X-ray crystallography

CS 279 / BIOE 279 / BIOPHYS 279 / BMDS 245 / CME 279

Nov. 12 and 19, 2026

Ron Dror

Outline

- Overview of x-ray crystallography
- Crystals
- Electron density
- Diffraction patterns
- The computational problem: determining structure from the diffraction pattern

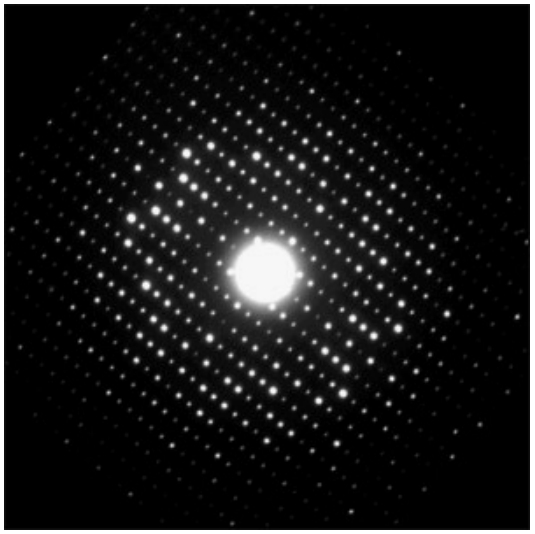
Overview of x-ray crystallography

X-ray crystallography is the most common way to determine 3D molecular structures

- Almost 80% of the structures in the PDB were determined through x-ray crystallography
- X-ray crystallography is also frequently used to determine 3D structures of small molecules (including drugs) on their own
- Over 30 Nobel Prizes awarded for work that involved or related to crystallography
- Why are we covering it in this course?
  - So you know where biomolecular structures come from
  - Because determining a structure this way involves solving a challenging computational problem
    - When crystallographers determine a structure, they typically say they “solved” the structure

The basic idea

- Make a crystal composed of the molecule whose structure you wish to determine
- Shine an intense beam of x-rays through the crystal, giving rise to a “diffraction pattern” (a pattern of spots of varying brightnesses)

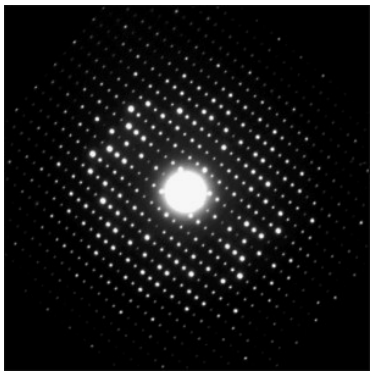


Note: In this lecture, I will often refer to the molecule being studied as “the protein”, but it could be any molecule, or a complex of multiple molecules bound together (e.g., protein+ligand)

http://lacasadeloscristales.trianatech.com/wp-content/uploads/2014/09/image005-300x300.jpg

The basic idea

- From that pattern, infer the 3D structure of the molecule
  - The diffraction pattern is 3D: one uses 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 that was crystallized
  - Before talking about what that means, let's go back and discuss what a crystal is and what electron density is



http://lacasadeloscristales.trianatech.com/wp-content/uploads/2014/09/image005-300x300.jpg

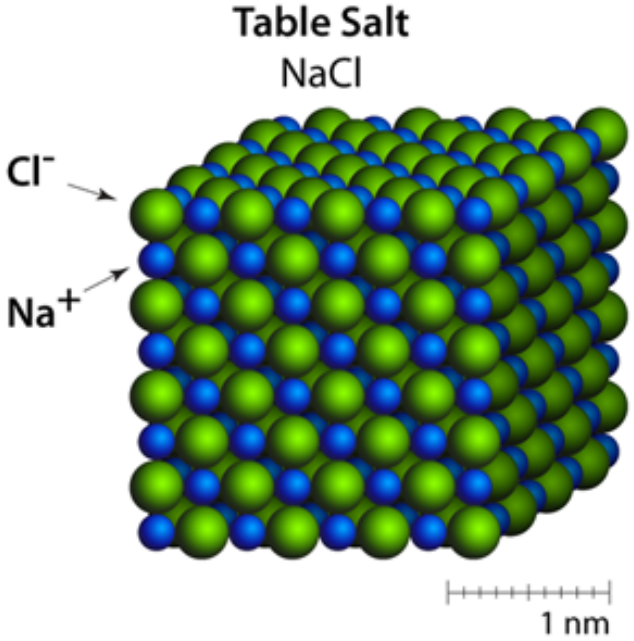
Crystals

What's a crystal?

