

# X-ray crystallography

CS/CME/BioE/Biophys/BMI 279

Nov. 9 and 11, 2021

Ron Dror

# Outline

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

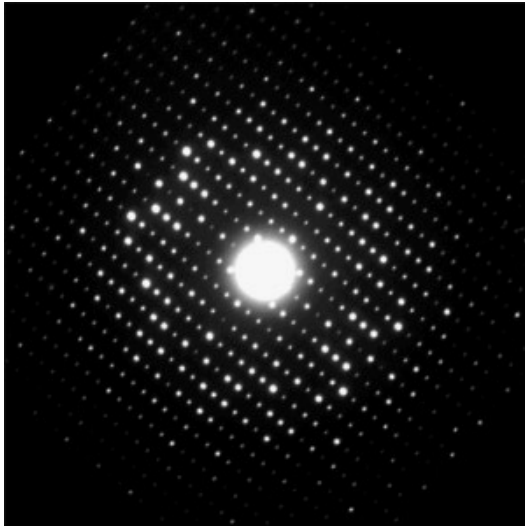
# Overview of x-ray crystallography

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

- Nearly 90% 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)
- 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

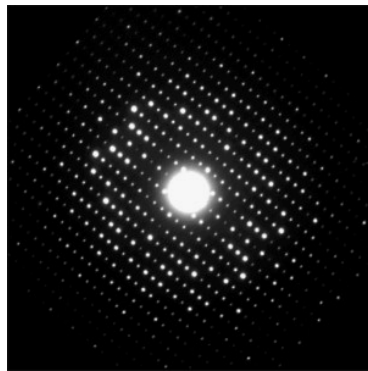
- Make a crystal composed of the molecule whose structure you wish to determine
- Shine an intense beam of x-rays through the crystal, giving rise to a “diffraction pattern” (a pattern of spots of varying brightnesses)



<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 we even worry 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>

