Electron Crystallography Apr 5, 2017

Iban Ubarretxena Department of Structural and Chemical Biology Icahn School of Medicine at Mount Sinai *iban.ubarretxena@mssm.edu*

Electron crystallography of biological macromolecules (2007). Glaeser et al. Oxford university press

Molecular Cryo-EM specimens



Ribosome

Hepatitis B Virus

Actin-Myosin

Aquaporin

Specimens and Resolution

Electron tomography

Arbitrary samples, as long as they are thin enough for the electrons (whole cells, asymmetric viruses, etc)

- Rather complicated sample preparation (usually)
- ▶Data collection can be quite tedious
- ▶ Resolution achieved so far ~20Å for ideal specimens (after motif averaging)



Single particles (large molecules or complexes)
No crystallization required, >200 KDa limit
Relatively simple sample preparation (after purification!)
Highly homogeneous sample required, ~ 1mg/ml, ~100 µl per batch of cryo grids
Resolution achieved so far ~2.5-3Å for icosahedral particles and ~ 2-3Å for non-symmetrical samples



- Helical arrangements/crystals
- Filaments or tubes required
- Commonly found in viruses, cytoskeletal proteins, or some spontaneously formed tubes
- Component molecules can be small
- Rather complicated image processing, but 3D volumes can be calculated from single image
 Resolution achieved so far ~3Å for ideal samples (bacterial flagella, TMV, acetylcholine receptor)



- 2D crystals
- Crystals needed (obviously!)
- Some proteins show tendency to form single-layered crystals (common for membrane proteins)
- Component molecules can be small
- Image processing is quite involved but well documented.
- ▶ Resolution achieved so far ~2Å for ideal samples (aquaporins)



Examples of structures solved by EC



2D crystals



Electron crystallography of 2D crystals



2D crystals under the microscope



ow to prepare EM specimens of membrane crystals

Negative staining

- Embedding a specimen in a layer of heavy metal salts, such as uranyl acetate, phosphotungstic acid, and ammonium molybdate.
- Provides high contrast for imaging
- Very quick and easy procedure
- (1) 2.5µl sample solution is adsorbed to a carboncoated grid (made hydrophilic by a glow discharge)
- 2 blot the grid with filter paper
- (③ wash with several drops of water)
- ④ stain with two drops of stain
- ⑤ blot the grid with filter paper and completely dry



Nobuhiko Gyobu

Negative staining

- Screening of crystallization conditions. The information on the morphology and quality of the specimen. Detection of crystalline arrays.
- Crystallographic study at 2-3 nm resolution. Rough estimate of the molecular surface, shape and the packing arrangement.
- Staining and drying results in distortions of the molecules. Incomplete stain embedding gives artifacts.



Atomic models of biological macromolecules by cryo-electron microscopy

Protein	Year	Sample Preparation Embedding Medium		
Bacteriorhodopsin	1990	2D crystals	Glucose	
Plant light-harvesting complex (LHC-II)	1994	2D crystals	Tannin	
α,β-tubulin	1998	2D crystals	Tannin-glucose	
Aquaporin-1	2000	2D crystals	Trehalose	
Acetylcholine receptor	2003	helical crystals	Ice	
Bacterial flagellar filament	2003	helical crystals	Ice	
Aquaporin-0	2004	2D crystals	Glucose, Trehalose	
Aquaporin-4	2005	2D crystals	Trehalose	
Microsomal Glutathione transferase 1 (MGST1)	2006	2D crystals	Trehalose	
Microsomal prostaglandin E synthase 1 (MPGES1)	2008	2D crystals	Trehalose	

Preparing flat specimens

Lack of specimen flatness is caused by:

•the roughness of the carbon support film

•the wrinkling of carbon film supported by EM grid upon cooling (cryo-crinkling)

Solutions:

 Use of molybdenum grids (the thermal expansion coefficient is similar to that of carbon)

·Use of flat carbon support films by spark-less evaporation

High-quality carbon support film



Spark-less evaporation

Evaporation with sparkling

Back injection method



Carbon sandwich method

Gyobu et al. J. Struct. Biol. (2004) 146, 325



2D crystals under the microscope



Image processing of 2D crystals at different stages

After threefold crystallographic averaging and replacement of image amplitudes by diffraction amplitudes

CTF corrected but amplitudes can be improved

After merging of several images to increase SNR

Result after averaging of unit cells by unbending in real space and filtering in reciprocal space Area of digitized micrograph of a BR 2D crystal in which only electron noise is visible (plus hair)

Courtesy of Richard Henderson

Single particles



Crystal



Why bother with 2D crystalline specimens?