- Under certain conditions, molecules come together to form a regular grid (a "lattice").
  - Example: table salt



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

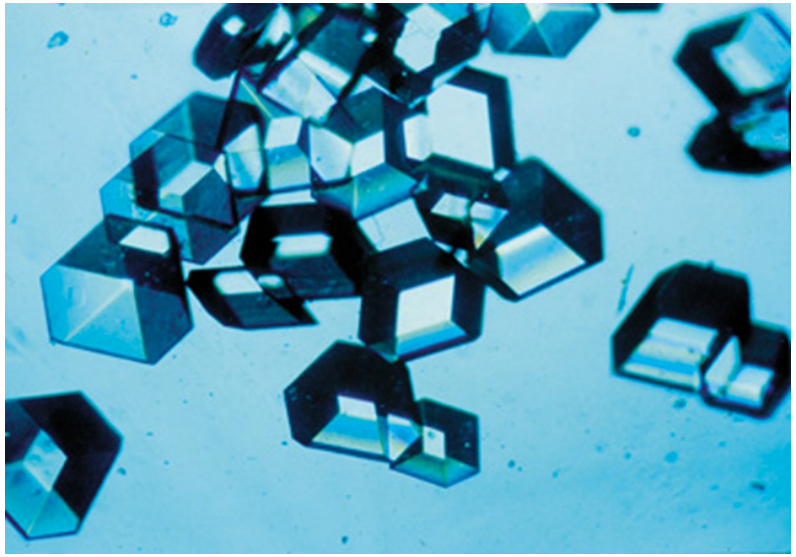


http://www.atomsinmotion.com/book/chapter4/rockSalt.png

Macromolecules can also form crystals

- Under certain conditions, proteins and other macromolecules will pack into a regular grid (a lattice)