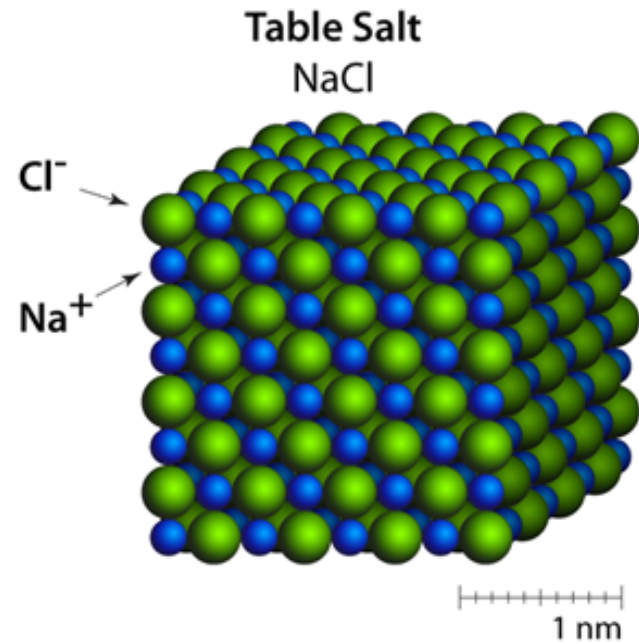
# Crystals

# What's a crystal?

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



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

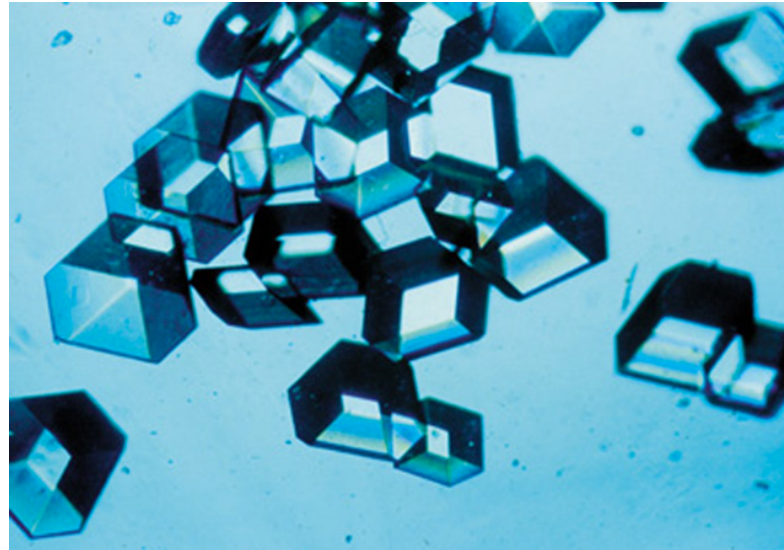


<http://www.atomsinmotion.com/book/chapter4/rockSalt.png> 8

# Macromolecules can also form crystals

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

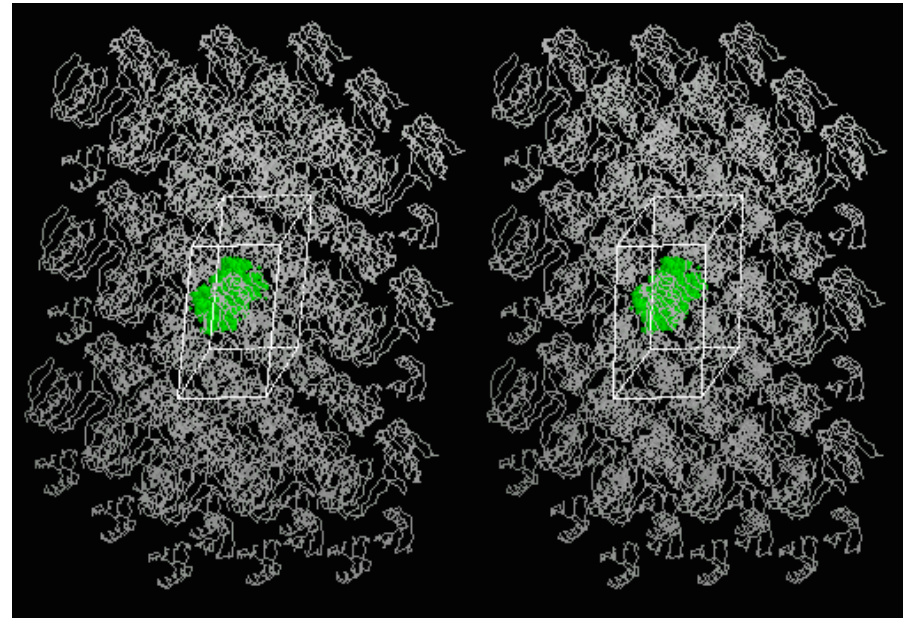
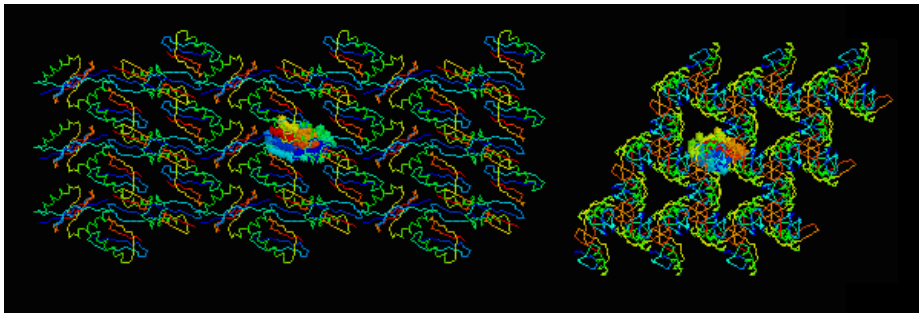
Insulin crystals