Biological macromolecules are sensitive to electron radiation. A crystal has lots of molecules related by an easily determined relation, thus allowing collection of data with less radiation used (higher S/N).

To preserve high resolution details and minimize the radiation damage, specimen images are recorded at low dose (5-20 e^{-/A^2}).

There are caveats!

•To be useful, crystals have to be large to have as many protein subunits as possible; homogeneous

•The order in the crystals has to be as perfect as possible

•There has to be only one layer.

Averaging is key for resolution



one unit cell



projection map from 1 image (100s of unit cells)



3d from several images (10,000s of unit cells)

Why are e.g. SPA and 2DX far superior in resolution to ET?

specimen damage

dose

AVERAGING!

> symmetrized class sum

single image

unsymmetrized class sum

Definition of symmetry?

- Something is symmetric when it is invariant (i.e., does not change) under some transformation
- For material objects, there is a myriad of possibilities: a cylinder is symmetric under arbitrary rotations about its axis, a sphere under any rotation, etc.
- There are continuous symmetries, and discrete, such as in polyhedra, a crystal, helical object, etc.
- A symmetry group is the set of all symmetry operations applicable to an object

Molecular Symmetry (applies to SPEM, and 2DX)



Tetrahedral (4) symmetry Octahedral (8) symmetry Icosahedral (20) symmetry







Virus

Hsp16.5

A crystal is an array of translationally repeating units

A 2D crystal is generated by translation of a unit cell along a linear combination of two fundamental, linearly independent vectors (a and b).



Courtesy of Michael Landsberg

Symmetry operations: Rotation

n-fold rotational symmetry dictates that rotation about a point by an angle of $360^{\circ}/n$ generates an image indistinguishable from the original





Symmetry operations: Reflection



- aka mirror or bilateral symmetry
- Any two points perpendicular to and equidistant from the axis (in 3D, or a line in 2D) of reflection are identical

Symmetry operations: Glide reflection



Translation by $\frac{1}{2}$ unit cell combined with a reflection about the axis of translation

Crystallographic Symmetry

- A crystallographic space group is the mathematical group of symmetry operations which apply to <u>both</u> the given unit cell and the crystal array
- There are 230 possible crystallographic space groups in 3D (65 for proteins and chiral molecules)
- For 2D projection maps, there are 17 plane groups
- These are different (but correlate somewhat trivially) to the 17 2D space groups which describe all possible 2D crystal arrangements

Primitive and centered cells

15 of the 17 plane groups are primitive cells, the remaining 2 are centered cells

- A primitive cell is a minimal region repeated by lattice translations (15/17)
- A face-centered cell is larger than the alternative primitive cell, and has internal repetition



There are 17 plane groups



Plane groups and 2D space groups

Plane group	Unit cell geometry	Highest order	Point group	Glide/screw	2d space group
	(crystal system)	rotation			
p1	rhomboid (oblique)	1	1	Ν	P1
p2	rhomboid (oblique)	2	2	Ν	P2
pm	rectangle	1	m	Ν	P12
pg	rectangle	1	m	Y	P12 ₁
cm	rectangle	1	m	Ν	C12
p2mm	rectangle	2	2mm	Ν	P222
p2mg	rectangle	2	2mm	Y	P222 ₁
p2gg	rectangle	2	2mm	Y	P22 ₁ 2 ₁
c2mm	rectangle	2	2mm	Ν	C222
p4	square	4	4	Ν	P4
p4mm	square	4	4mm	Ν	P422
p4gm	square	4	4mm	Y	P42 ₁ 2
р3	rhombus (hexagonal)	3	3	Ν	P3
p3m1	rhombus (hexagonal)	3	3m	Ν	P321
p31m	rhombus (hexagonal)	3	3m	Ν	P312
p6	rhombus (hexagonal)	6	6	Ν	P6
p6mm	rhombus (hexagonal)	6	6mm	Ν	P622

How can we take advantage of the symmetry?

In single particles, we average the parts that are equivalent due to the symmetry of the problem, which translates into a better signal to noise ratio.

In helical aggregates and crystals, the Fourier transform shows the symmetry of the problem, with enhanced features visible in the power spectra.