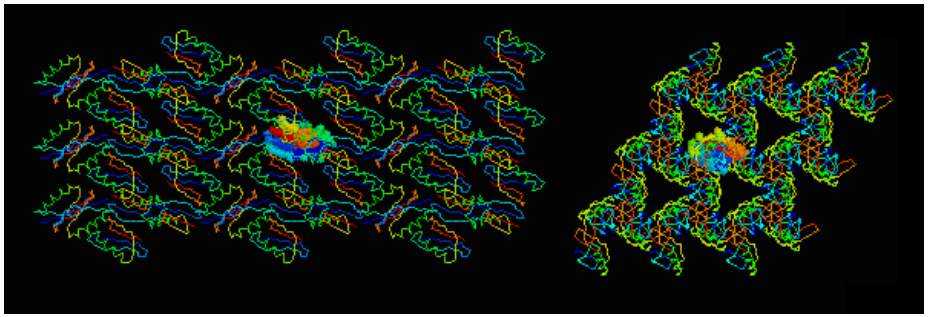
Insulin crystals

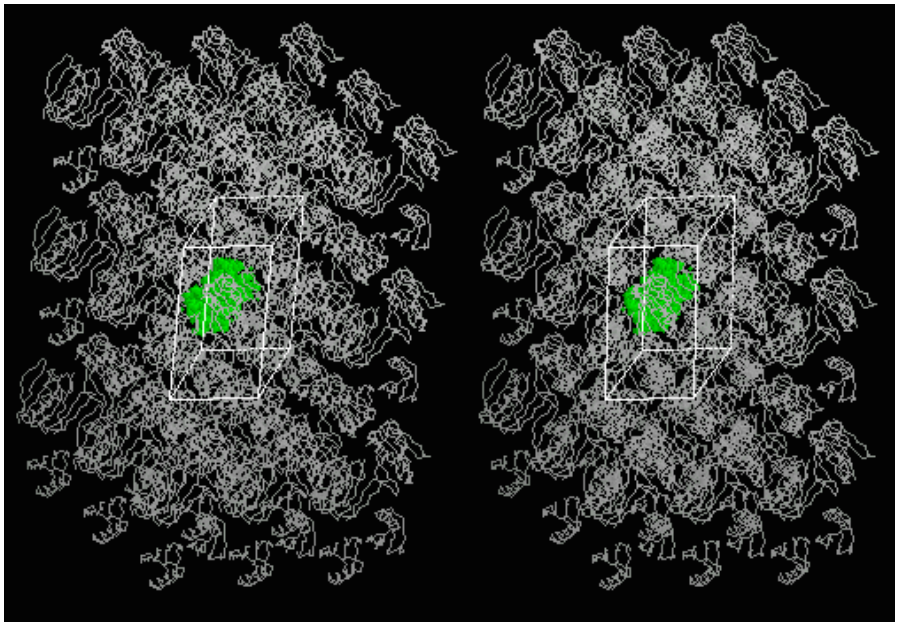


http://science.nasa.gov/media/medialibrary/1999/09/10/msad20sep99_1_resources/9901879.jpg

Macromolecules can also form crystals

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





http://www.umass.edu/molvis/decatur/pe2.727/protexpl/xtlcon.htm

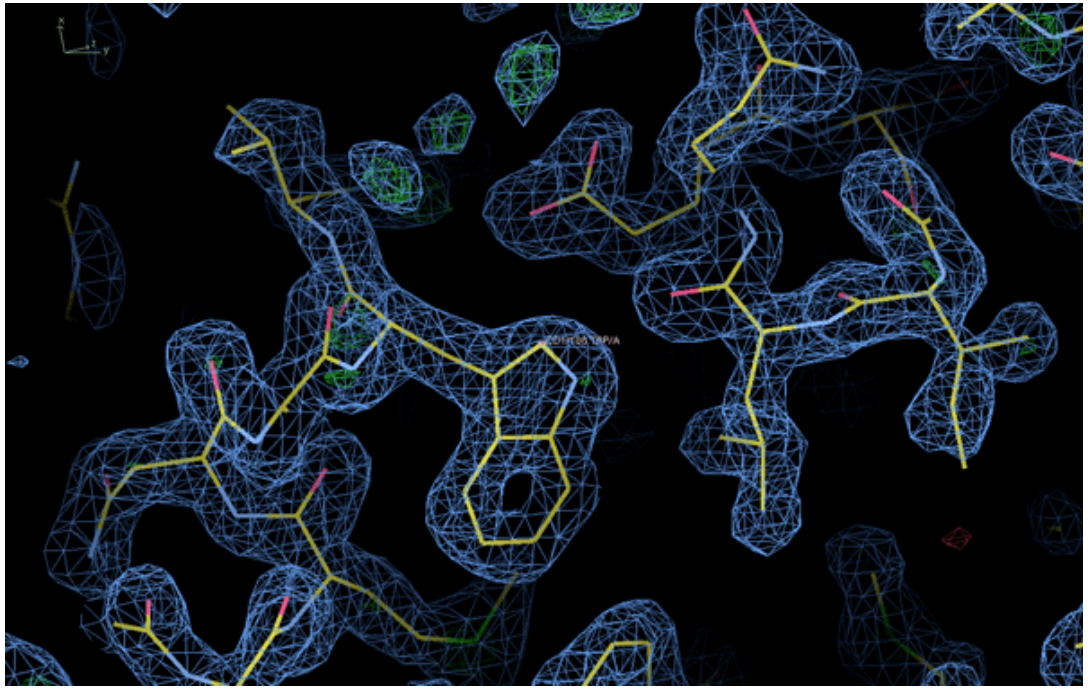
Caveats

- Getting macromolecules to form crystals can be hard
  - Historically, crystallographers sometimes worked for years or decades to get good crystals of a particular molecule (though it's become easier over time thanks to better experimental methods)
- Crystallography gives you a static snapshot of a molecule's structure
  - Usually this snapshot corresponds to the molecule's "average" structure
- Sometimes a molecule will adopt a different average structure in a crystal than it does in its natural environment
  - Fortunately, these differences are usually small