[http://science.nasa.gov/media/medialibrary/1999/09/10/msad20sep99\\_1\\_resources/9901879.jpg](http://science.nasa.gov/media/medialibrary/1999/09/10/msad20sep99_1_resources/9901879.jpg)

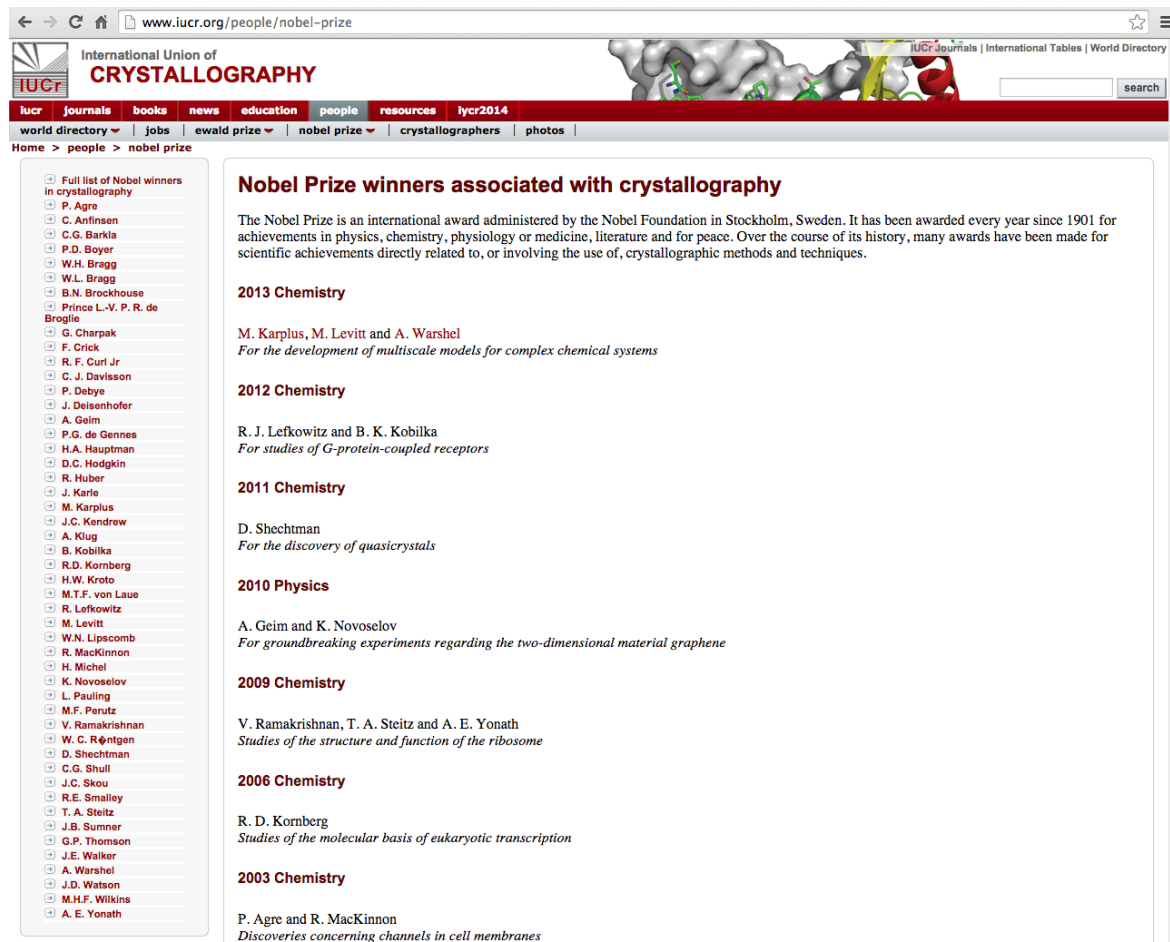
# Macromolecules can also form crystals

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



# Caveats

- Getting macromolecules to form crystals can be hard
  - Crystallographers have sometimes worked for decades to get good crystals of a particular protein