"DIFFRACTION"!!

What is the image of a 2D crystal according to linear image theory?



Representation of a 2D crystal.

-) Representation as the convolution of the object with the lattice.
- The FT of a has points on a reciprocal lattice ,which is related to the image lattice. The position of the spots in the transform is determined by the image lattice. Whereas the Amplitude and phase values depends on the structure of the repeating motif in the image.
 - Miller indices (h,k) are used to refer to a particular lattice point in the transform

... the image is the convolution of 3 factors:

- 1) molecular structure
- 2) the lattice structure
- 3) the CTF

Image processing is about:

- 1) increase SNR form molecular structure
- 2) correction of lattice disorder
- 3) correction for CTF modulation of diffracted waves.

Analysis of the Image into Fourier Components

Aim: To obtain reliable measures of the amplitude and phase of the frequency components.

Fourier transform Fourier synthesis

Original Image

Extraction and correction of Fourier components

Fourier transforms of images of crystalline specimens

- Each diffraction spot represents a different (spatial) frequency
- The amplitude of each spot is unique to the structure of the object
- When a Fourier transform of a crystal image is calculated with a computer, one also gets the phase at each spatial frequency



The Fourier transform plays a central role in understanding the analysis of diffraction data

The electron density intensity $\rho_c(\vec{R})$ everywhere in a protein crystal, can be represented as a sum of cosines.

This Fourier series illustrates the point in 1D:

$$\rho(x) = \sum_{n=0}^{N} F(n) \cos\left\{-2\pi \frac{n}{L}x + \alpha(n)\right\}$$

Each cosine must have its own amplitude F(n), phase $\alpha(n)$ and periodicity L.



. 1/s Frequency tells you about image spacings

Amplitude tells you "how much" of a frequency component is present

Phase tells you "where" the frequency components are located in the image

3 cosine functions are characterized by frequencies, amplitudes and phases



Real-space 1D image as a superposition of 3 cosine functions





3 Fourier components

3 cosine functions are characterized by different frequencies





Chiu et al. Biophysical Journal (1993) vol. 64 (5) pp. 1610-25

The Fourier transform plays a central role in understanding the analysis of diffraction data

The electron density intensity $\rho_c(\vec{R})$ everywhere in a protein crystal, can be represented as a sum of cosines.

This Fourier series illustrates the point in 1D:

$$\rho(x) = \sum_{n=0}^{N} F(n) \cos\left\{-2\pi \frac{n}{L}x + \alpha(n)\right\}$$

Each cosine must have its own amplitude F(n), phase $\alpha(n)$ and periodicity L.

$$\rho_{c}(\vec{R}) = \sum_{h,k,l=0} F(\vec{g}_{h,k,l}) \cos\{-2\pi \vec{g}_{h,k,l} \cdot \vec{R} + \alpha(\vec{g}_{h,k,l})\}$$

$$e^{i\theta} = \cos\theta + i\sin\theta$$



Resolution and spatial frequency



FIGURE 1 An illustration of a one-dimensional object potential function v(x) at the top of the figure which can be decomposed into three cosine waves with different frequencies (s), amplitudes (A), and phases (α) with respect to a common origin. The mathematical expression is $v(x) = \sum A_i(s) \cos (2\pi S_i x + \alpha_i)$. (provided by A. Avila-Sakar and V. Mootha)

RESOLUTION, "d", AND SPATIAL FREQUENCY, "s = 1/d"

ARE "THE SAME THING"

The Fourier transform plays a central role in understanding the analysis of diffraction data (\vec{r})

The electron density intensity $\rho\left(\vec{R}\right)$ everywhere in a protein crystal, can be represented as a sum of cosines.

$$\rho(\vec{R}) = \sum_{h,k,l=0} F(\vec{g}_{h,k,l}) e^{i\alpha(\vec{g}_{h,k,l})} e^{-i2\pi(\vec{g}_{h,k,l})\cdot\vec{R}}$$

$$\rho(\vec{R}) = \int F(\vec{S}) e^{i\alpha(\vec{S})} e^{-i2\pi\vec{S}\cdot\vec{R}} d\vec{S}$$
amplitude phase spatial frequency

 $F(\vec{S}) = F(\vec{S})e^{i\alpha(\vec{S})}$ is the structure factor

S is the 3D spatial frequency vector, which replaces the discrete vector g_{hkl}

$$\rho(\vec{R}) = \int \mathbf{F}(\vec{S}) e^{-i2\pi\vec{S}\cdot\vec{R}} d\vec{S}$$

