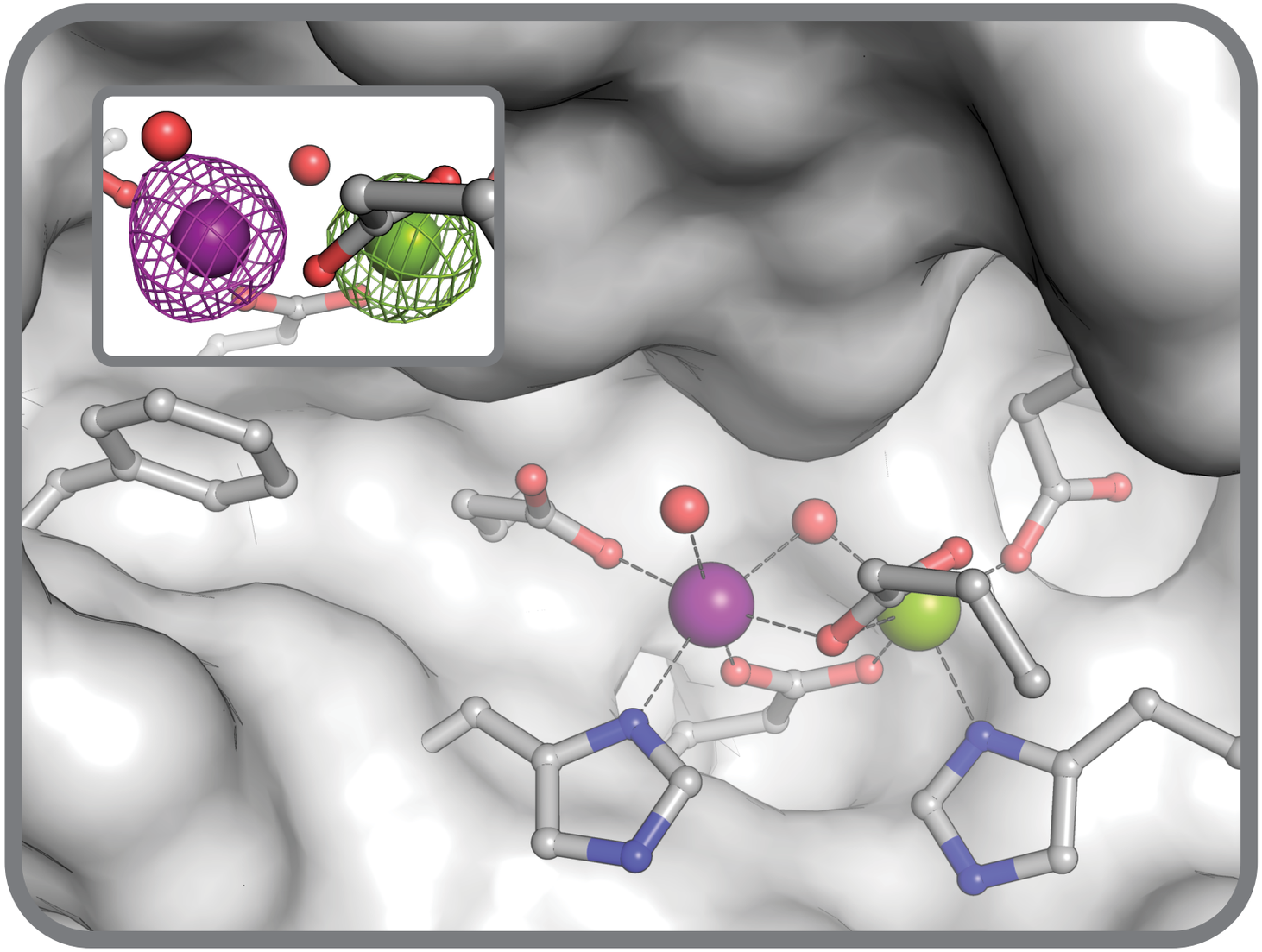
X-ray crystallography



Amie Boal
Penn State University
Oliver Einsle
University of Freiburg

May 31, 2014
2014 Penn State Bioinorganic Workshop

How do we visualize macromolecules?

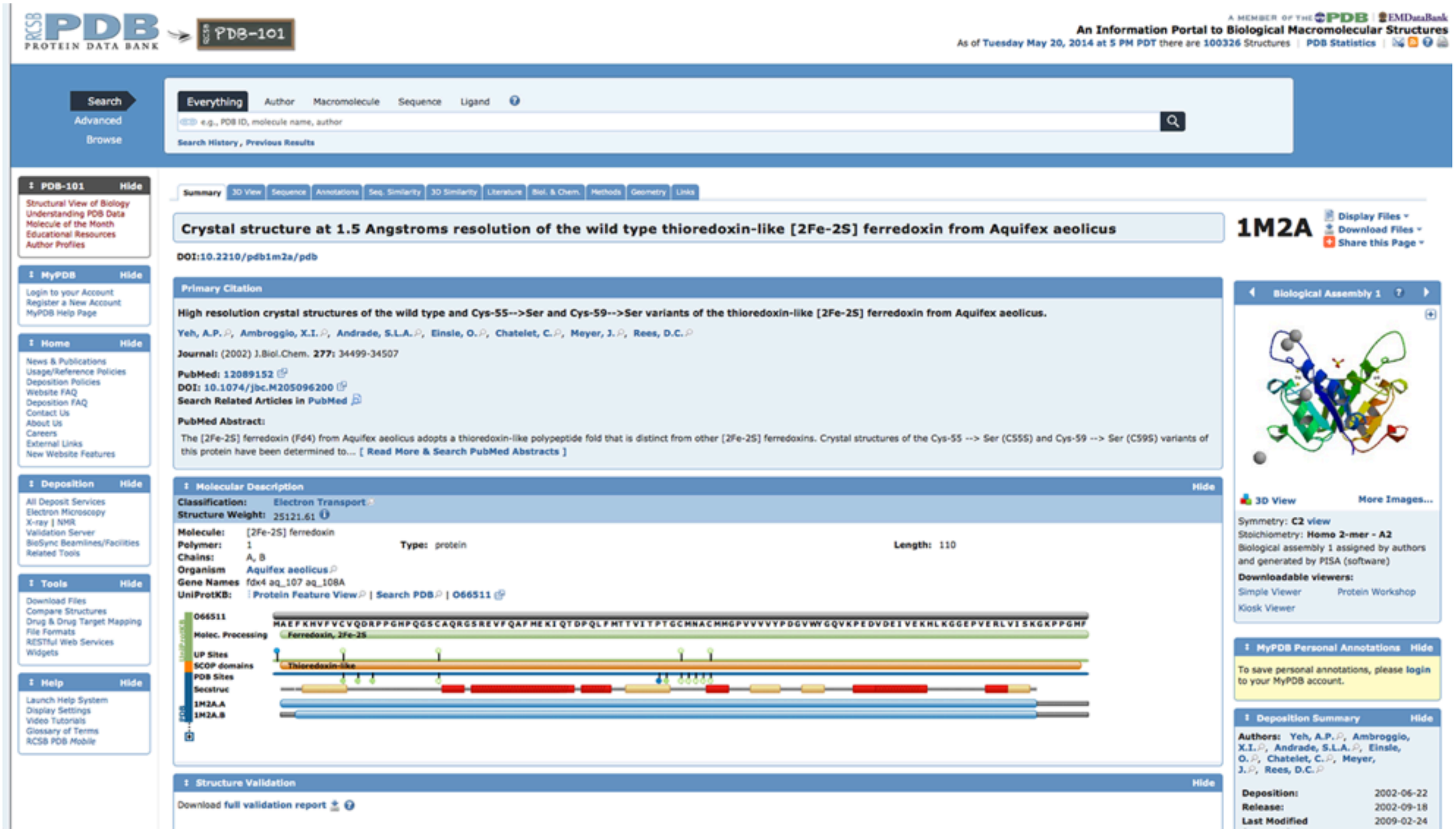
- X-ray crystallography
- NMR spectroscopy
- Electron microscopy
- Small-angle X-ray scattering
- Circular dichroism, analytical ultracentrifugation

Resolution, size range, ease of use

X-ray crystallography

- Useful with a large range of molecular sizes
- High-resolution information possible
- Ongoing effort to improve user accessibility
- Many available tools to view and assess results
  - Molecular models
  - Electron density maps

Protein Data Bank



Protein Data Bank

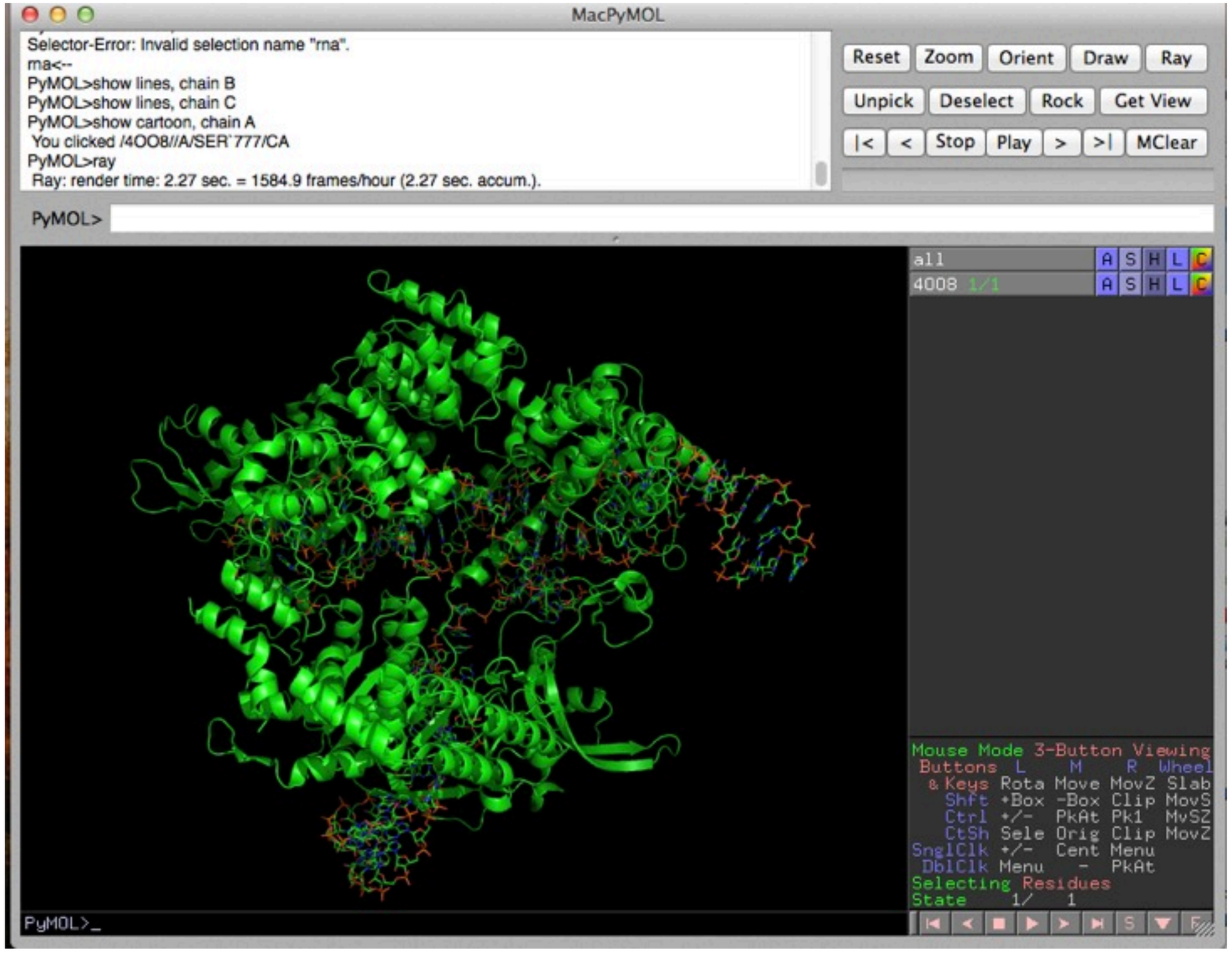
```
ATOM      2  CA  ALA A   1       8.154  -3.613  48.765  1.00 84.41           C
ATOM      3  C   ALA A   1       8.846  -3.642  47.425  1.00 84.01           C
ATOM      4  O   ALA A   1       9.909  -4.243  47.280  1.00 86.46           O
ATOM      5  CB  ALA A   1       6.821  -2.853  48.659  1.00 85.53           C
ATOM      6  N   GLU A   2       8.269  -2.965  46.448  1.00 80.47           N
ATOM      7  CA  GLU A   2       8.865  -2.949  45.122  1.00 74.28           C
ATOM      8  C   GLU A   2       9.176  -1.507  44.702  1.00 64.00           C
ATOM      9  O   GLU A   2      10.119  -0.904  45.208  1.00 72.07           O
ATOM     10  CB  GLU A   2       7.900  -3.631  44.156  1.00 80.94           C
ATOM     11  CG  GLU A   2       6.701  -4.229  44.880  1.00 85.35           C
ATOM     12  CD  GLU A   2       5.427  -3.401  44.735  1.00 88.86           C
ATOM     13  OE1 GLU A   2       5.478  -2.199  44.371  1.00 91.34           O
ATOM     14  OE2 GLU A   2       4.357  -3.977  45.001  1.00 91.79           O
ATOM     15  N   PHE A   3       8.400  -0.962  43.779  1.00 35.90           N
ATOM     16  CA  PHE A   3       8.551   0.438  43.321  1.00 19.11           C
ATOM     17  C   PHE A   3       7.248   0.816  42.655  1.00 19.44           C
ATOM     18  O   PHE A   3       6.737   0.028  41.843  1.00 23.14           O
ATOM     19  CB  PHE A   3       9.620   0.549  42.257  1.00 26.89           C
ATOM     20  CG  PHE A   3       9.663   1.875  41.590  1.00 36.11           C
ATOM     21  CD1 PHE A   3      10.017   3.026  42.299  1.00 33.73           C
ATOM     22  CD2 PHE A   3       9.341   1.976  40.234  1.00 36.63           C
ATOM     23  CE1 PHE A   3      10.037   4.269  41.654  1.00 36.70           C
ATOM     24  CE2 PHE A   3       9.366   3.223  39.586  1.00 36.98           C
ATOM     25  CZ  PHE A   3       9.697   4.352  40.276  1.00 28.43           C
ATOM     26  N   LYS A   4       6.683   1.977  43.023  1.00 16.63           N
ATOM     27  CA  LYS A   4       5.483   2.481  42.374  1.00 12.41           C
ATOM     28  C   LYS A   4       5.799   3.924  42.003  1.00 13.56           C
ATOM     29  O   LYS A   4       6.515   4.626  42.674  1.00 14.35           O
ATOM     30  CB  LYS A   4       4.244   2.433  43.261  1.00 16.86           C
ATOM     31  CG  LYS A   4       3.743   0.964  43.429  1.00 20.15           C
ATOM     32  CD  LYS A   4       2.576   0.856  44.350  1.00 21.42           C
ATOM     33  CE  LYS A   4       2.339  -0.576  44.706  1.00 35.30           C
ATOM     34  NZ  LYS A   4       2.149  -1.299  43.403  1.00 38.95           N
ATOM     35  N   HIS A   5       5.218   4.321  40.862  1.00 12.52           N
ATOM     36  CA  HIS A   5       5.430   5.647  40.353  1.00  9.21           C
ATOM     37  C   HIS A   5       4.027   6.216  40.137  1.00  8.86           C
```