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

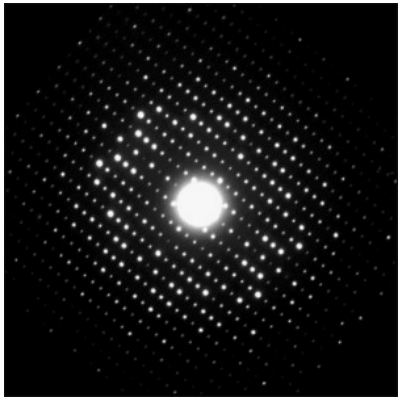


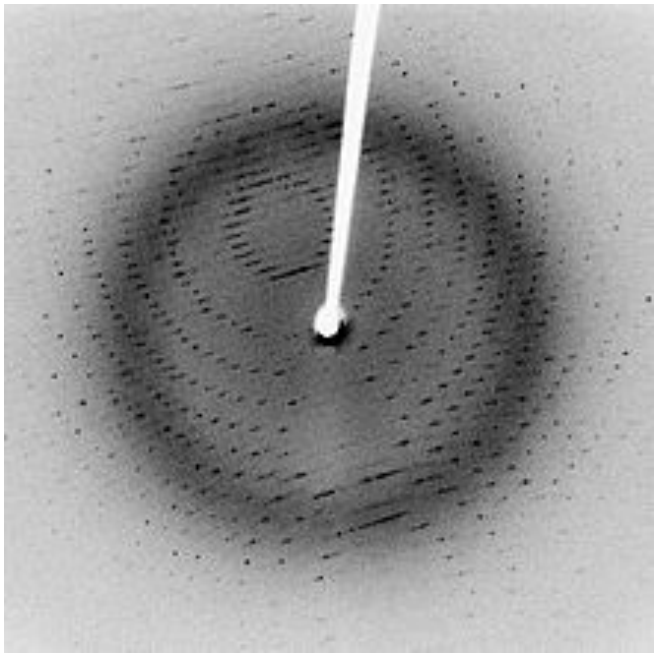
http://www.lynceantech.com/images/electron_density_map.png

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

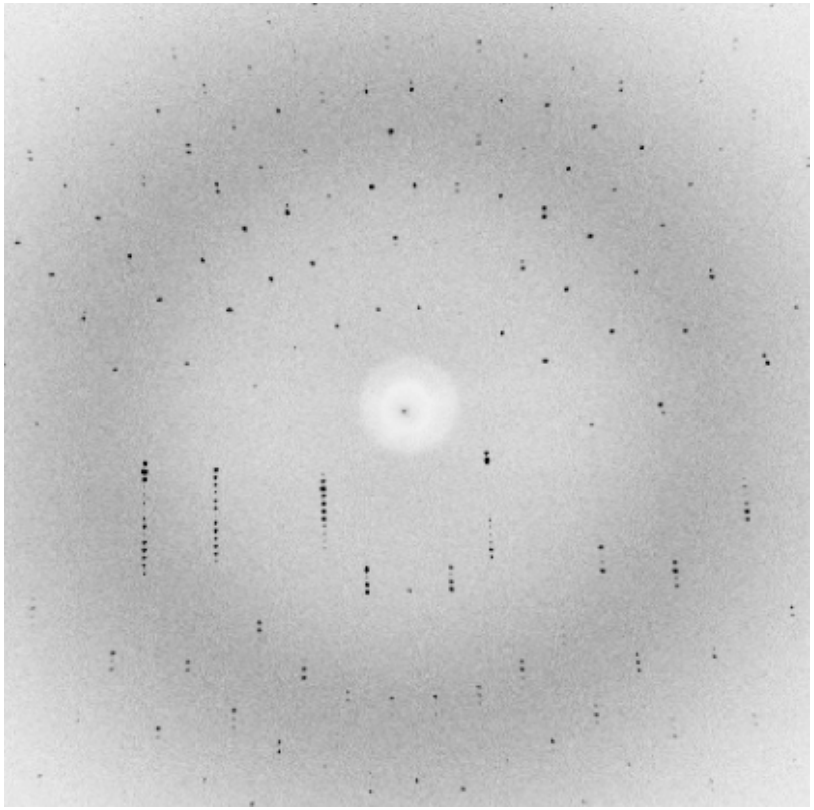




Note that the bright spots are sometimes pictured in light/white shades (left) and sometimes in dark/black shades (right)