The screenshot shows the IUCr website page titled "Nobel Prize winners associated with crystallography". The page features a navigation menu at the top with categories like "iucr", "journals", "books", "news", "education", "people", "resources", and "iucr2014". Below the menu, there is a search bar and a list of navigation links including "world directory", "jobs", "ewald prize", "nobel prize", "crystallographers", and "photos". The main content area is divided into two columns. The left column contains a list of Nobel winners in crystallography, including names like P. Agre, C. Anfinsen, C.G. Barkla, P.D. Boyer, W.H. Bragg, W.L. Bragg, B.N. Brockhouse, Prince L.-V. P. R. de Broglie, G. Charpak, F. Crick, R. F. Curl Jr, C. J. Davison, P. Debye, J. Deisenhofer, A. Geim, P.G. de Gennes, H.A. Hauptman, D.C. Hodgkin, R. Huber, J. Karle, M. Karplus, J.C. Kendrew, A. Klug, B. Kobilka, R.D. Kornberg, H.W. Kroto, M.T.F. von Laue, R. Lefkowitz, M. Levitt, W.N. Lipscomb, R. MacKinnon, H. Michel, K. Novoselov, L. Pauling, M.F. Perutz, V. Ramakrishnan, W. C. Röntgen, D. Shechtman, C.G. Shull, J.C. Skou, R.E. Smalley, T. A. Steltz, J.B. Sumner, G.P. Thomson, J.E. Walker, A. Warshel, J.D. Watson, M.H.F. Wilkins, and A. E. Yonath. The right column contains the main text, which includes a general introduction to the Nobel Prize and a list of winners from 2003 to 2013, each with their names and a brief description of their work.

International Union of  
**CRYSTALLOGRAPHY**

www.iucr.org/people/nobel-prize

Home > people > nobel prize

### Nobel Prize winners associated with crystallography

The Nobel Prize is an international award administered by the Nobel Foundation in Stockholm, Sweden. It has been awarded every year since 1901 for achievements in physics, chemistry, physiology or medicine, literature and for peace. Over the course of its history, many awards have been made for scientific achievements directly related to, or involving the use of, crystallographic methods and techniques.