PDB accession code 1M2A

Model viewing

- PyMOL Molecular Graphics System
  - www.pymol.org
- Chimera
  - www.cgl.ucsf.edu/chimera/
- Coot
  - www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/
- Other resources
  - www.rcsb.org/pdb/static.do?p=software/software_links/molecular_graphics.html

Model viewing



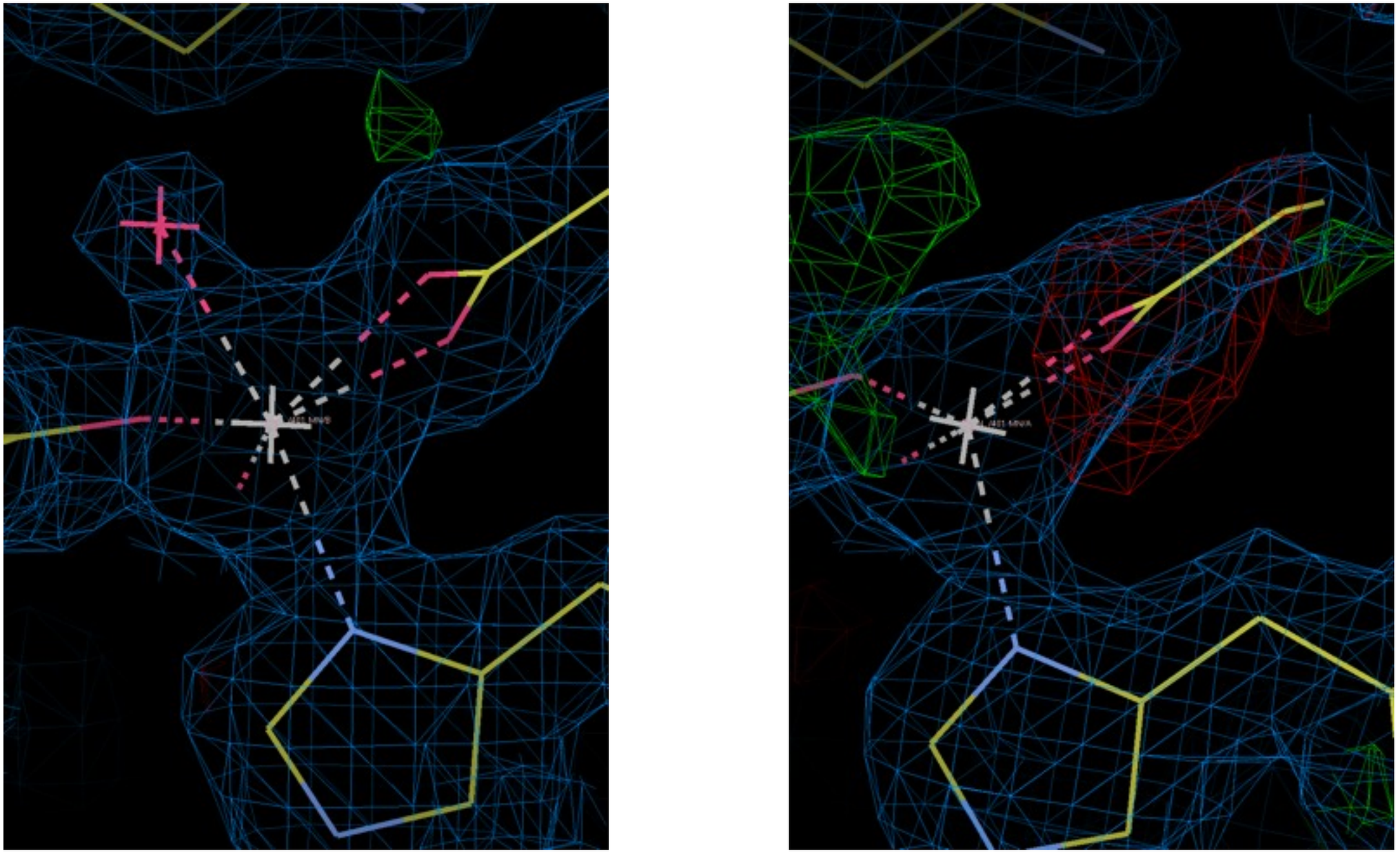
Limitations

- Static snapshot of dynamic molecules
- Model is an average of all molecules in crystal lattice
- Obtained under specific conditions - possibly far from biological relevance
- Structure can be influenced by crystallization and data collection process
- Resulting model is an interpretation of data

Electron density map viewing

- Electron density server
  - eds.bmc.uu.se/eds/
- Swiss PDB viewer
  - http://spdbv.vital-it.ch/
- Coot
  - www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/

Map viewing



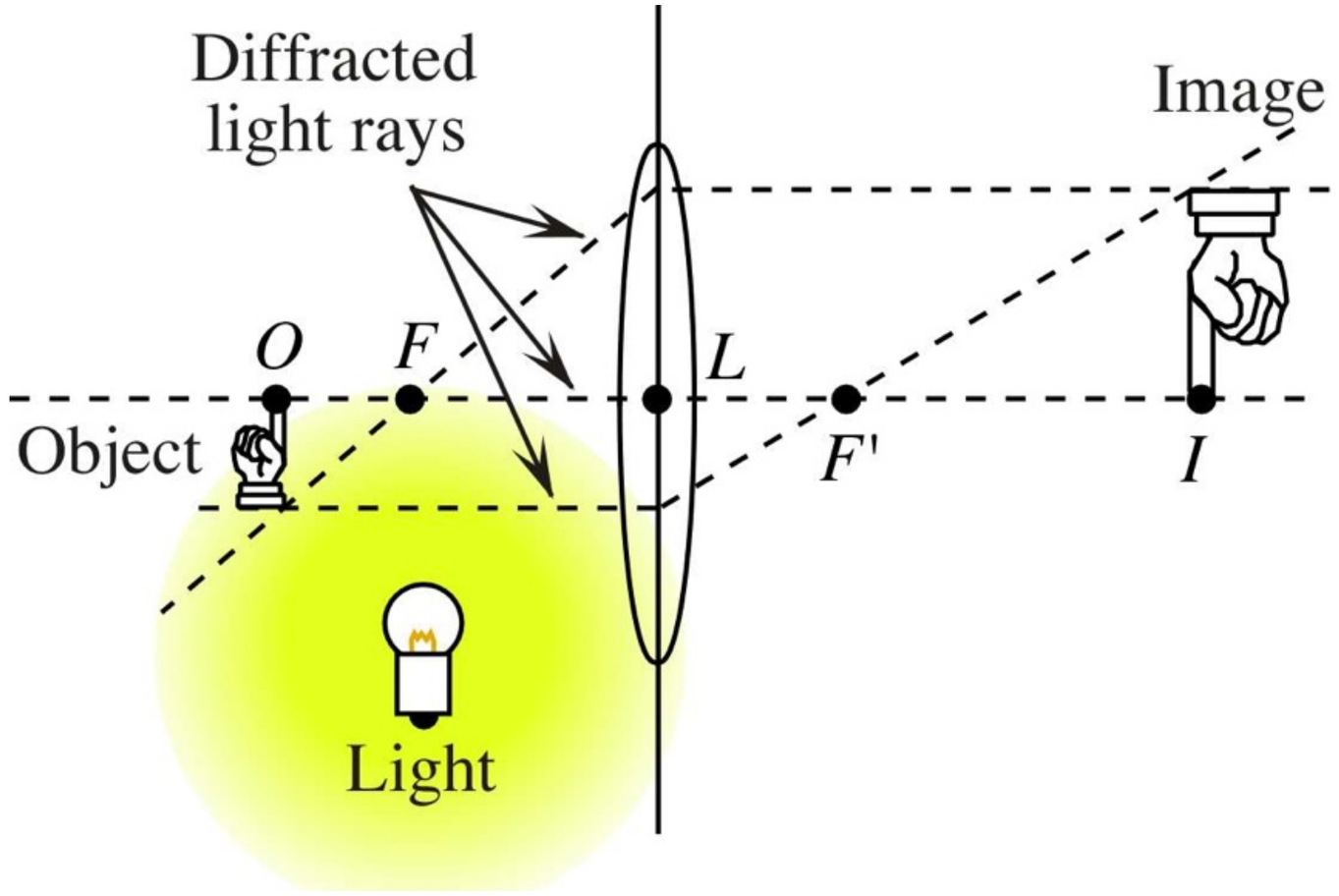
PDB accession code 4DR0

Metalloprotein crystallography

- Starting point for mechanistic/computational study
- Inspiration for synthetic models
- Models for other biophysical techniques
- Location, identity, stoichiometry of metallocofactors
- Geometry and identity of ligands
- Structures of complex metallocofactors

Sommerhalter, M., Lieberman, R.L., and Rosenzweig, A.C. (2005) *Inorg. Chem. 44*, 770-778.

Microscopy analogy



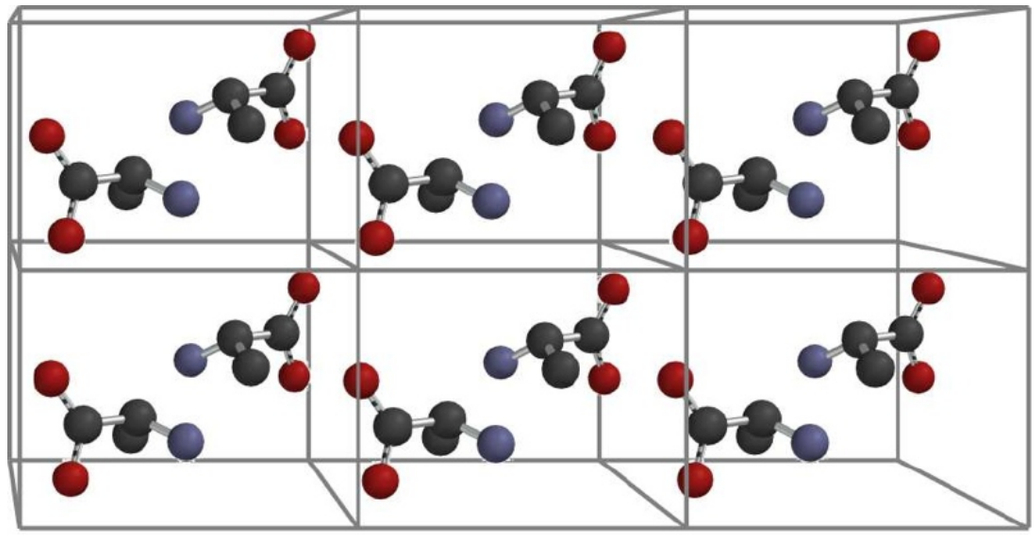
Wavelength of light rays must be appropriate for object size

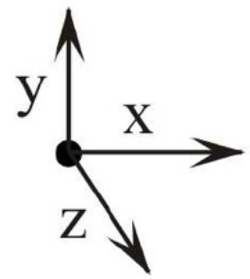
X-rays are used to visualize atomic-level structure in macromolecules

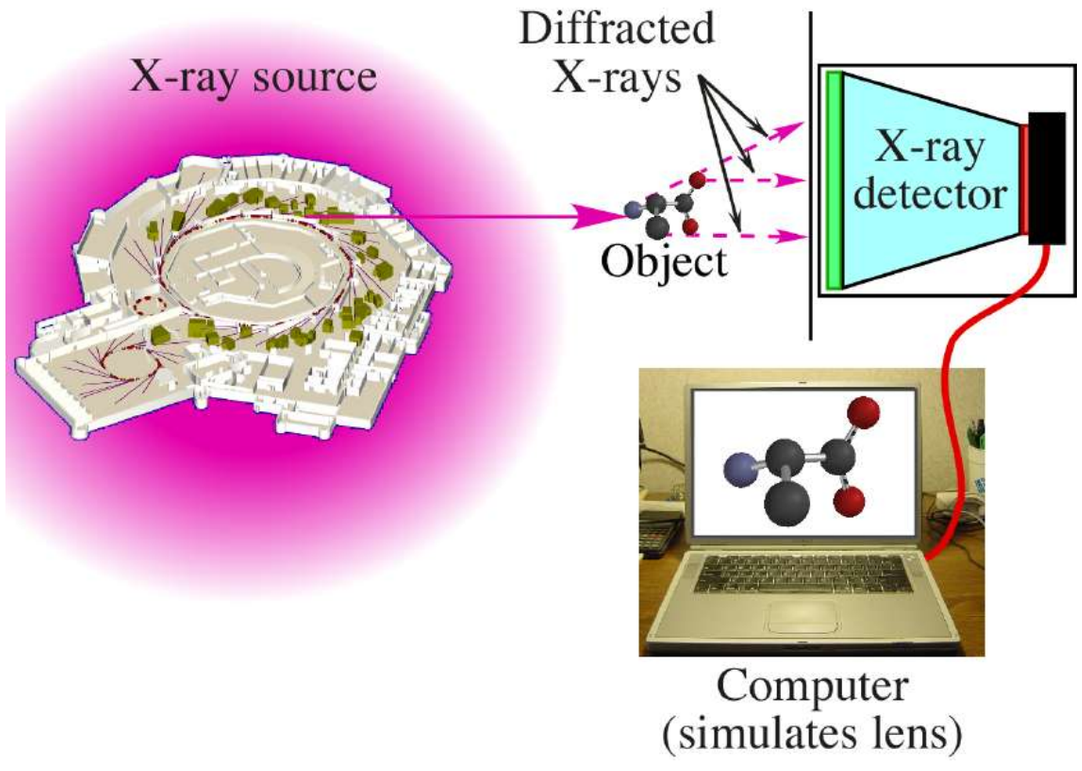
Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

1. Single molecules diffract X-rays weakly

Crystal lattice amplifies signal







2. X-rays can't be focused by lenses

Structure solution facilitated by mathematical methods

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Crystallography workflow

- Biological sample production
- Crystallization
- X-ray diffraction data collection
- Structure solution/phase determination
- Model building and validation

Other resources

- Workshops and courses
  - APS CCP4 crystal school - www.ccp4.ac.uk/schools/APS-school/
  - NSLS RapidData - www.bnl.gov/rapidata/content/announcement.asp
  - CSH course - meetings.cshl.edu/courses/2014/c-crys14.shtml

Other resources

- Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed.
- Rupp, B. (2010) Biomolecular Crystallography
- Metalloprotein crystallography
    - Sommerhalter, M., Lieberman, R.L., and Rosenzweig, A.C. (2005) *Inorg. Chem. 44*, 770-778.