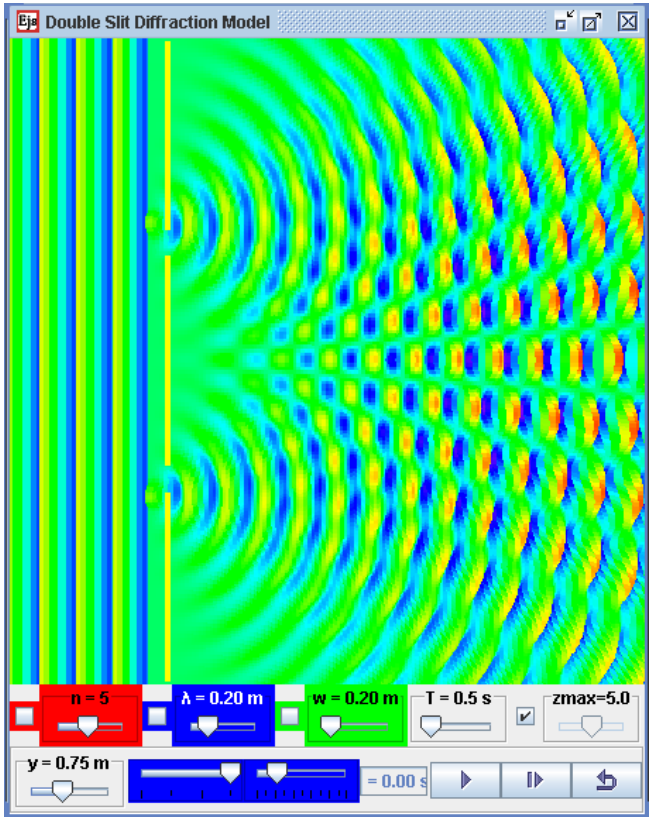
Diffraction patterns

- This pattern is actually three dimensional.
  - If you rotate the crystal (or move the camera), you see different parts of it



What causes diffraction patterns?

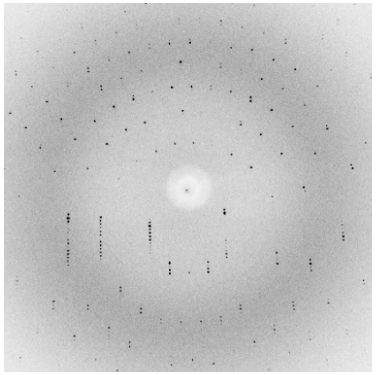
- Short answer: interference of light
  - The bright spots are places where light interferes constructively. Elsewhere it tends to interfere destructively (cancel out).



http://weelookang.blogspot.com/2011/10/ejs-open-source-double-slit-diffraction.html

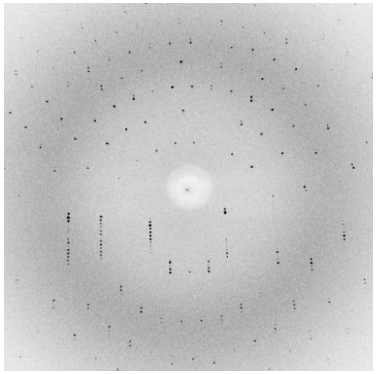
Relationship between diffraction pattern and electron density