#### 2013 Chemistry

M. Karplus, M. Levitt and A. Warshel  
*For the development of multiscale models for complex chemical systems*

#### 2012 Chemistry

R. J. Lefkowitz and B. K. Kobilka  
*For studies of G-protein-coupled receptors*

#### 2011 Chemistry

D. Shechtman  
*For the discovery of quasicrystals*

#### 2010 Physics

A. Geim and K. Novoselov  
*For groundbreaking experiments regarding the two-dimensional material graphene*

#### 2009 Chemistry

V. Ramakrishnan, T. A. Steitz and A. E. Yonath  
*Studies of the structure and function of the ribosome*

#### 2006 Chemistry

R. D. Kornberg  
*Studies of the molecular basis of eukaryotic transcription*

#### 2003 Chemistry

P. Agre and R. MacKinnon  
*Discoveries concerning channels in cell membranes*

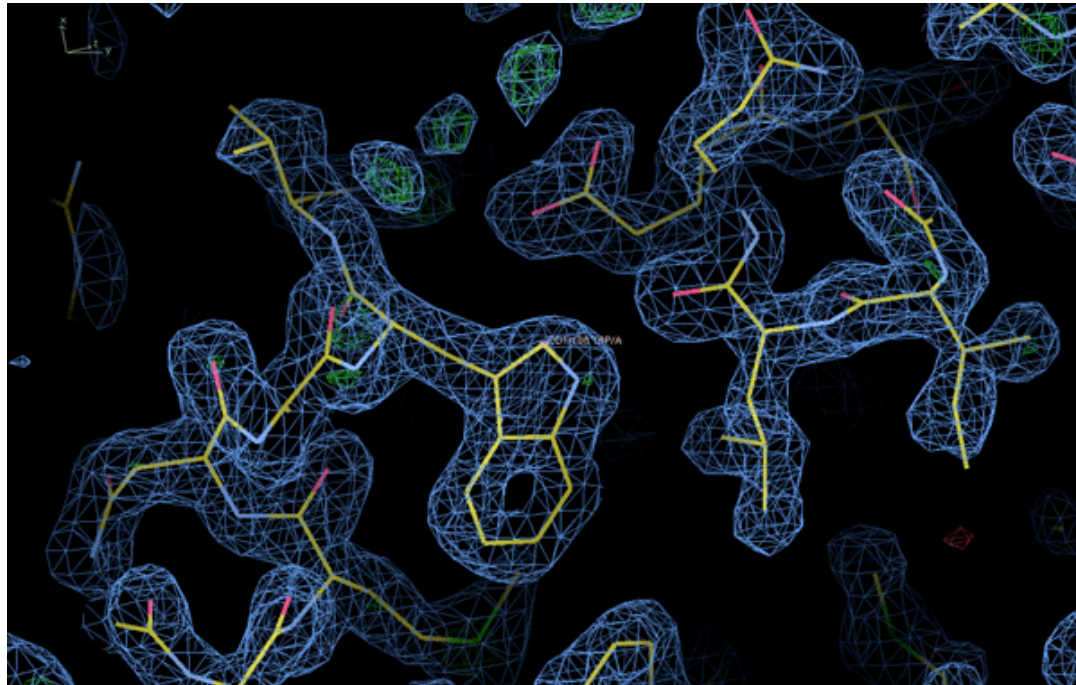
# Caveats

- Sometimes a molecule will adopt a different structure in a crystal than it does in its natural environment
- Crystallography gives you a static snapshot of a molecule's structure
  - Usually (but not always) this snapshot corresponds to the molecule'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

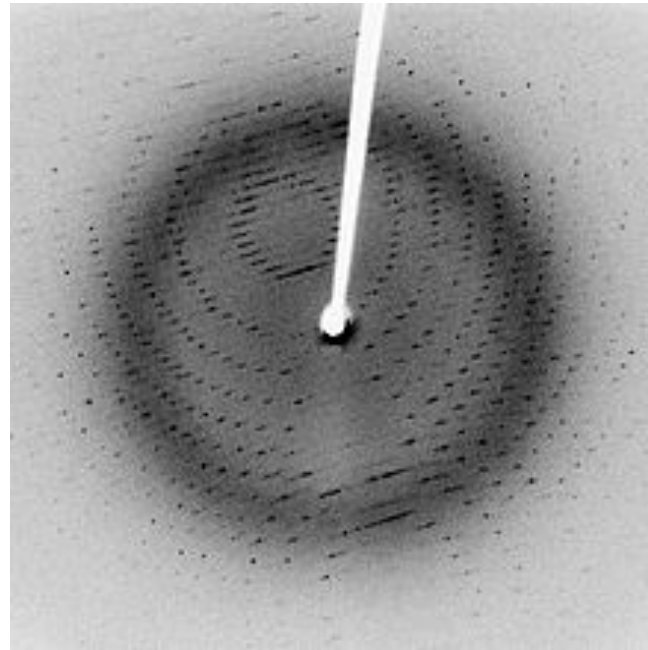
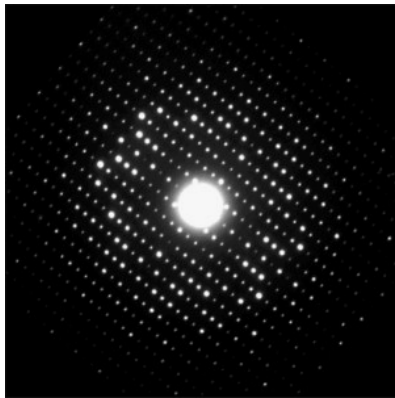