Crystallography workflow

- Biological sample production
- Crystallization
- X-ray diffraction data collection
- Structure solution/phase determination
- Model building and validation

Biological sample production

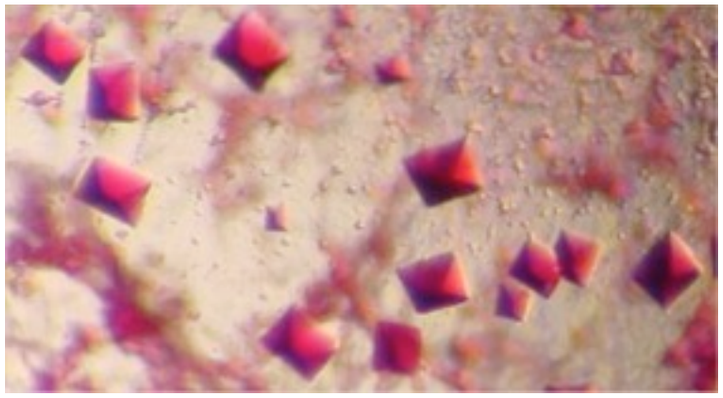
- Construct/sequence selection
- Source organism
- Metallocofactor incorporation
- Affinity tags and purification scheme

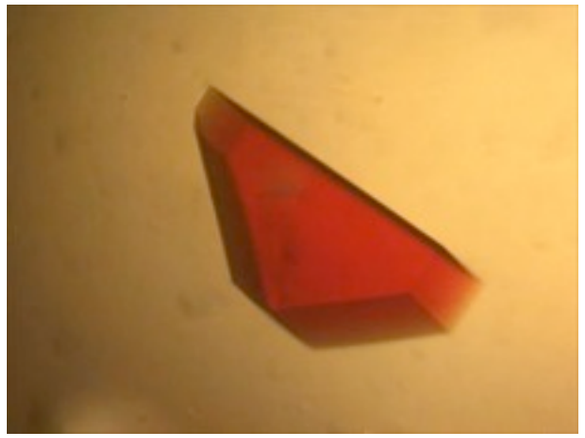
Must assess purity, polydispersity, activity

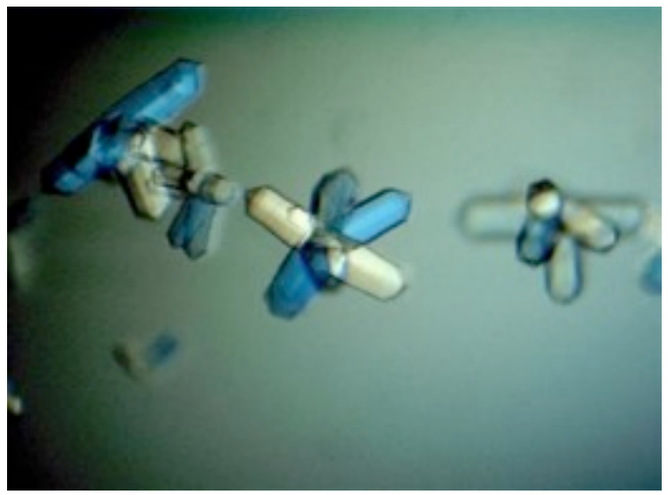
Crystallography workflow

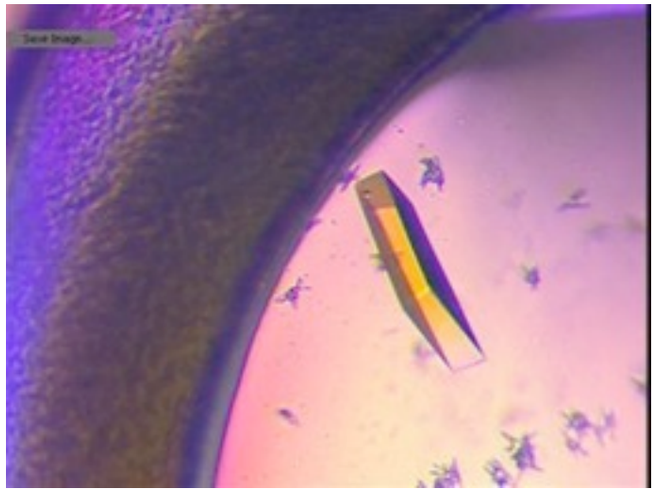
- Biological sample production
- Crystallization
- X-ray diffraction data collection
- Structure solution/phase determination
- Model building and validation

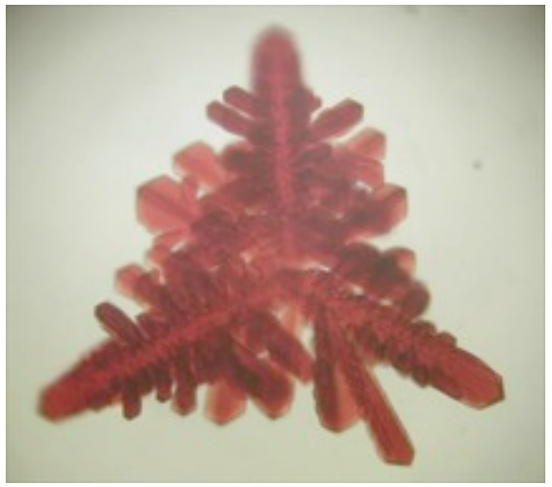
Macromolecular crystals







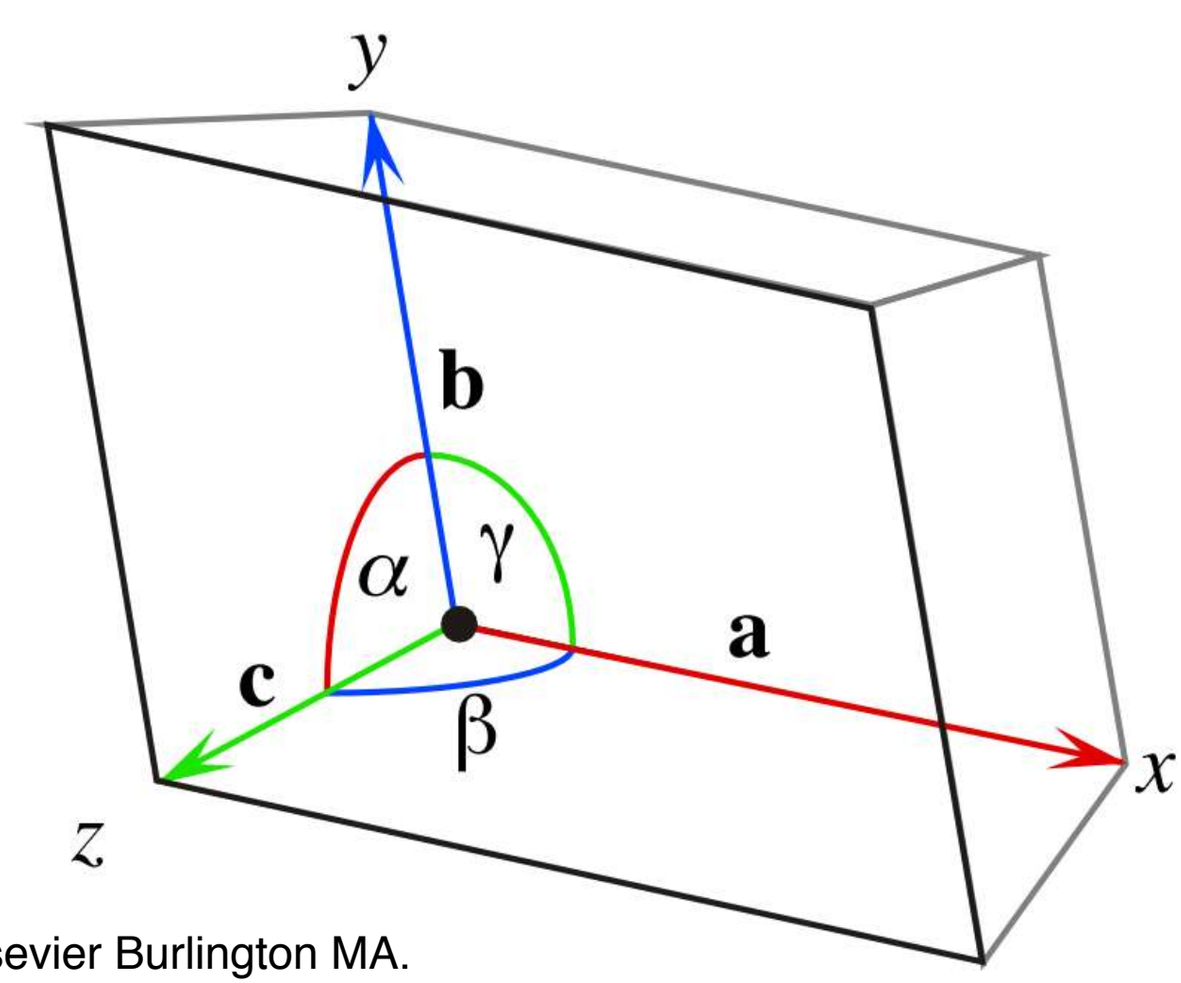




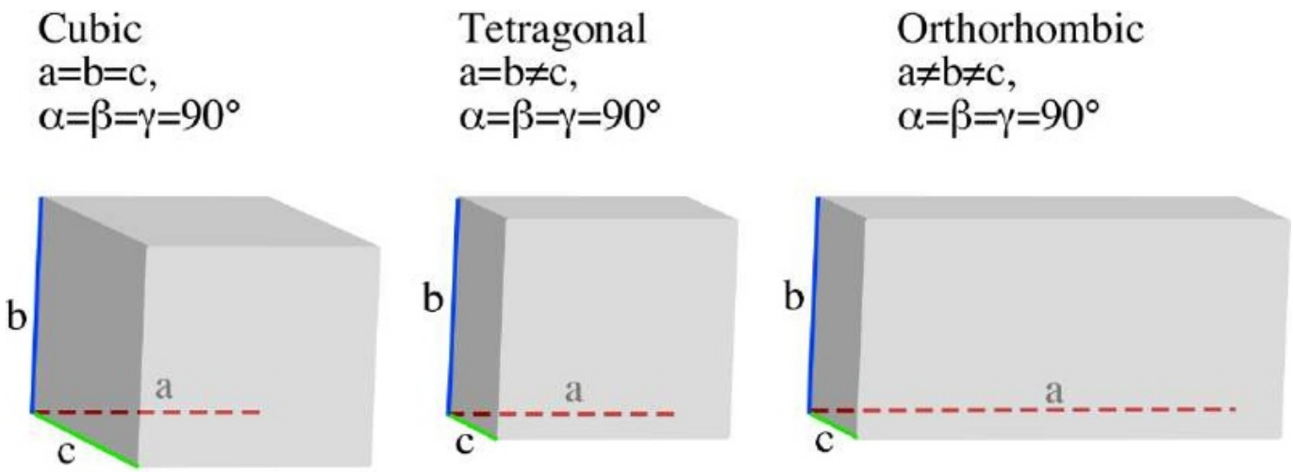
Hampton Research

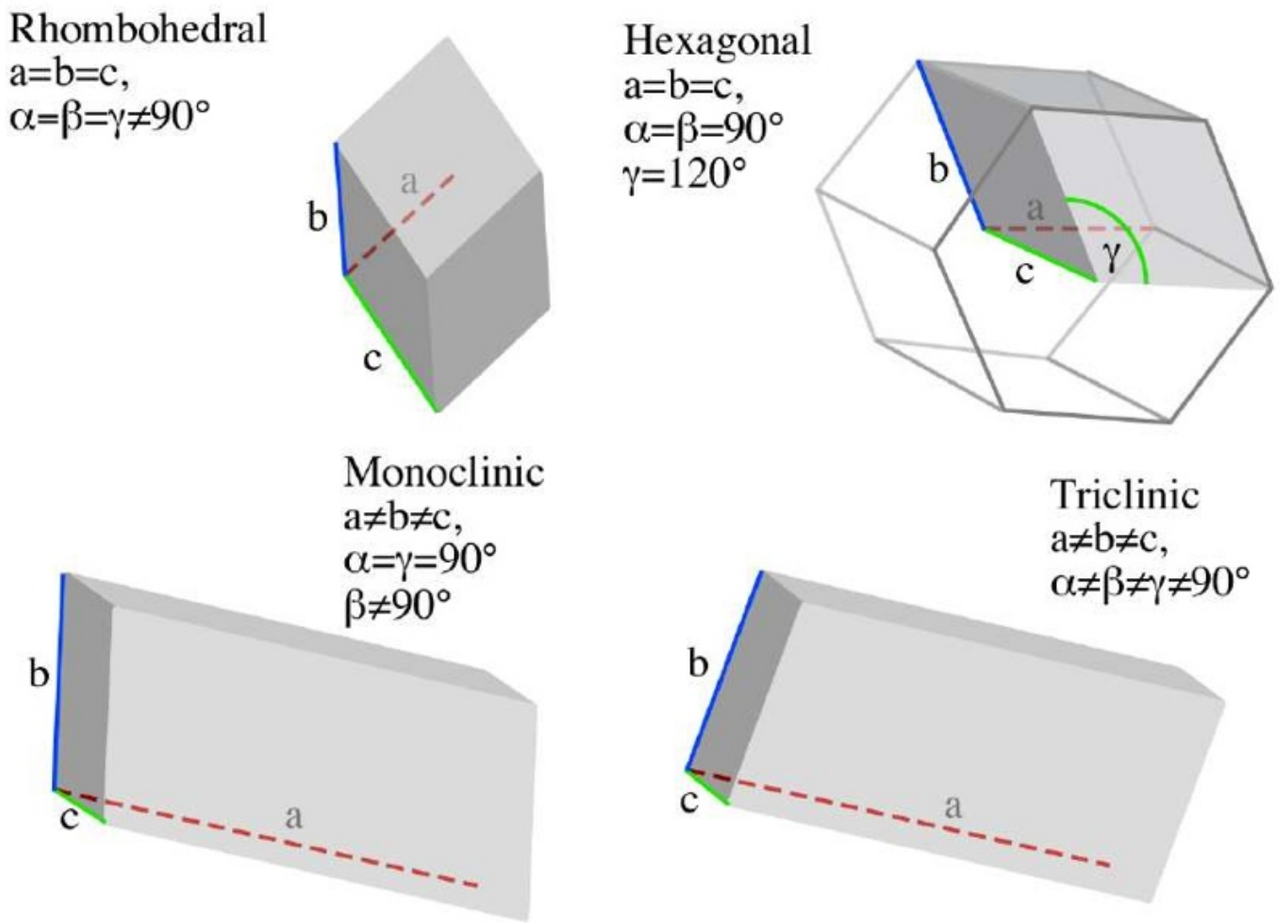
Crystallization

- A macromolecular crystal is an ordered 3D array of molecules held together by weak interactions
- Made up of unit cells
- Cells are the repeating unit
- Unit cells have a defined shape
- x, y, z coordinate system



Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.