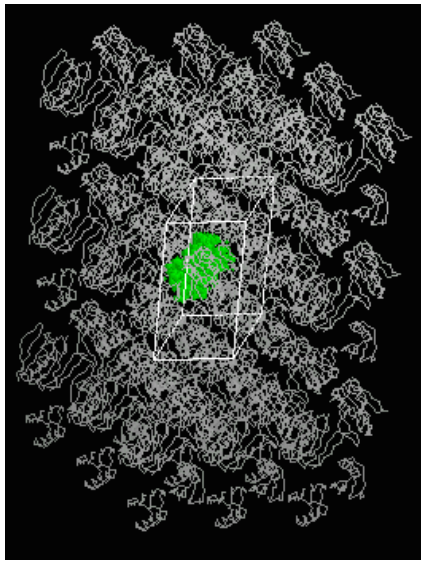
- It turns out that the diffraction pattern is the *Fourier transform* of the electron density
  - Both the electron density and the diffraction pattern are functions of three dimensions (i.e., defined at every point in a 3D volume)
  - Each bright spot in the diffraction pattern corresponds to one sinusoidal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



Why is the diffraction pattern the Fourier transform of the electron density?

- If you work out how much each point in the electron density contributes to a particular diffraction peak, the answer is a sinusoid (whose period fits into the unit cell an integer number of times in each dimension, so that each unit cell contributes equally).
- In between the peaks, each unit cell contributes differently, resulting in destructive interference.
- For a more detailed explanation, see Notes on course web site.





You need to understand the Fourier transform relationship, but not why it 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 diffraction pattern (i.e., same magnitudes for each spot)

Why is it possible to solve this problem?

- In principle, multiple electron densities could give rise to the observed diffraction pattern
- **But the vast majority of those densities won't correspond to plausible 3D molecular structures**
- For example, we know that:
  - Electron density should never be negative
  - Electron density should correspond to the atoms in the crystallized protein/macromolecule
    - And, perhaps, other atoms present when the crystal formed
  - Atoms connected by a covalent bond will be near one another
- The goal is to find a 3D structure that is both physically reasonable and consistent with the observed diffraction pattern
- **Discuss: How would you do this?**

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—by far—is *molecular replacement*
  - Start with a computational model of the protein's structure, such as an AlphaFold prediction
  - 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, then one can try various experimental methods to help determine phases, but this is rarely necessary today
  - Examples: anomalous dispersion and isomorphous replacement

Phase refinement

- Once we have an initial model, we can search for perturbations to that model that improve the fit to the experimental data
  - **One usually restrains the search to "realistic" molecular structures using a molecular mechanics force field**
    - Search for a structure that minimizes the sum of two terms: (1) the difference between the observed diffraction pattern and the diffraction pattern calculated from the structure, (2) the calculated energy of the structure
    - This dramatically improves the accuracy of the results
    - The idea was introduced by Axel Brunger, now on the Stanford faculty
  - The search is usually done by Monte Carlo sampling, with simulated annealing

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 overfitting, one generally ignores a small subset of the experimental data during the refinement process, then checks how well one can predict it at the end
  - Just like cross-validation in machine learning
  - This idea also came from Brunger (who introduced the term R_{free} to quantify the error in the prediction)

Optional material

Modern structure prediction methods can help in phase refinement

- Several recent methods improve phase refinement using AlphaFold (or related structure predictors). Examples:
  - Start with an AF2 model, refine it using standard methods, use the refined model as a template for a new AF2 prediction, and iterate
  - Optimize the MSA cluster profile for a protein (with a linear bias including scaling and offset factors) to maximize agreement between the AF2 prediction and the experimental data

Serial crystallography with XFEL

Optional material

- Using an extremely bright x-ray beam, one can detect diffraction patterns from extremely small crystals (which are easier to obtain) or, in principle, even single particles (e.g., viruses)
- This is now possible thanks to x-ray free electron lasers (XFELs), pioneered at Stanford/SLAC
- Challenge:
  - The tiny crystal (or particle) disintegrates when the laser beam hits it
  - For each crystal or particle, one can only capture a single image. One repeats this for many crystals/particles.
  - This makes the computational reconstruction problem more challenging!
    - We'll cover related problems in the cryo-EM lecture

A few additional notes

- Protein crystals contain water
  - Often half the crystal is water (filling all the empty spaces between copies of the protein)
  - Usually only a few water molecules are visible in the structure, because the rest are too mobile
- One generally 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