[http://www.lynceantech.com/images/electron\\_density\\_map.png](http://www.lynceantech.com/images/electron_density_map.png)

# Diffraction patterns

# Diffraction patterns

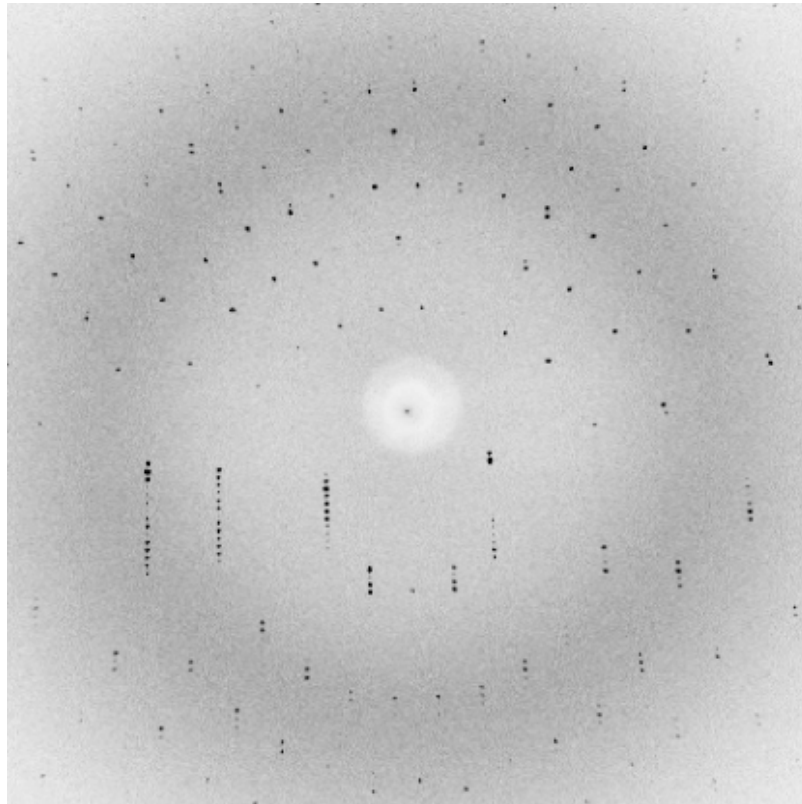
- When you shine a light beam through a crystal, you get a distinctive pattern of bright spots called a diffraction pattern