14 different lattice systems
- Bravais lattice

Defined by unit cell shape

a = b = c relationships are identities

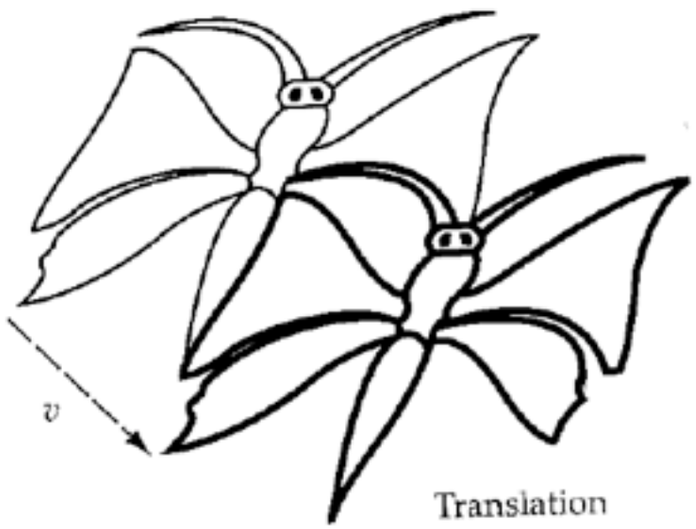
If a = b, unit cell contents along axes are identical

Can have internal symmetry
- multiple copies of protein in unit cell
- related by symmetry operators

Final model only describes structure of asymmetric unit

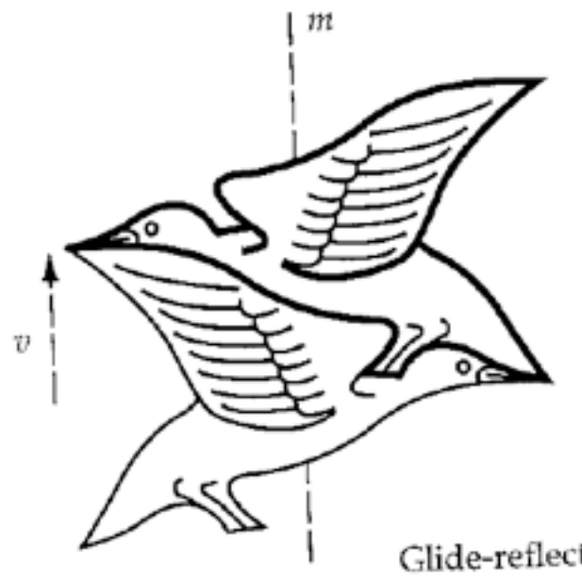
Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Symmetry operators



Translation

Reflection



Glide-reflection

Rotation

Schattschneider, D. (2004) M.C. Escher: Visions of Symmetry, 2nd Ed. University of Michigan Press.

Space groups

- 230 possible space groups
- 65 for chiral molecules
- Designated by symbols that describe lattice type and symmetry operators
  - $P2_12_12_1$
  - $C2$
- Must assign space group correctly to solve structure

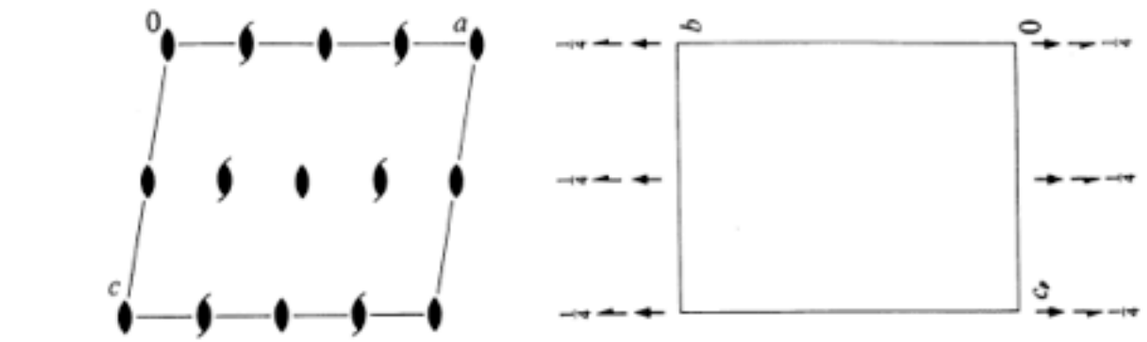
Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

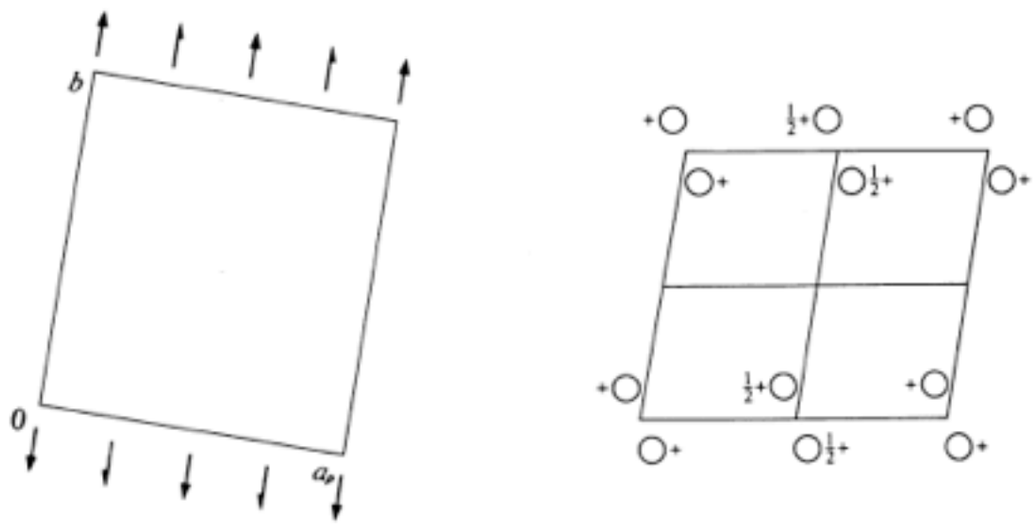
Wukovitz, S., and Yeates, T. (1995) *Nat. Struct. Biol. 2*, 1062-1067.

C2 space group

$C2$	C_2^3	2	Monoclinic
No. 5	$C121$		Patterson symmetry $C12/m1$

UNIQUE AXIS b, CELL CHOICE 1





Origin on 2

Asymmetric unit $0 \le x \le \frac{1}{2}$; $0 \le y \le \frac{1}{2}$; $0 \le z \le 1$

Symmetry operations

For $(0,0,0)+$ set

(1) 1 (2) 2 0,y,0

For $(\frac{1}{2},\frac{1}{2},0)+$ set

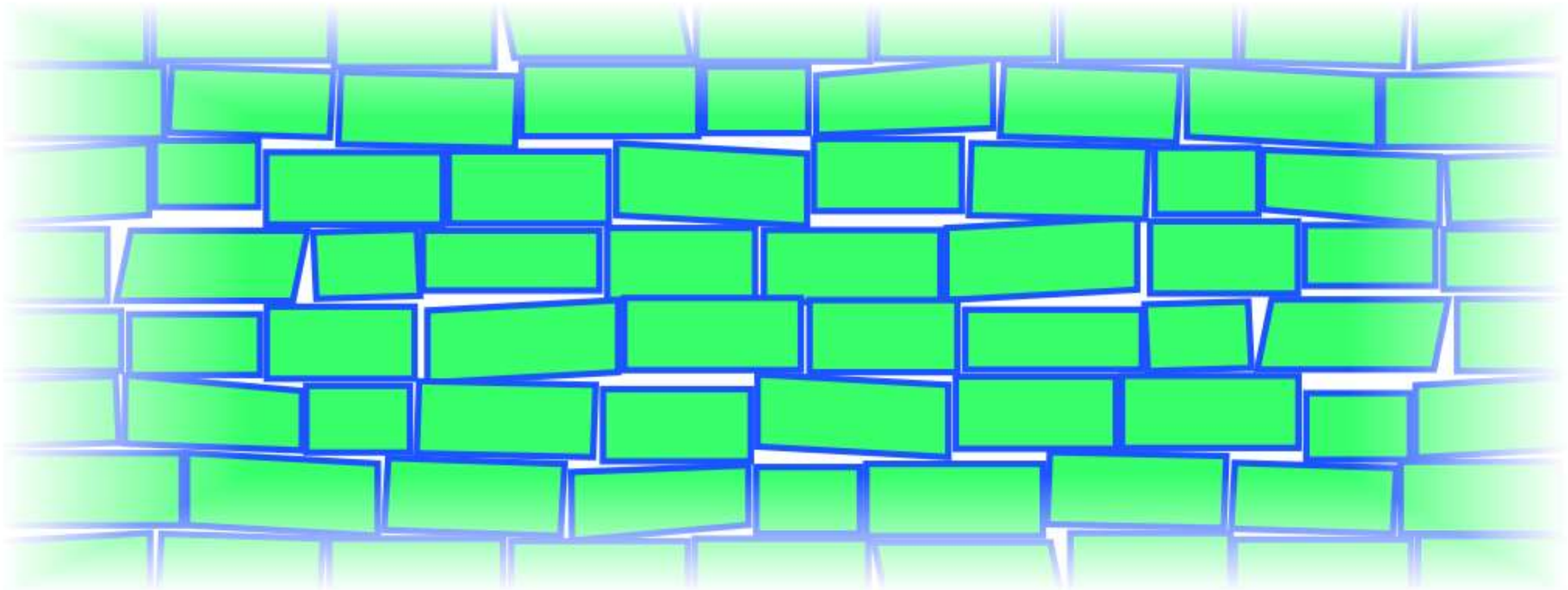
(1) $t(\frac{1}{2},\frac{1}{2},0)$ (2) $2(0,\frac{1}{2},0)$ $\frac{1}{4}$,y,0

Positions

Multiplicity, Wyckoff letter, Site symmetry		Coordinates	
		$(0,0,0)+$	$(\frac{1}{2},\frac{1}{2},0)+$
4 c 1	(1) x,y,z	(2) $\bar{x},y,\bar{z}$	

International Union of Crystallography (2006) Vol. A, pp. 124-131.

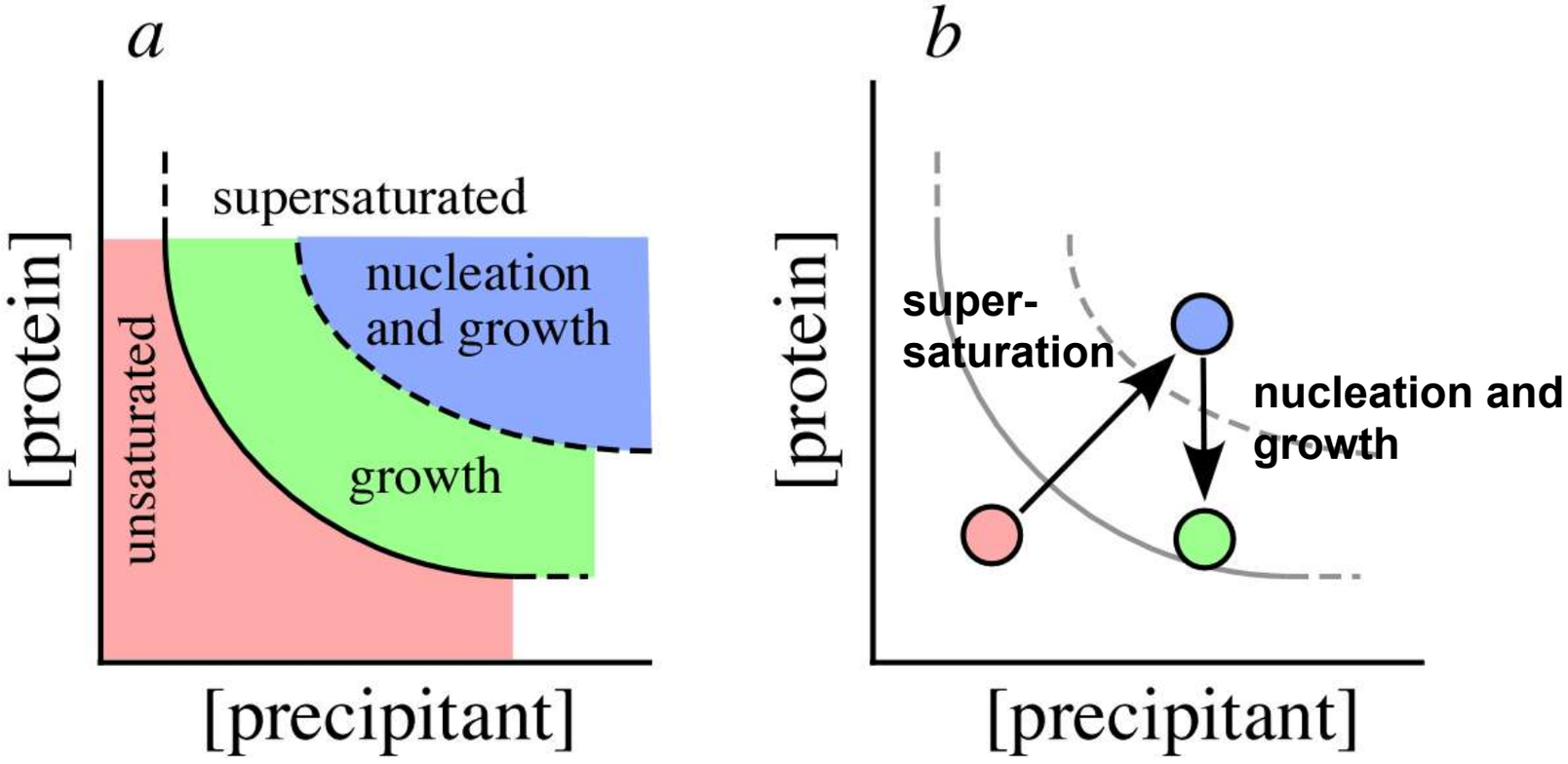
Crystallization



- Lattice is array of unit cells - an imperfect one!
- Crystallographic experiment provides an image of average electron density in the asymmetric unit

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Crystallization



- Supersaturation, nucleation, growth
- Slow controlled precipitation from solution without unfolding the molecule

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Precipitants

- Salts (ammonium sulfate)
- Polyethylene glycol
- Small organic compounds
- Additives, pH, temperature important

Disrupt biomolecule hydration layer

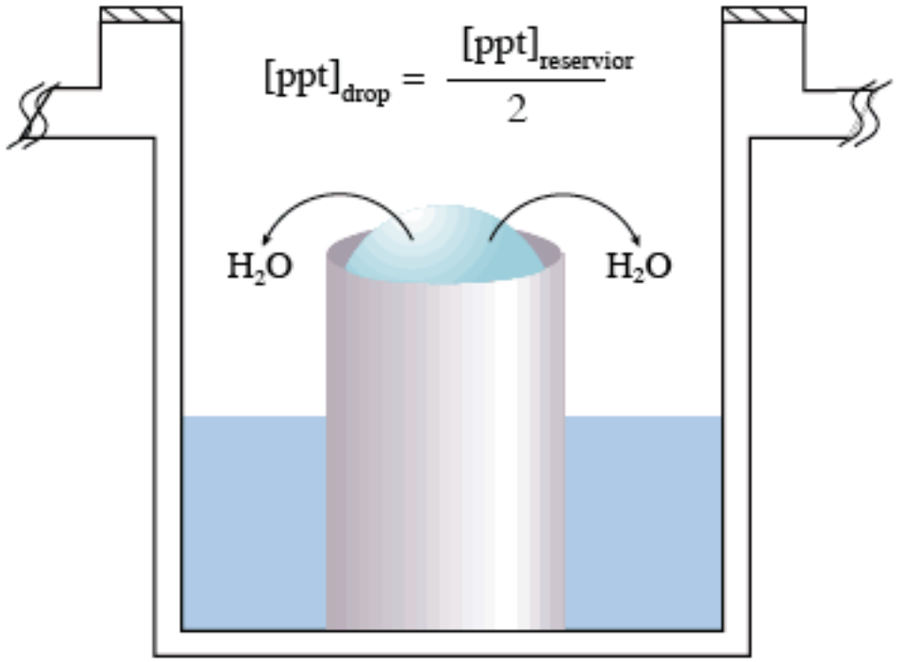
Force self-association without denaturation