In other words, the electron density $\rho(\vec{R})$ is the (inverse) Fourier transform of the structure factor **Experimentally we need to measure** $F(\vec{S}) = F(\vec{S})e^{i\alpha(\vec{S})}$ by measuring amplitude and phase at each discrete spatial frequency

Convolution of one unit cell with a 2-D lattice produces a 2-D crystal



The electron density function of a crystal, $\rho_c(\vec{R})$ can be described in terms of two separate functions $\rho_u(\vec{R})$ and $L(\vec{R})$. $\rho_u(\vec{R})$ is the electron density function of the unit cell

 $L(\vec{R})$ is the lattice function that marks the position of every unit cell

$$\rho_{c}\left(\vec{R}\right) = \rho_{u}\left(\vec{R}\right) \otimes L\left(\vec{R}\right) = \int \rho_{u}\left(\vec{R}'\right) L\left(\vec{R} - \vec{R}'\right) d\vec{R}'$$

This integral is difficult to solve in real space!!!

Convolution: the convolution of an integral expresses the amount of overlap of one function f_2 as it is shifted by another function f_1 , is denoted $f_1 * f_2$ and defined over an infinite range as:





$$\rho_{c}\left(\vec{R}\right) = \rho_{u}\left(\vec{R}\right) \otimes L\left(\vec{R}\right) = \int \rho_{u}\left(\vec{R}'\right) L\left(\vec{R} - \vec{R}'\right) d\vec{R}'$$

But much easier to solve in Fourier or reciprocal space!!!

$$FT\left[\rho_{u}\left(\vec{R}\right)\otimes L\left(\vec{R}\right)\right]=FT\left[\rho_{u}\left(\vec{R}\right)\right]\cdot FT\left[L\left(\vec{R}\right)\right]$$



 $L(\vec{R}) = \sum_{j} \delta(\vec{R} - \vec{R}_{j})$ is a sum of Dirac delta functions, one at every lattice point

... and its FT is another sum of Dirac delta functions but this time in reciprocal space

$$FT\left[L\left(\vec{R}\right)\right] = FT\left[\sum_{j}\delta(\vec{R}-\vec{R}_{j})\right] = \sum_{h,k,l}\delta(\vec{S}-\vec{g}_{h,k,l})$$



is the reciprocal lattice of the crystal lattice



$$\mathbf{FT}\left[L\left(\vec{R}\right)\right] = FT\left[\sum_{j}\delta(\vec{R}-\vec{R}_{j})\right] = \sum_{h,k,l}\delta(\vec{S}-\vec{g}_{h,k,l})$$

The points in the reciprocal lattice are determined by reciprocal lattice vectors **a***, **b*** and **c***, which are themselves determined by the crystal lattice (unit cell) vectors.

$$a^* = \frac{b \times c}{a \cdot b \times c}$$
 $b^* = \frac{c \times a}{a \cdot b \times c}$ $c^* = \frac{a \times b}{a \cdot b \times c}$

Then the vector to an arbitrary reciprocal lattice point in Fourier space can be written in terms of reciprocal lattice vectors \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* and of Miller indices.

$$\vec{g}_{h,k,l} = ha^* + kb^* + lc^*$$

The Miller indices (hkl) provide a unique identification for every point in the reciprocal lattice.

THUS they serve to identify every diffraction spot in a diffraction pattern

$$FT\left[\rho_{u}\left(\vec{R}\right)\otimes L\left(\vec{R}\right)\right]=FT\left[\rho_{u}\left(\vec{R}\right)\right]\cdot FT\left[L\left(\vec{R}\right)\right]$$

$$FT\left[\rho_{c}\left(\vec{R}\right)\right] = FT\left[\rho_{u}\left(\vec{R}\right) \otimes L\left(\vec{R}\right)\right] = \sum_{h,k,l} F(g_{hkl})\delta(S - g_{hkl})$$

In other words, the Fourier Transform of a crystal is ZERO everywhere except at the reciprocal lattice points g_{hkl} , where it has the same value as the Fourier Transform of the unit cell F(S) would have at that spatial frequency (S).

What is the image of a 2D crystal according to linear image theory?



Representation of a 2D crystal.

