X-ray crystallography

CS/CME/Biophys/BMI 279

Nov. 12, 2015

Ron Dror
Outline

• Overview of x-ray crystallography
• Crystals
• Electron density
• Diffraction patterns
• The computational problem: determining structure from the diffraction pattern
Overview of x-ray crystallography
X-ray crystallography is the most common way to determine 3D molecular structures

- 90% of the structures in the PDB were determined through x-ray crystallography
- X-ray crystallography is also frequently used to determine structures of other biomolecules (e.g., RNA) or of small molecules (including drugs)
- Why are we covering it in this course?
  - So you know where protein structures come from
  - Because determining a structure this way involves solving a challenging computational problem
The basic idea

- Get the molecule whose structure you want to determine to form a crystal
- Shine an intense beam of x-rays through the crystal, giving rise to a “diffraction pattern” (a pattern of spots of varying brightnesses)
The basic idea

• From that pattern, infer the 3D structure of the molecule
  – In fact, we use multiple images, with the x-rays shining through the crystal at different angles

• This is a challenging computational problem!

• It turns out the diffraction pattern is closely related to the *Fourier transform* of the electron density of the molecule we crystallized
  – Before we even worry about what that means, let’s go back and discuss what a crystal is and what electron density is

Crystals
What’s a crystal?

- Under certain conditions, molecules line up into a regular grid (a “lattice”).
  - Example: table salt

http://www.bigfoto.com/miscellaneous/photos-16/salt-crystals-94jf.jpg

Proteins can also form crystals

• Under certain conditions, entire proteins will pack into a regular grid (a lattice)

Insulin crystals

http://science.nasa.gov/media/medialibrary/1999/09/10/msad20sep99_1_resources/9901879.jpg
Proteins can also form crystals

- Under certain conditions, entire proteins will pack into a regular grid (a lattice)

Multiple views of the crystal formed by an immunoglobulin-binding domain (PDB entry 1PGB)

http://www.umass.edu/molvis/decatur/pe2.727/protextlcon.htm
Caveats

- Getting proteins to form crystals can be hard
  - Crystallographers sometimes work for decades to get good crystals of a particular protein
Caveats

• Sometimes a protein will adopt a different structure in a crystal than it does in its natural environment

• Crystallography gives you a static snapshot of a protein’s structure
  – Usually (but not always) this snapshot corresponds to the protein’s “average” structure
Electron density
Electron density of a molecule

- The *electron density* corresponding to the 3D structure of a molecule gives the probability of finding an electron at each point in space.
- X-rays bounce off electrons they hit.
Diffraction patterns
Diffraction patterns

- When you shine a light beam through a crystal, you get a distinctive pattern of bright spots called a diffraction pattern.

The dark spots are sometimes pictured in light shades (white) and sometimes in dark shades (black).
Diffraction patterns

- This pattern is actually three dimensional.
  - If you move the imaging plane (or rotate the crystal), you see different parts of it
What causes diffraction patterns?

• Short answer: interference of light
  – The bright spots are places where light interferes constructively. Elsewhere it tends to interfere destructively (cancel out).

You’re not responsible for this

Relationship between diffraction pattern and electron density

• It turns out that the diffraction pattern is the *Fourier transform* of the electron density
  – Both the electron density and the diffraction pattern are functions of three dimensions (i.e., defined at every point in a 3D volume)
  – Each bright spot in the diffraction pattern corresponds to one sinusoidal component of the electron density
  – The Fourier transform gives a magnitude and a phase (shift) for each sinusoid, but it’s only practical to measure the amplitude, not the phase
• Brightness of the spot gives the magnitude

• You need not understand why this relationship holds
The computational problem: determining structure from the diffraction pattern
The challenge

• Given a diffraction pattern, determine the electron density and/or the position of each atom
• If we had a magnitude and a phase associated with each spot in the diffraction pattern—and thus with each 3D sinusoid—then we could just sum up appropriately scaled and shifted 3D sinusoids to recover the electron density
• But we don’t have the phases
  – This makes the problem “underdetermined”—in principle, multiple electron densities could give rise to the same set of diffraction pattern magnitudes
  – But the vast majority of those won’t correspond to reasonable 3D structures of the protein
General approach to solution

• **Step 1: Initial phasing**
  – Come up with an approximate solution for the structure (and thus an approximate set of phases)

• **Step 2: Phase refinement**
  – Then consider perturbations to the structure
  – Search for perturbations that improve the fit to the experimental data (the diffraction pattern)
Initial phasing

• The most common method for initial phasing is *molecular replacement*
  – Start with a computational model of the protein structure (often the structure of a homologous protein)
  – Search over the possible ways that a protein with this structure could be packed into a crystal, and find the one that gives the best fit to the data

• If one can’t build a good computational model of the protein, then one can try various experimental methods to help determine phases
  – Example: *isomorphous replacement*, where one replaces several atoms of the protein with heavier atoms (usually metals), and then uses the change in the diffraction pattern to solve for the phases
    • You’re not responsible for this
  – Even with additional experimental information, one generally still needs to solve a computational problem
Phase refinement

• Once we have an initial model, we can search for perturbations to that model that improve the fit to the experimental data
  – This is usually done through a Monte Carlo search (via simulated annealing)
  – One usually restrains the search to “realistic” molecular structures using a molecular mechanics force field
    • This dramatically improves the accuracy of the results
    • The idea was introduced by Axel Brunger, now on the Stanford faculty
Phase refinement

• A major challenge in the phase refinement process is to avoid overfitting—i.e., fitting to the noise in the experimental measurements.

• To avoid this, one generally ignores a small subset of the experimental data during the refinement process, then sees how well one can predict it at the end.
  – Just like cross-validation in machine learning.
  – This idea also came from Brunger.
Computational methods continue to improve

• Although the phasing problem is decades old, researchers are still inventing better solutions
A few additional notes

• Protein crystals contain water
  – Often half the crystal is water
  – Usually only a few water molecules are visible in the structure, because the rest are too mobile

• One usually can’t determine hydrogen positions by x-ray crystallography
  – But one can model them in computationally

• Some high-profile, published crystal structures have turned out to be completely incorrect, due to computational problems/errors