Can affect structure

Crystallization

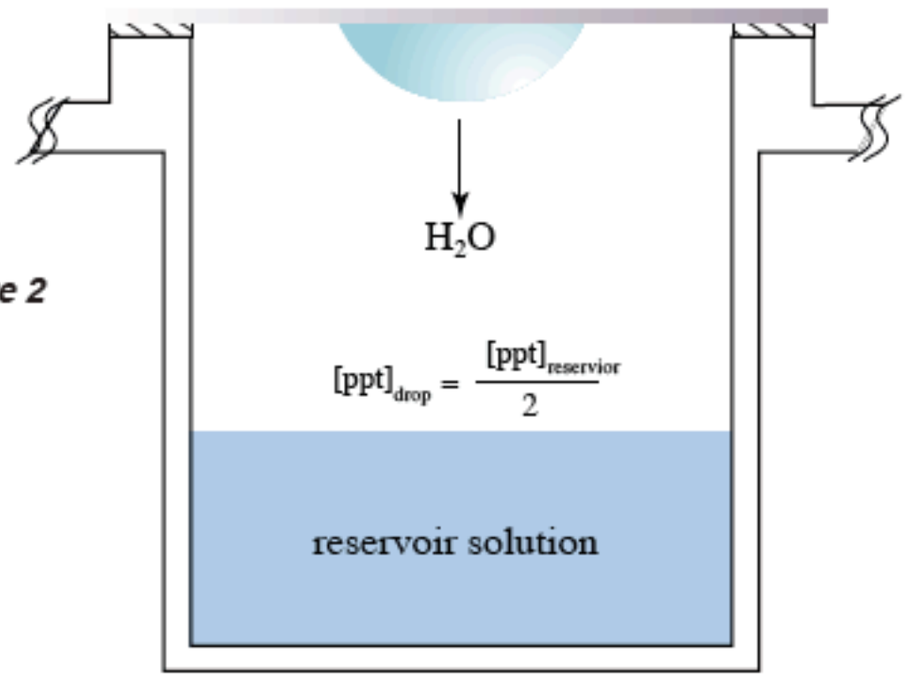
Vapor diffusion is the most commonly used method

figure 1



Sitting drop

figure 2



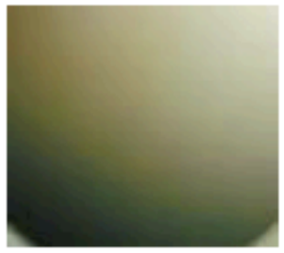
Hanging drop

A crystallization experiment takes days/weeks/months

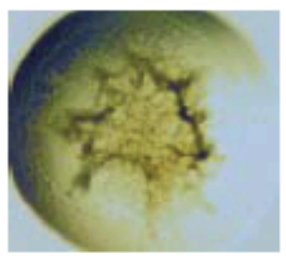
Hampton Research

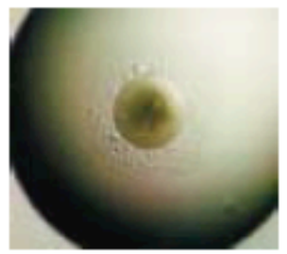
Crystallization

Spectrum of possible outcomes in a crystallization screen

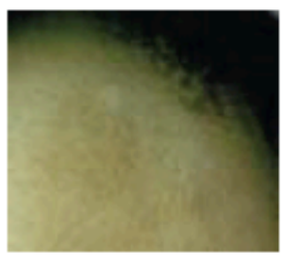

Clear Drop


Skin / Precipitate


Precipitate


Precipitate / Phase

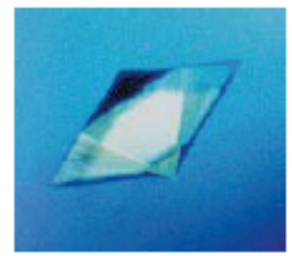

Quasi Crystals


Microcrystal


Needle Cluster


Plates

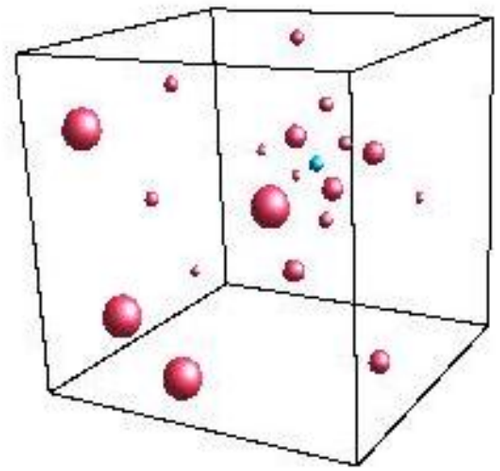

Rod Cluster

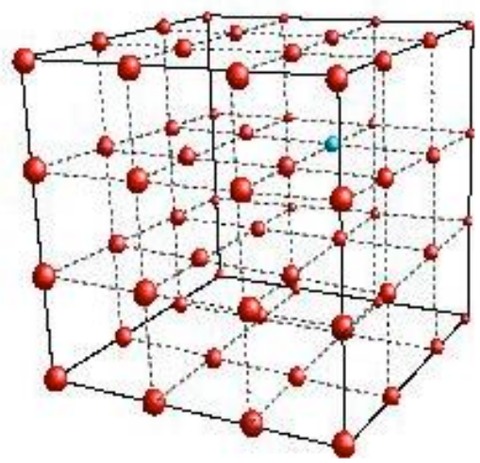

Single Crystal

Hampton Research

Finding the precipitant

- Trial-and-error approach
- Sparse-matrix screens sample “condition space”
- May also require sampling of “sequence space”

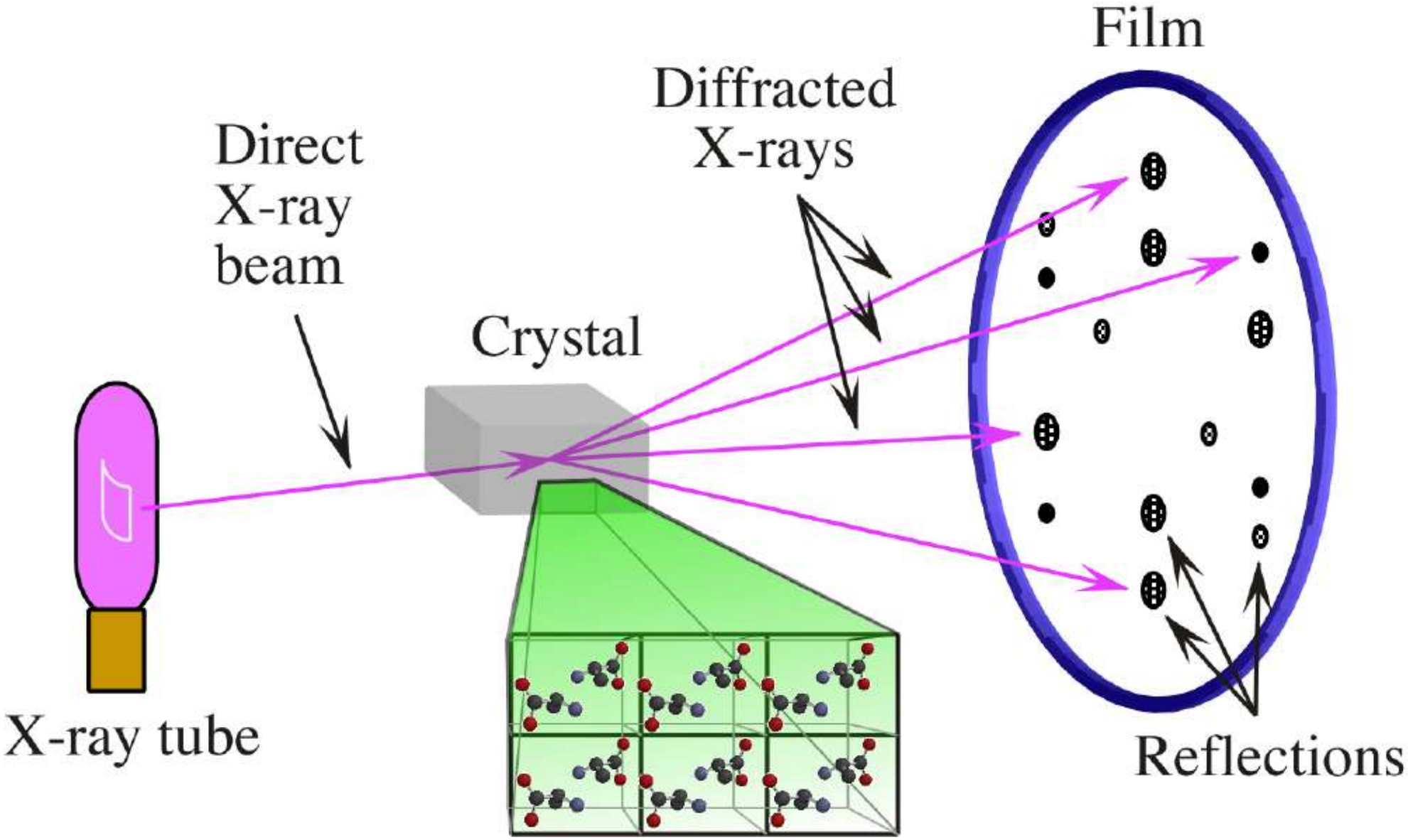




Crystallography workflow

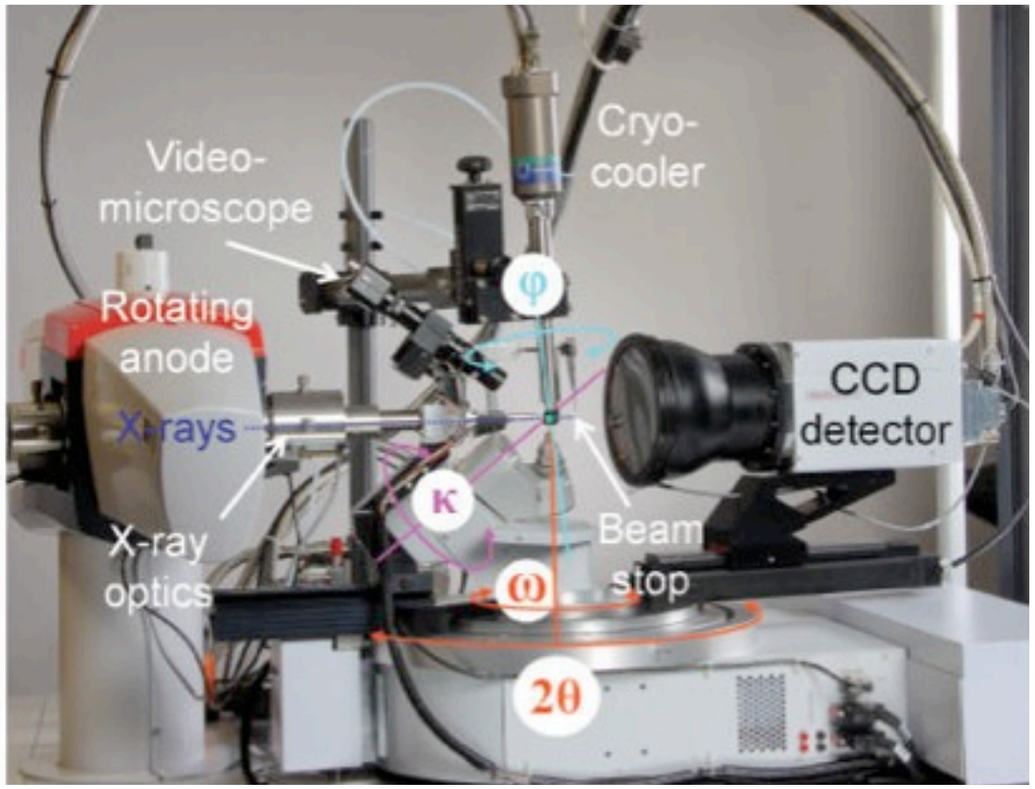
- Biological sample production
- Crystallization
- X-ray diffraction data collection
- Structure solution/phase determination
- Model building and validation

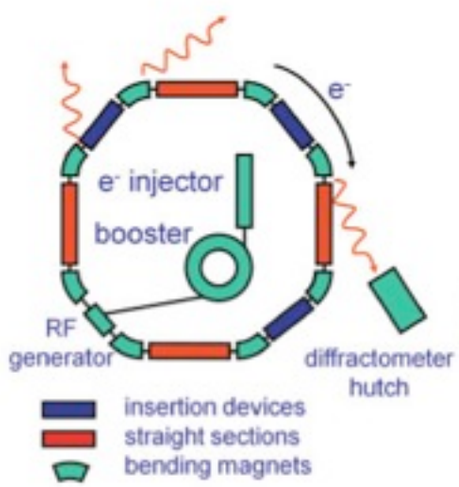
X-ray diffraction data

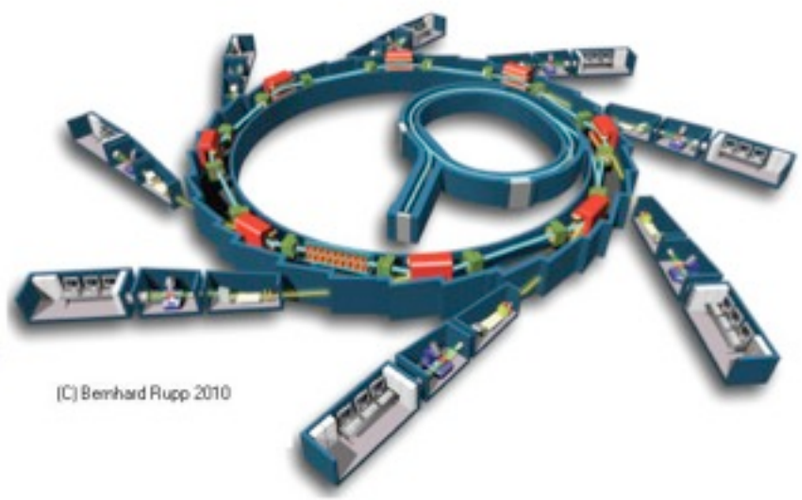


Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Where are data collected?

