. At best you will get one of several conformations, that which crystallized best, which is usually lowest in energy. However the one that biologically active state might be a higher energy state.

**Advantages of Crystallography of DNA**

- Local variations within one conformation are seen. These are of interest for understanding sequence-specific interactions with ligands (proteins, drugs, metal ions).
- Local variation can be caused by
  - base sequence (interesting)
  - crystal contacts (not interesting)
  - inadequate data (not interesting)
- To know about crystal contacts, must compare two sequences in different space groups (or one sequence that happens to crystal pack in two or more ways).
- Often characterize structures by space group and unit cell dimensions. (See Kielkopf et al. Table 1).

**Specific Features Found in Nucleic Acid Crystal Structures**

- Fig 4-4 (p89) shows major features of A-, B-, and Z-form dsDNA. See also Table 1 (p90).

Some terms:
- **Pitch**—Distance along the axis per turn of the helix. See Table 4.1 (BCT). These values are typically about 34 Å for B-form DNA. This is also some times reported in ‘per bp’ units in the literature (i.e. different people sometimes use it differently). See for example Table 3 of Kielkopf et al. (3.4 Å). Note that BCT refers to ‘per bp’ units as Rise/bp in Table 4.1.
- **bp/turn**—typically 10 for B-form DNA. See Table 4.1 of BCT.
- **bp inclination**—tilt of bp from vertical axis. Close to zero for B-form DNA. (Like a spiral staircase.) About 12° for A-form DNA. (Like a spiral staircase with tilted stairs.)
• Twist—This is the number of degrees turned about the dyad axis when going from one bp to the next. As you might imagine, for B-form DNA it is about 36° (=360° / 10 bp). However, again the microheterogeneity of DNA rears its head and the variation in twist is large, ranging from 24° to 51°.

• Note that there is a strong dependence of twist on sequence. See Fig 4-6. In general there is low twist for purine-purine steps (this is because it maximizes stacking of the adjacent purines). Other generalizations on twist are given on p99.

• Fig 4-5 defines translations and rotations of the bases. Note that these vary along a helix (Fig 4-6). Note that shear, opening and propeller twist are discussed in Table 3 of Kielkopf et al.

• Unusual structures often result from highly repetitive sequences.

\[ \text{d(G}_4\text{C}_4\text{)}_2 \text{ is near A-form due to stacking of the purine steps.} \]

\[ \text{d(CGCGCG)} \text{ is Z-form (this is recognized by TLR9 in innate immunity).} \]

• oligo dA-oligo dT tracts propeller twist so much that each base forms bifurcated hydrogen bonds to its Watson-Crick partner and the base to the 5’ side of it. This likely occurs because the AT base pair has only 2 hydrogen bonds and so retains enough flexibility to highly propeller twist. (GC base pairs don’t propeller twist as much.) See Table 3 of Kielkopf for extra propeller twisting of AT base pairs.

• A practical point: Avoid strongly repeating sequences for your experiments.

• Importantly, DNA helices with imperfections in base pairing usually have the perturbation localized to the mismatch and maybe its nearest-neighbor. This allows the nearest neighbor theory to work reasonably well on large nucleic acid structures.

• Sequences which form hairpins in solution will often crystallize as a duplex due to high effective local concentration of strands.

• dsRNA is A-form and similar to B-form dsDNA. dsRNA has been of strong interest recently in the literature due to its involvement in RNA interference (RNAi). In general, dsRNA is more stable than dsDNA due to O2’ hydrogen bonding with O4’ of following residue. These leads to a more ‘squat’ structure, with a smaller rise/bp. (see Table 4.1).

The local mobility of double helical nucleic acids

• The crystal gives a ‘thermal parameter’ or B-factor for each residue, which is a measure of mobility. A high B-factor is indicative of local disorder, which means that part of the structure is less-well defined, and might be mobile.

tRNA

• tRNA is a treasure trove of structural information on RNA.

• The structure is ‘L-shaped’, with about 90% of the bases being stacked. Tertiary interactions largely occur in the ‘L’.

---

\(^5\) See Table 4.1 to find that twist of A-DNA is less than that of B-DNA (ca. 32 and 36°, respectively).

\(^6\) We will see this in coming lectures. The nearest-neighbor theory treats thermodynamics of nucleic acids as the sum of additive local interactions acting over adjacent base pairs only.
• The ends of the molecule are more flexible than the middle.

• Multivalent ions (especially Mg\(^{2+}\)) help fold the RNA by neutralizing the close approach of the phosphate backbone. Mg\(^{2+}\) acts like a ‘glue’ to hold the structure together.