-) Representation as the convolution of the object with the lattice.
- The FT of a has points on a reciprocal lattice ,which is related to the image lattice. The position of the spots in the transform is determined by the image lattice. Whereas the Amplitude and phase values depends on the structure of the repeating motif in the image.
 - Miller indices (h,k) are used to refer to a particular lattice point in the transform

... the image is the convolution of 3 factors:

- 1) molecular structure
- 2) the lattice structure
- 3) the CTF

Image processing is about:

- 1) increase SNR form molecular structure
- 2) correction of lattice disorder
- 3) correction for CTF modulation of diffracted waves.

Analysis of the Image into Fourier Components

Aim: To obtain reliable measures of the amplitude and phase of the frequency components.

Fourier transform Fourier synthesis

Original Image

Extraction and correction of Fourier components

Amplitude and phases



Fourier space

Amplitude and phases



Imaging for phases and/or for amplitudes



Figure 1.1. Formation of an image and diffraction pattern by a lens from an infinite periodic object. For parallel, monochromatic illumination, the transmitted and diffracted beams are focused into spots, F_0 , F_1 , F_1' etc., which form the diffraction pattern in the back focal-plane of the lens at a distance *f*, the focal length, from the lens. Only three diffraction spots are shown for clarity. The diffracted beams recombine to form a magnified image in the image plane. Note that the image is inverted with respect to the object, but the diffraction pattern is not. In effect, the screen of the electron microscope is in the plane I_1-I_3 in the imaging mode and in the plane F_1-F_1' in the diffraction mode.

Imaging for phases and/or for amplitudes



Diffraction amplitudes are better than image amplitudes

We can collect diffraction data directly where the CTF modulation is minimal: Amplitudes are good but phases are lost



The phases of the structure factors are lost when diffraction intensities are recorded.

$$\rho(\vec{R}) = \int F(\vec{S}) e^{-i2\pi \vec{S} \cdot \vec{R}} d\vec{S}$$
$$F(\vec{S}) = F(\vec{S}) e^{i\alpha(\vec{S})}$$

$$\Psi_{scattered} \propto F(\vec{S})e^{i\alpha(\vec{S})}$$

 $I(\vec{S}) = \Psi \cdot \Psi^*$

$$\Psi_{sca} \propto F(\vec{S})e^{i\alpha(\vec{S})} \cdot F(\vec{S})e^{-i\alpha(\vec{S})} = F^2(\vec{S})$$

Image processing of 2D crystals at different stages

After threefold crystallographic averaging and replacement of image amplitudes by diffraction amplitudes

CTF corrected but amplitudes can be improved

After merging of several images to increase SNR

Result after averaging of unit cells by unbending in real space and filtering in reciprocal space Area of digitized micrograph of a BR 2D crystal in which only electron noise is visible (plus hair)

Courtesy of Richard Henderson

UNBENDING



Anchi Cheng

correlations

 $f * f = f^{+}(-t) \otimes f(t)$

 $f * g = f^{+}(-t) \otimes g(t)$



MERGING



Stahlberg

MERGING



Fig. 7. Graphical comparison of all phases determined for six spots with resolution beyond 4.7 A. Phases are plotted as vectors on a polar diagram with the length of each vector being proportional to $1/IQ^2$. Thus the strongest spots show up as longer vectors. The result of the summation of all the vectors is also shown as a vector, but this is plotted at 1/4 of its true length. The direction of this **graphically**. vector is our best estimate of the phase of the Fourier component in the structure, and its length can be used to provide an estimate of the error.

Stahlberg

3D structure analysis requires views from multiple directions

Projection theorem: When a 3-D object is projected to produce a 2-D image, its Fourier transform is a 2-D slice (NOT a projection) of the 3-D transform of the oject

These 2-D slices always pass through the origin and thus are called"central sections" When data from many different "central sections" are combined one builds up the full, 3-D Fourier transform



XBL 925-4723

The reciprocal lattice of a 2D crystal consists of lattice lines





Accumulating merged images with tilts



Accumulating merged images with tilts

low tilt

high tilt and various tilt axes





Interpolation along Lattice Lines

To apply the 3D FT¹, lattice lines have to be uniformly sampled.



Examples of Lattice Lines



Fig. 1. Lattice line data. Plots of amplitudes (lower panels) and phases (upper panels) along the z^* axis for three selected reflections. The fitted lattice lines were produced by weighted least squares fitting and the resulting errors are shown.

Sampling of Fourier Space after combination of all central sections