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

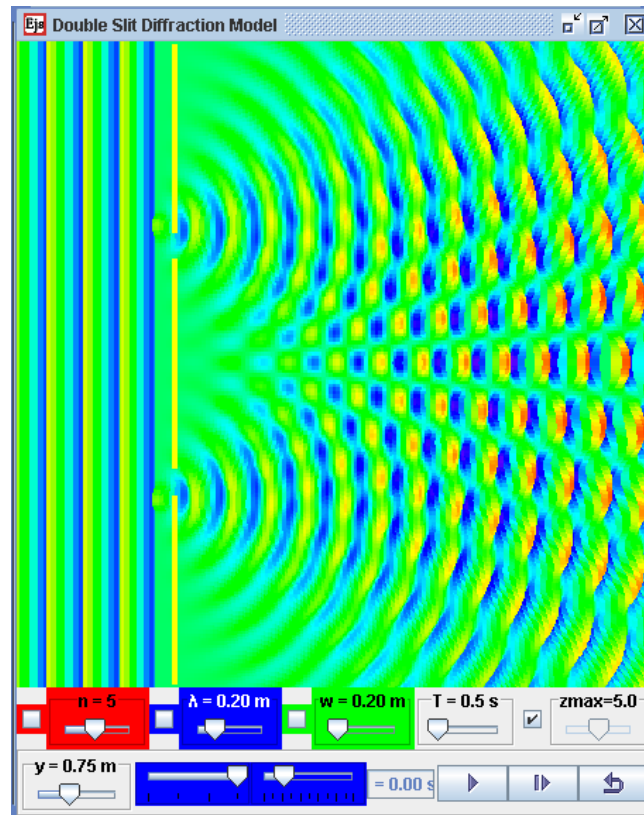
# Diffraction patterns

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



# What causes diffraction patterns?

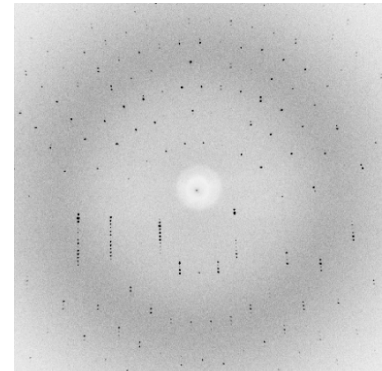
- Short answer: interference of light
  - The bright spots are places where light interferes constructively. Elsewhere it tends to interfere destructively (cancel out).



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

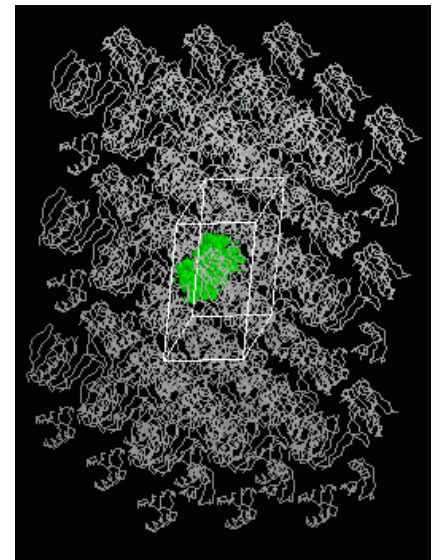
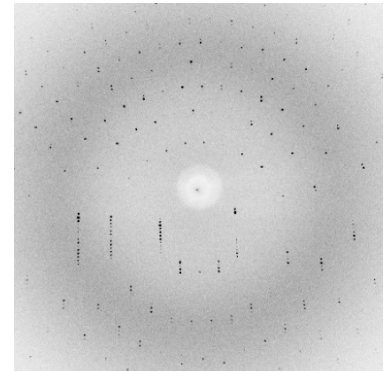
# Relationship between diffraction pattern and electron density

- 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
  - 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 width of the lattice an integer number of times in each dimension—i.e., 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.



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 won't correspond to reasonable 3D 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
  - Covalently bonded atoms will be near one another

# 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

# 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
    - Search for a structure that minimizes the sum of two terms: (1) 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

# 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 sees 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  $R_{\text{free}}$  to quantify the error in the prediction)

# Computational methods continue to improve

- Although the phasing problem is decades old, researchers are still inventing better solutions

nature

Vol 464 | 22 April 2010 | doi:10.1038/nature08892

## Super-resolution biomolecular crystallography with low-resolution data

Gunnar F. Schröder<sup>1,2</sup>, Michael Levitt<sup>2</sup> & Axel T. Brunger<sup>2,3,4,5,6</sup>

NOVEMBER 2013 | NATURE METHODS

## Improved low-resolution crystallographic refinement with Phenix and Rosetta

Frank DiMaio<sup>1,6</sup>, Nathaniel Echols<sup>2,6</sup>, Jeffrey J Headd<sup>2</sup>, Thomas C Terwilliger<sup>3</sup>, Paul D Adams<sup>2,4</sup> & David Baker<sup>1,5</sup>

Machine learning approaches  
(see paper on course website)

# Serial crystallography with XFEL

- Using an extremely bright x-ray beam, one could detect diffraction patterns from extremely small crystals or even single particles (e.g., a virus)
- This is now possible thanks to x-ray free electron lasers (XFELs), pioneered at Stanford/SLAC
- Challenge:
  - The particle (or tiny crystal) disintegrates when the laser beam hits it
  - One can only capture an image from a single angle from each crystal or particle
  - 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 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