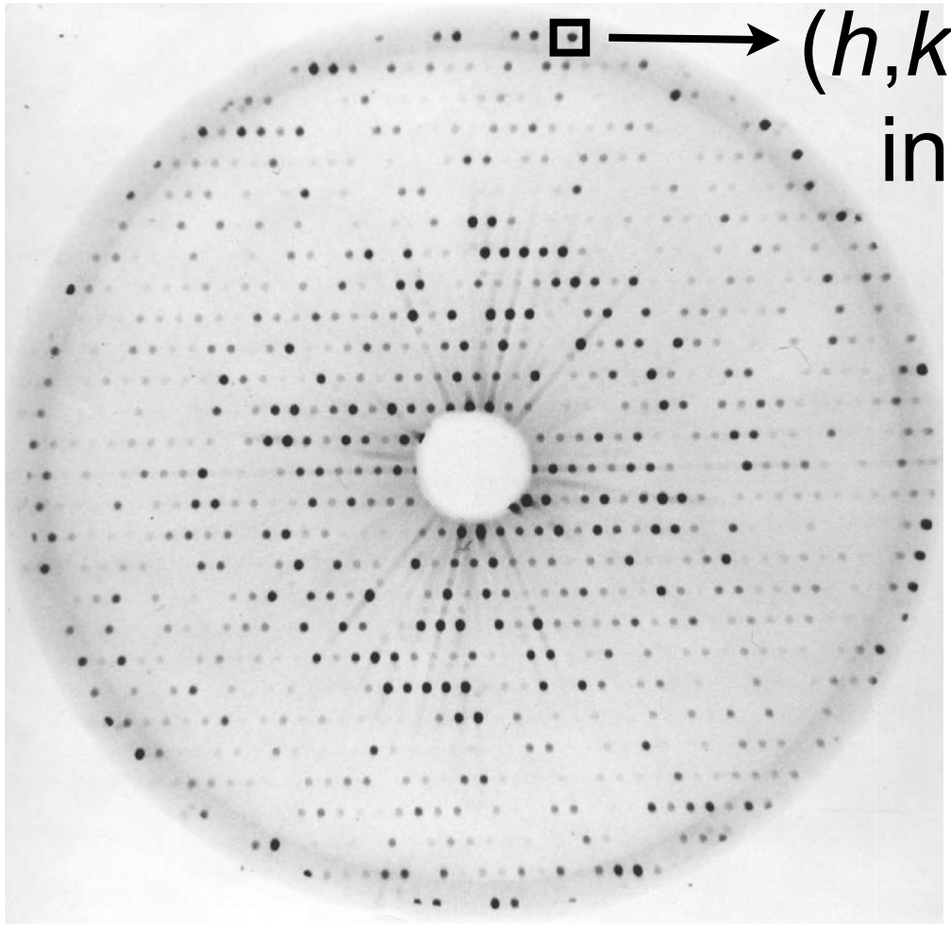






X-ray diffraction data

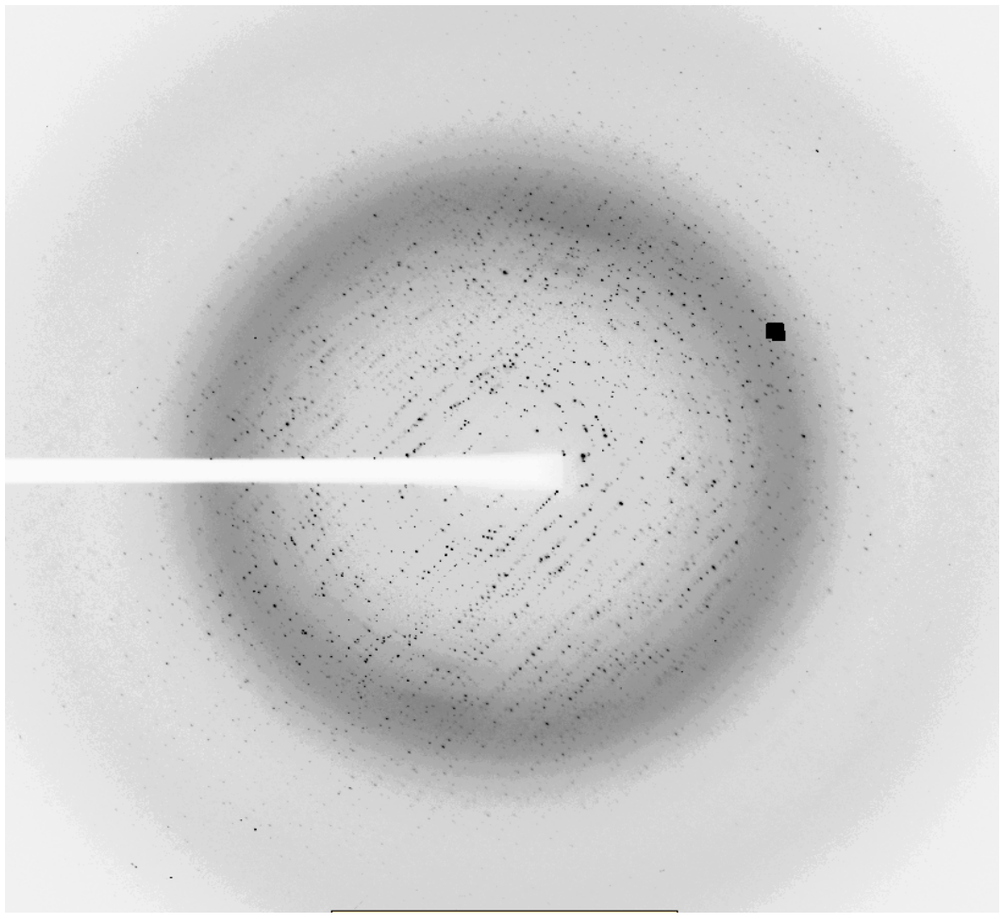
- What do we look for in a diffraction pattern?
  - spot shape
  - single lattice
  - resolution limit



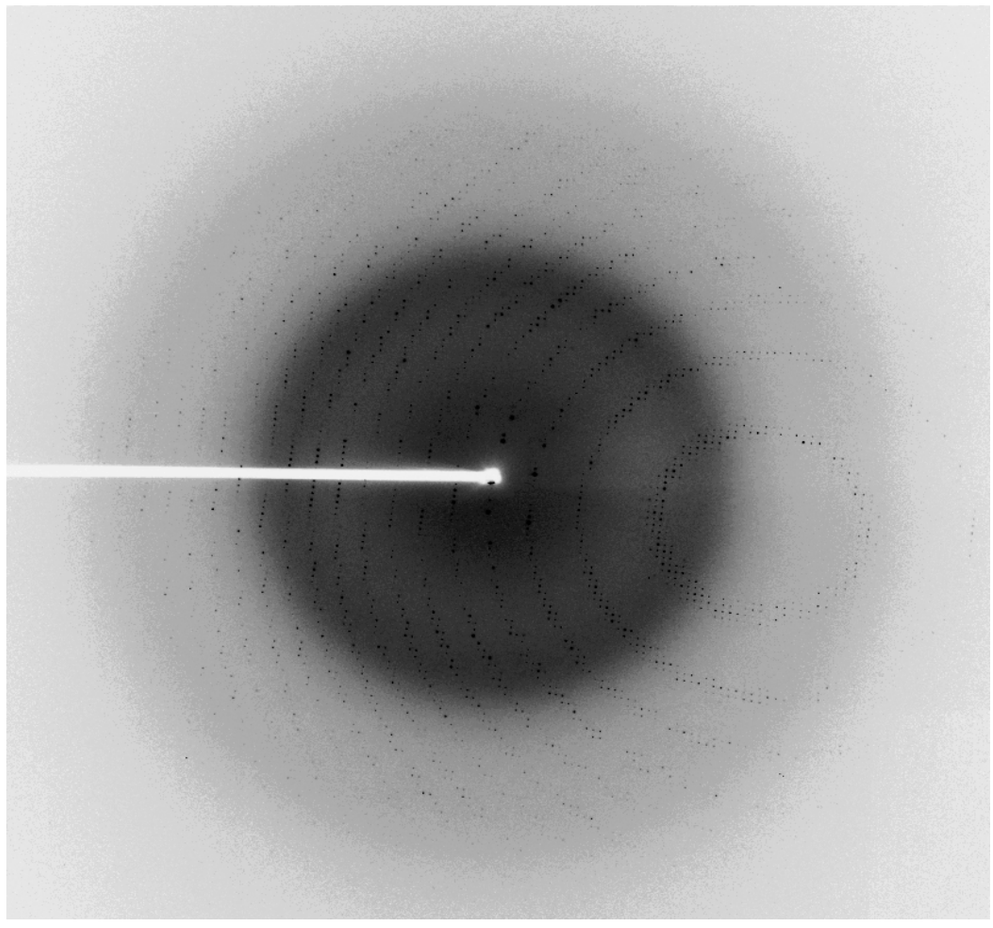
⊡ ⟶ (*h*,*k*,*l*) index, intensity

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Good and bad diffraction



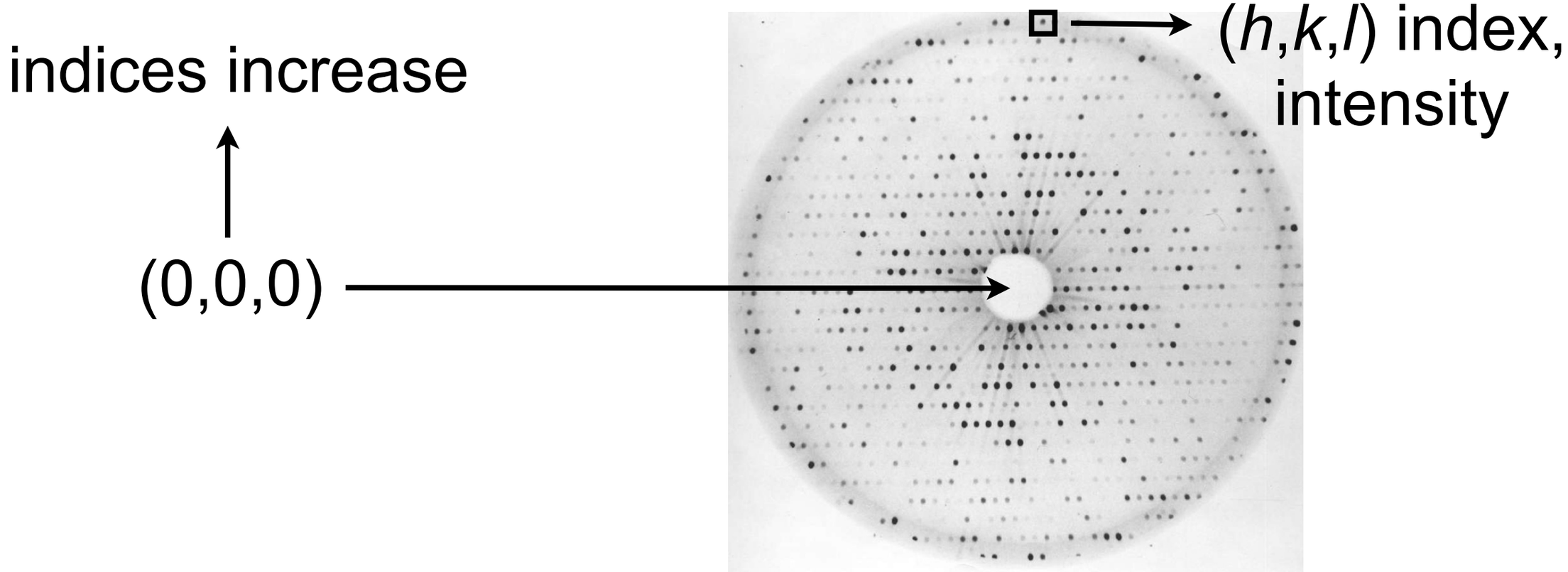
N-terminal His-tagged *B. subtilis* NrdF old prep



N-terminal His-tagged *B. subtilis* NrdF new prep
+ MonoQ purification step

X-ray diffraction data

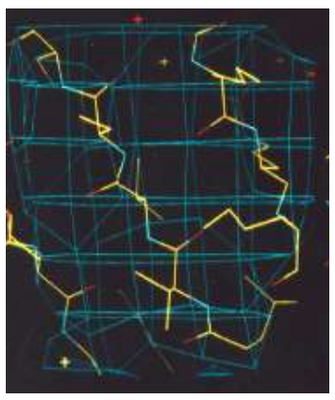
- Each spot is a reflected X-ray that is given an index
- Indices provide information about the crystal lattice
- Intensities provide information about positions of atoms



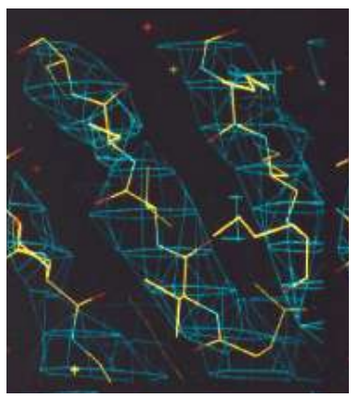
Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Importance of resolution

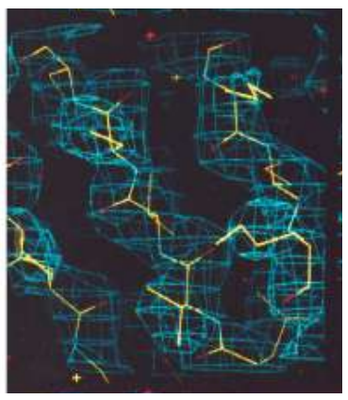
- Reflections far away from the origin provide high-resolution structural information
- Resolution affects level of detail in electron density maps



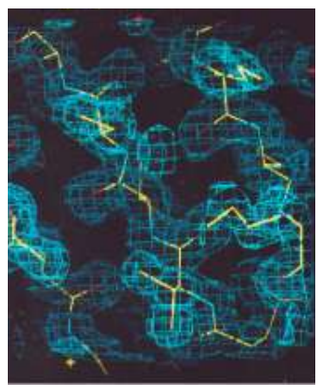
6.0 Å



4.5 Å



3.0 Å

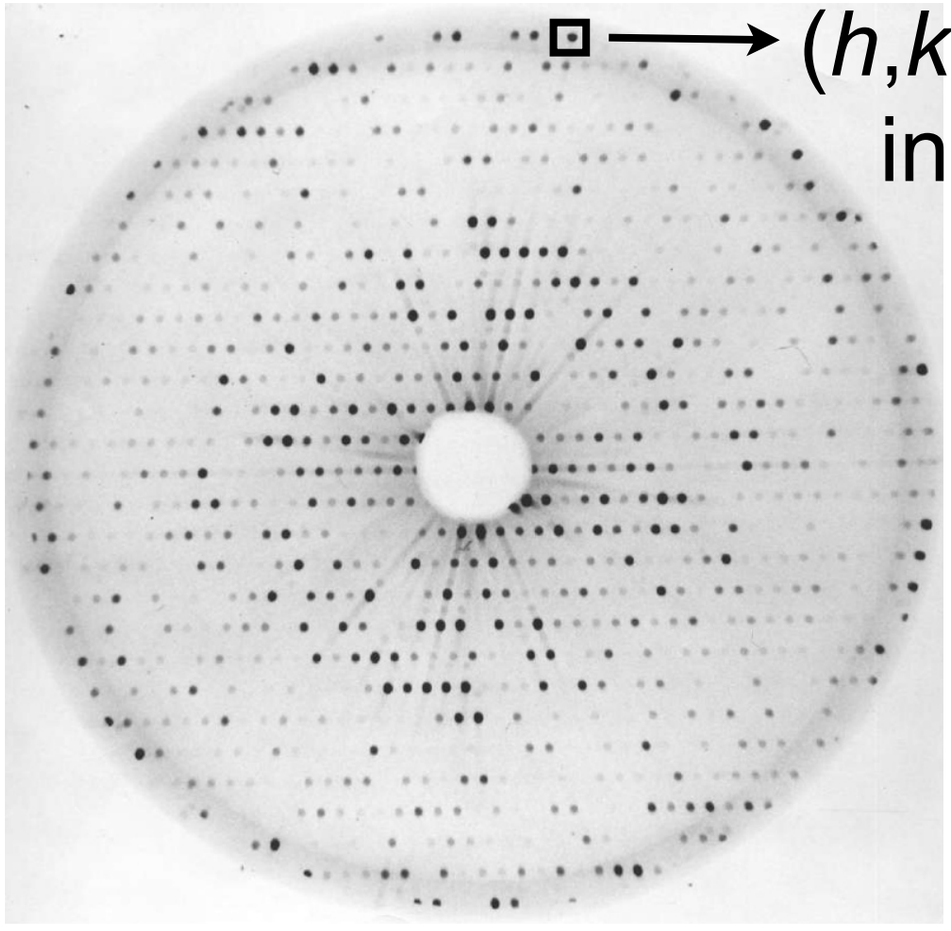


1.6 Å

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

X-ray diffraction data

- What do we look for in a diffraction pattern?
  - spot shape
  - single lattice
  - resolution limit



Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

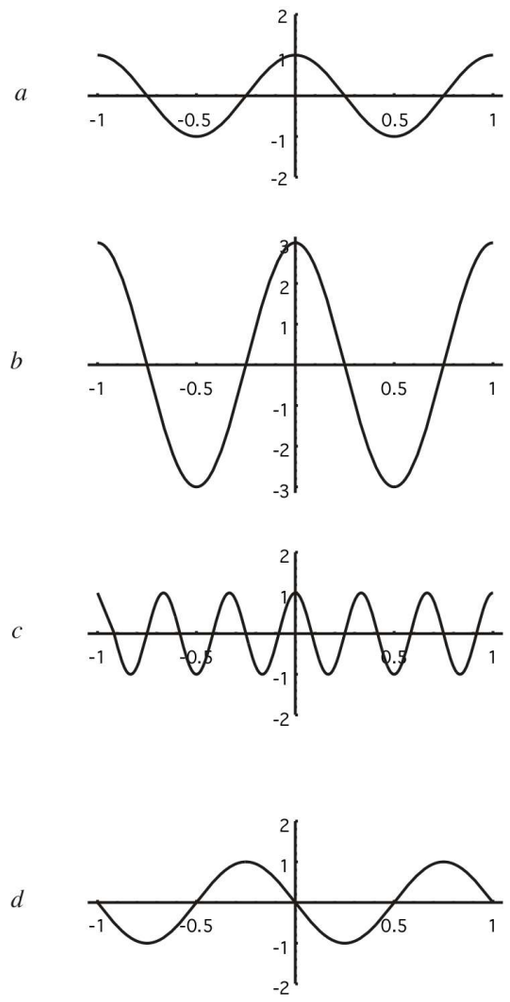
X-ray diffraction

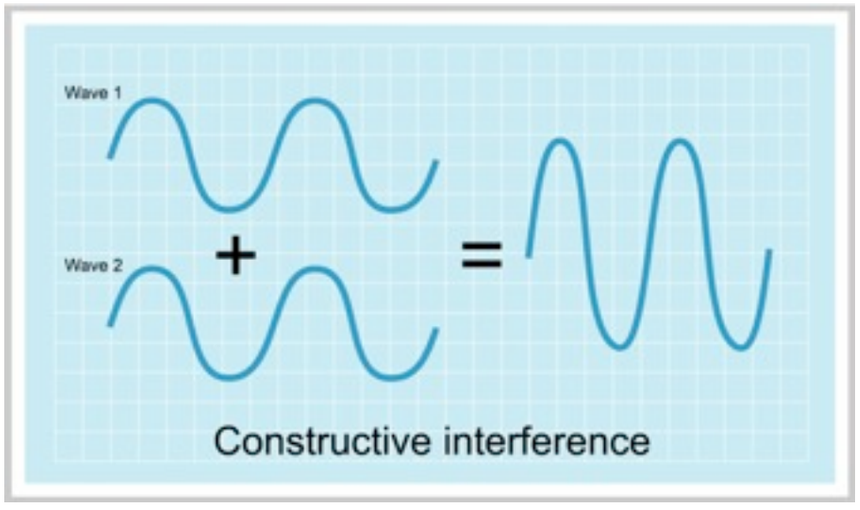
- Diffracted rays with strong constructive interference are observed
- Bragg’s law describes constructive interference in X-ray diffraction by crystals

X-ray diffraction

- X-rays have an amplitude, phase, frequency
- Constructive interference increases amplitude

$$f(x) = F \cos 2\pi(hx+\alpha)$$



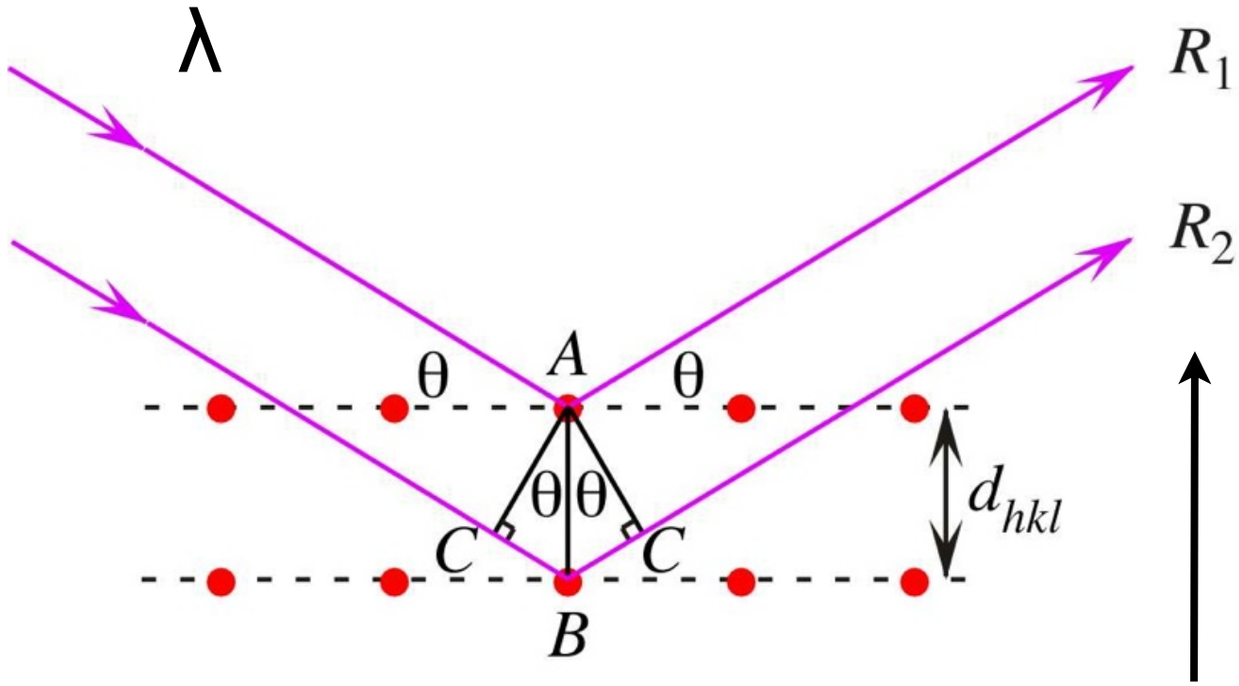


Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Bragg's law

- Parallel planes of atoms in crystals produce diffracted X-rays with strong constructive interference

- Conditions that satisfy Bragg's law lead to diffraction



$2d_{hkl} \sin \theta = n\lambda$ in phase

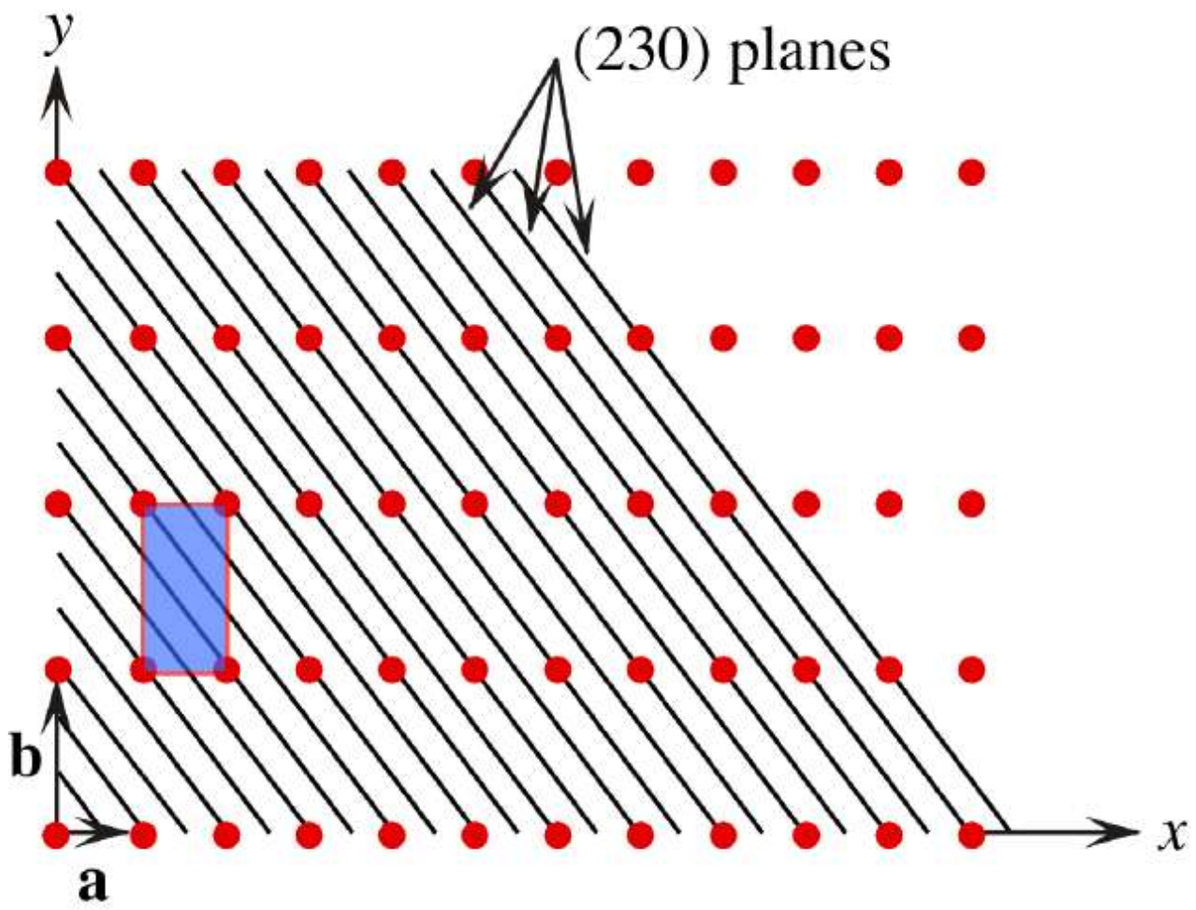
Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

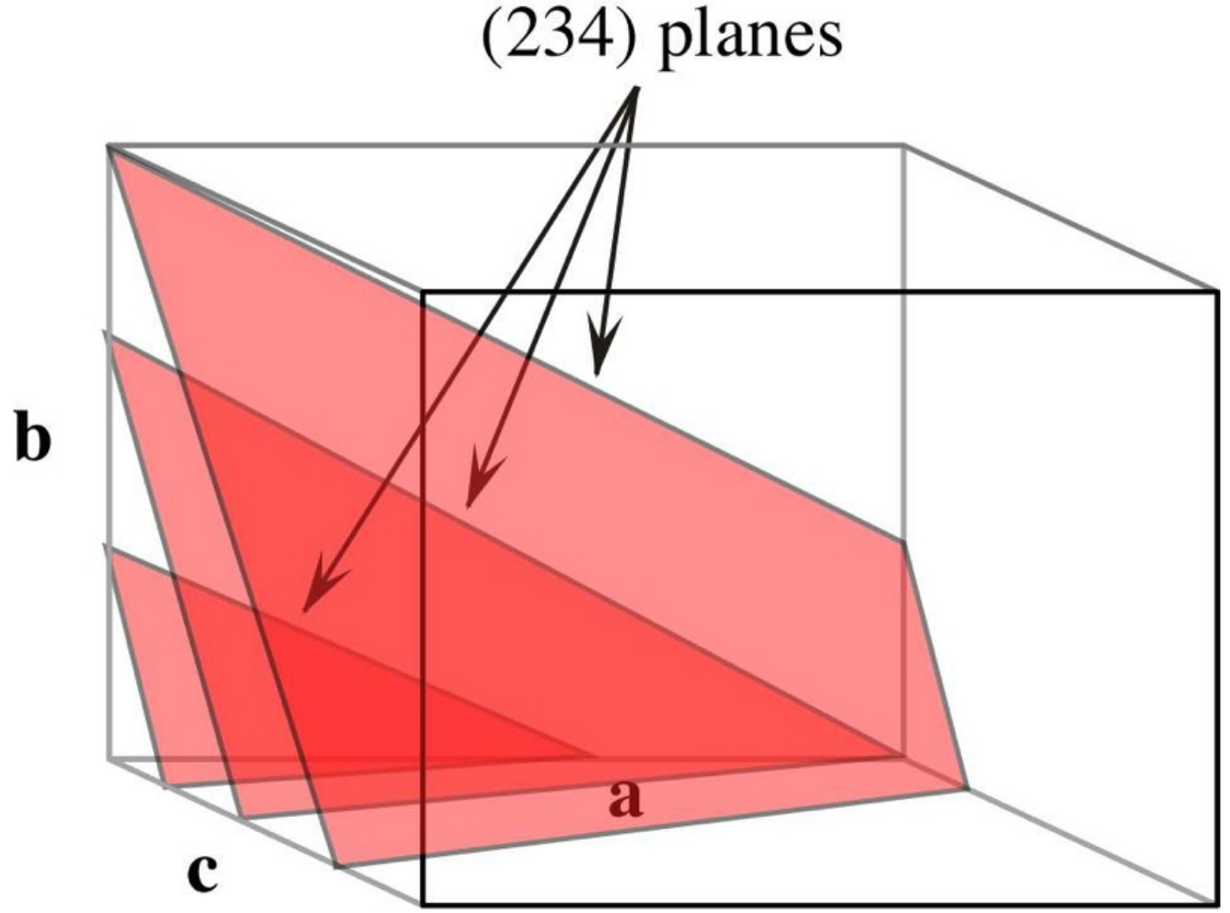
Bragg's law

- A crystal lattice can be divided into many different regularly spaced sets of parallel planes
- Indexed based on number of times planes intersect the unit cell on each a,b,c edge
- Index notation is *h*, *k*, *l* and corresponds the index of a spot in the diffraction pattern

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Examples of lattice planes

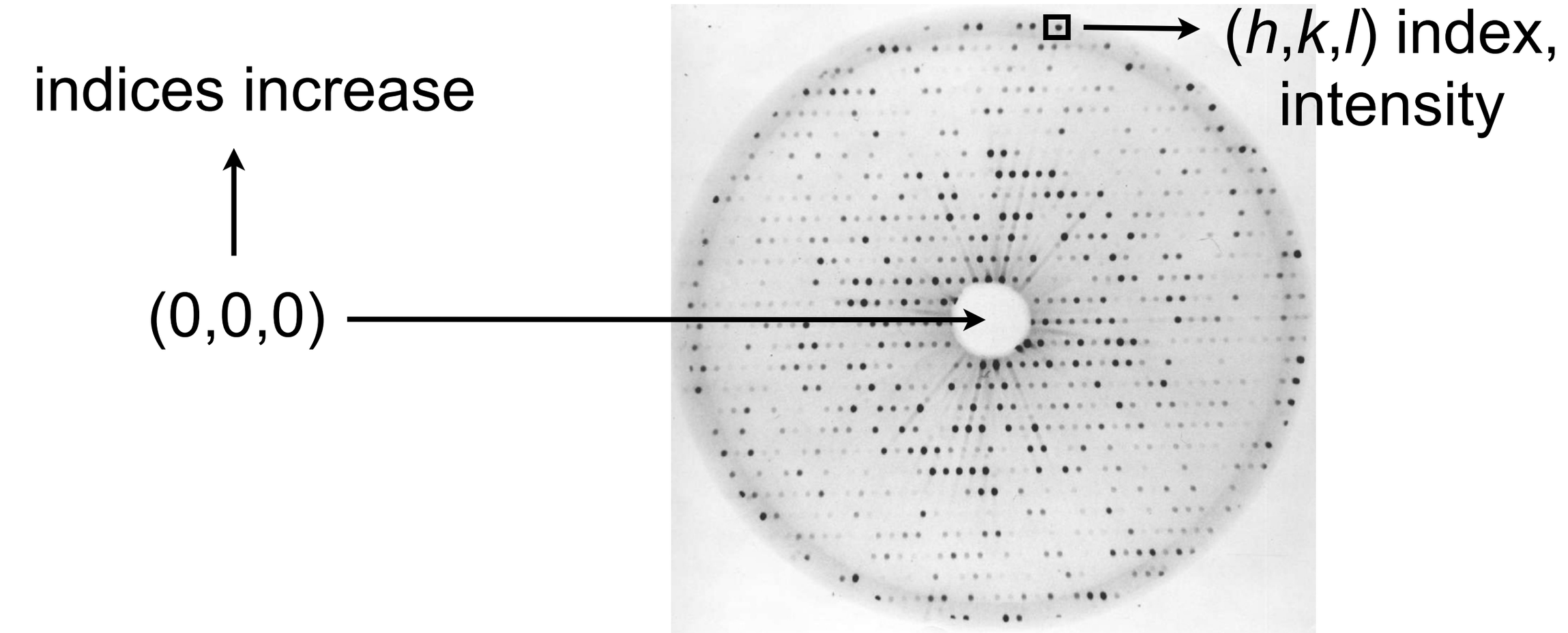




Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Diffraction data

- Each spot is a reflection with an index
- Indices provide information about the crystal lattice



Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

More about planes

- Each set of planes gives rise to a single reflection or spot
- Reflections form a reciprocal lattice
- The reciprocal lattice is related to the crystal lattice via indices of lattice planes
- Closely spaced planes provide more detailed information about electron density of atoms that lie on the planes

Diffraction patterns

- Can calculate unit cell dimensions from *h*,*k*,*l* indices
- Positions of spots in diffraction pattern contain information about lattice type and symmetry
- Intensity is a reporter of the electron density associated with given set of planes

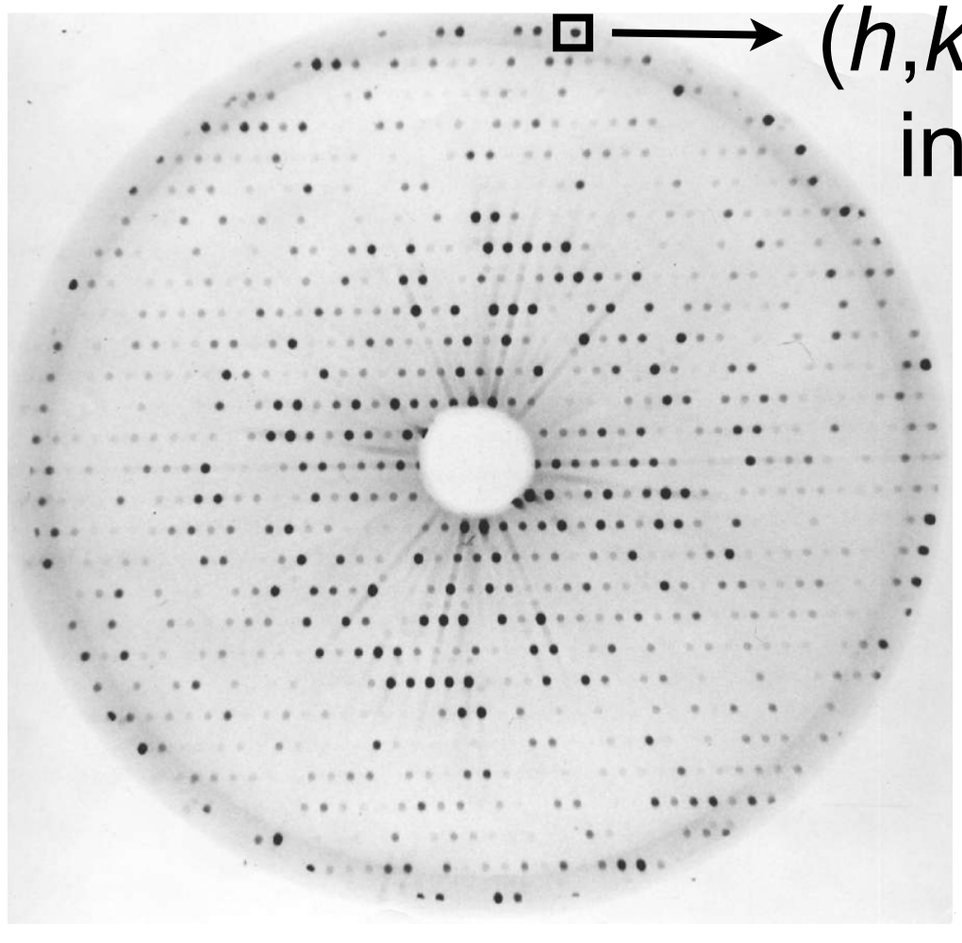
Diffraction data

- Each spot is a reflection with an index
- Indices provide information about the crystal lattice

Unit cell dimensions

Lattice type

Resolution



Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Matthews coefficient

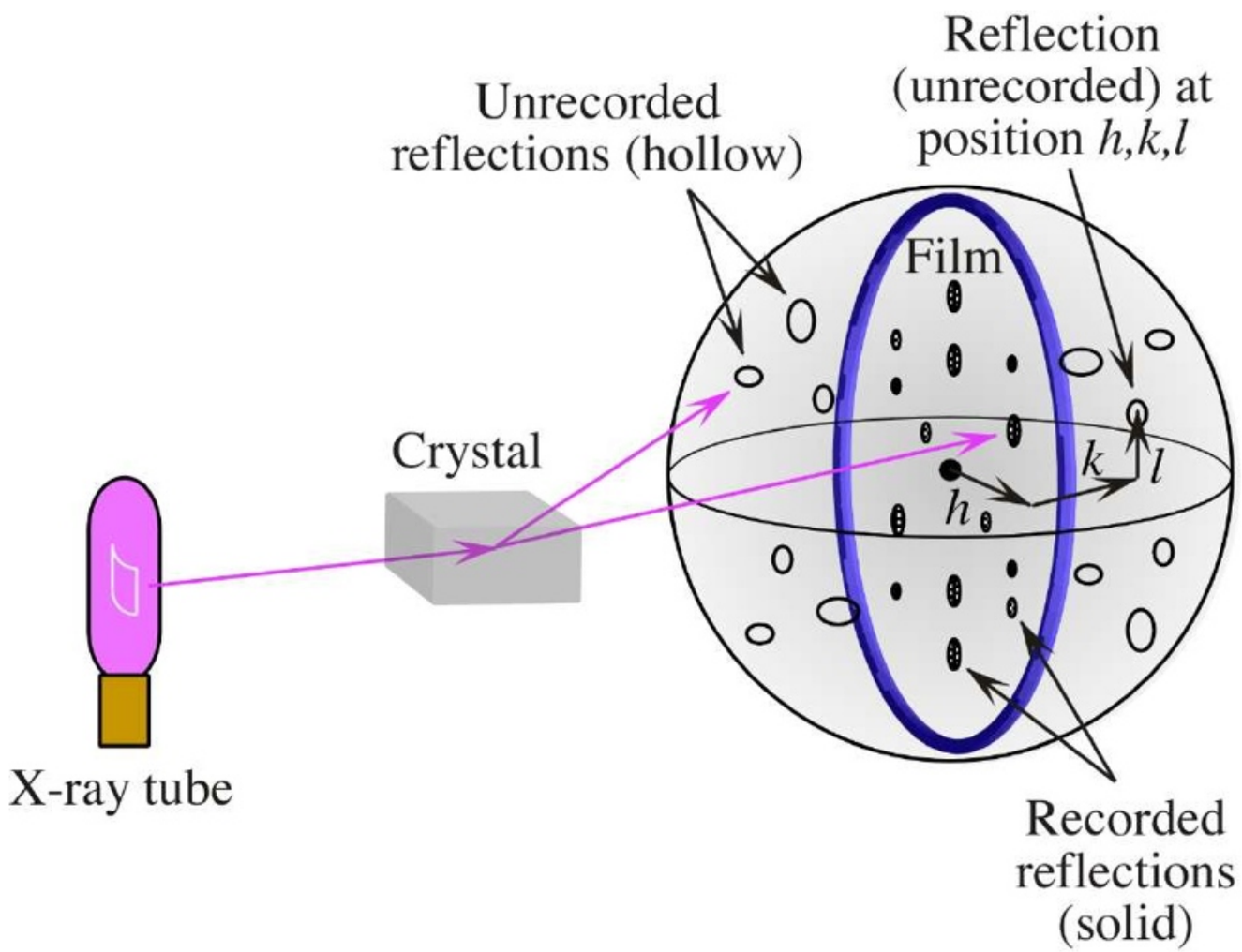
- Use unit cell dimensions to estimate number of residues in asymmetric unit
- Assume ~50% solvent content in protein crystals
- Provides useful information for structure solution and phasing

Kantardjieff, K.A., and Rupp, B. (2003) *Protein Science 12*, 1865-1871.

X-ray dataset

- Set of diffraction pattern images collected while rotating the crystal
- Goal is to collect a complete and redundant set of reflections and their intensities
- Diffraction patterns are typically not reported
- Use data collection statistics to judge quality
- Datasets are large (1-10 GB)

Data collection



Crystal rotation during data collection captures unrecorded reflections

Important parameters:

- oscillation angle
- number of images
- exposure time
- detector position

X-ray dataset

- Set of diffraction pattern images collected while rotating the crystal
- Goal is to collect a complete and redundant set of reflections and their intensities
- Diffraction patterns are typically not reported
- Use data collection statistics to judge quality
- Datasets are large (1-10 GB)

X-ray data processing

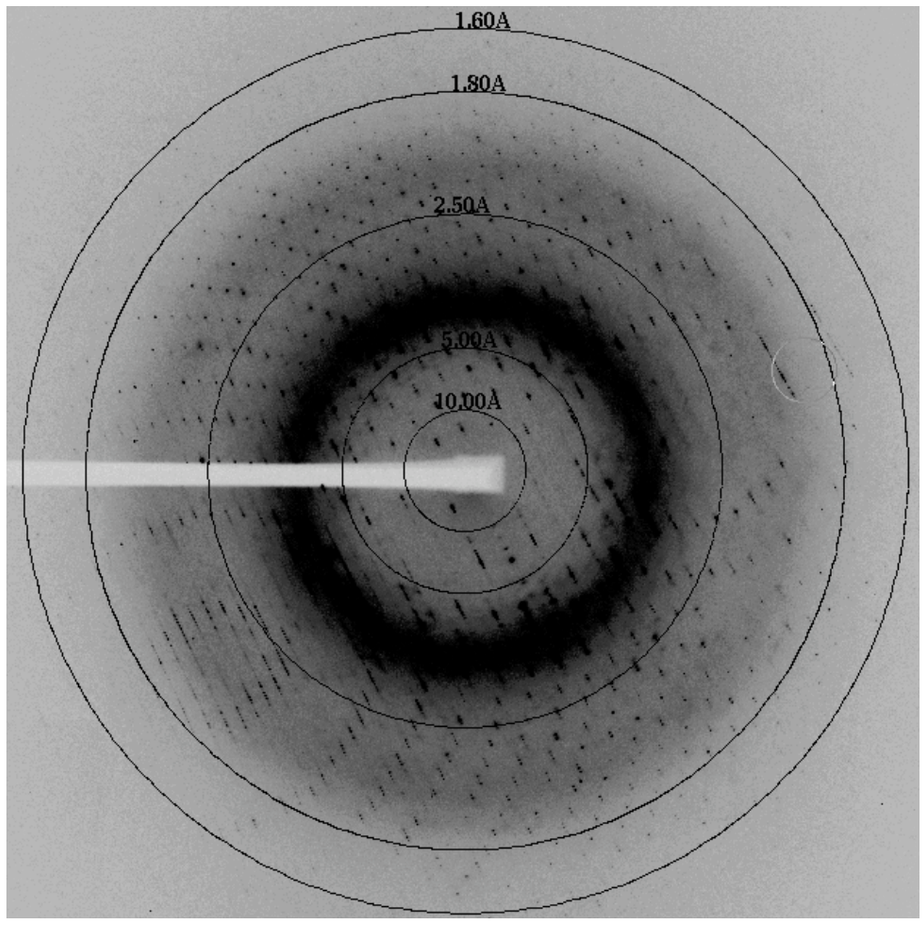
- Indexing
- Bravais lattice and unit cell assignment
- Integration
- Scaling and analysis of statistics
- Space group and resolution limit determination

Data collection statistics

- **Resolution**
  - ~3.5-4.0 Å to resolve secondary structure
  - ~2.5-3.5 Å to identify side chains
  - ~1.5-2.5 Å to model side chains/backbone confidently
- **Intensity to background ratio (*I*/sigma*I*)**
  - Sets resolution limit
- **R_{merge} or R_{sym}**
  - Measure of intensity differences between reflections with same index (<0.1 overall)

Resolution bins

- Spots in each diffraction image are binned by resolution
- Statistics are compiled for overall dataset and individual resolution bins



Data collection statistics

- **Resolution**
  - ~3.5-4.0 Å to resolve secondary structure
  - ~2.5-3.5 Å to identify side chains
  - ~1.5-2.5 Å to model side chains/backbone confidently
- **Intensity to background ratio (*I*/s*I*), completeness, redundancy**
  - Sets resolution limit
- **R_{merge} or R_{sym}**
  - Measure of intensity differences between reflections with same index (<0.1 overall)

Data collection statistics

	WT
Data collection statistics	
Maximum resolution (Å)	1.50
Wavelength (Å)	0.9918
Total reflections	98,554
Unique reflections	27,758
Completeness (%)[a]	98.9 (97.1)
$I/\sigma(I)$	7.8 (1.5)
R_{sym} (%)[b]	5.1 (40.4)
Refinement statistics	
Resolution limits (Å)	31.5–1.50
R-factor[c]	0.184
R-free	0.216
Estimated coordinate error (Å)[d]	0.09
RMS deviations from ideal values	
Bond lengths (Å)	0.024
Bond angles (°)	2.192
Dihedral angles (°)	25.11
Improper torsion angles (°)	1.60
Average temperature factor (Å^2)	
Protein	22.1, 20.6
Iron-sulfur	13.0, 12.8
Water	36.8
Zinc	37.1
Sulfate	44.3
Ramachandran plot,[e] residues in	
Most favored regions (%)	90.2
Additional allowed regions (%)	9.1
Generously allowed regions (%)	0.6
Disallowed regions (%)	0.0

Overall

< 0.1 overall

High res. bin

> 2.0 high res. bin

Yeh, A.P., Ambroggio, X.I., Andrade, S.L.A., Einsle, O., Chatelet, C., Meyer, J., and Rees, D.C. (2002) *J. Biol. Chem. 277*, 34499-34507.

Structure solution

- Structure factors relate diffraction intensities to electron density and atom position
  - Each reflection can be described by a structure factor (F_{hkl}) function
  - F_{hkl} is the sum of diffractive contributions of all atoms in the unit cell

no. of atoms in unit cell

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

atomic scattering factor

Structure solution

- Electron density map is 3D plot of ρ(*x, y, z*)
- ρ(*x, y, z*) is a Fourier sum

electron density function

structure factor

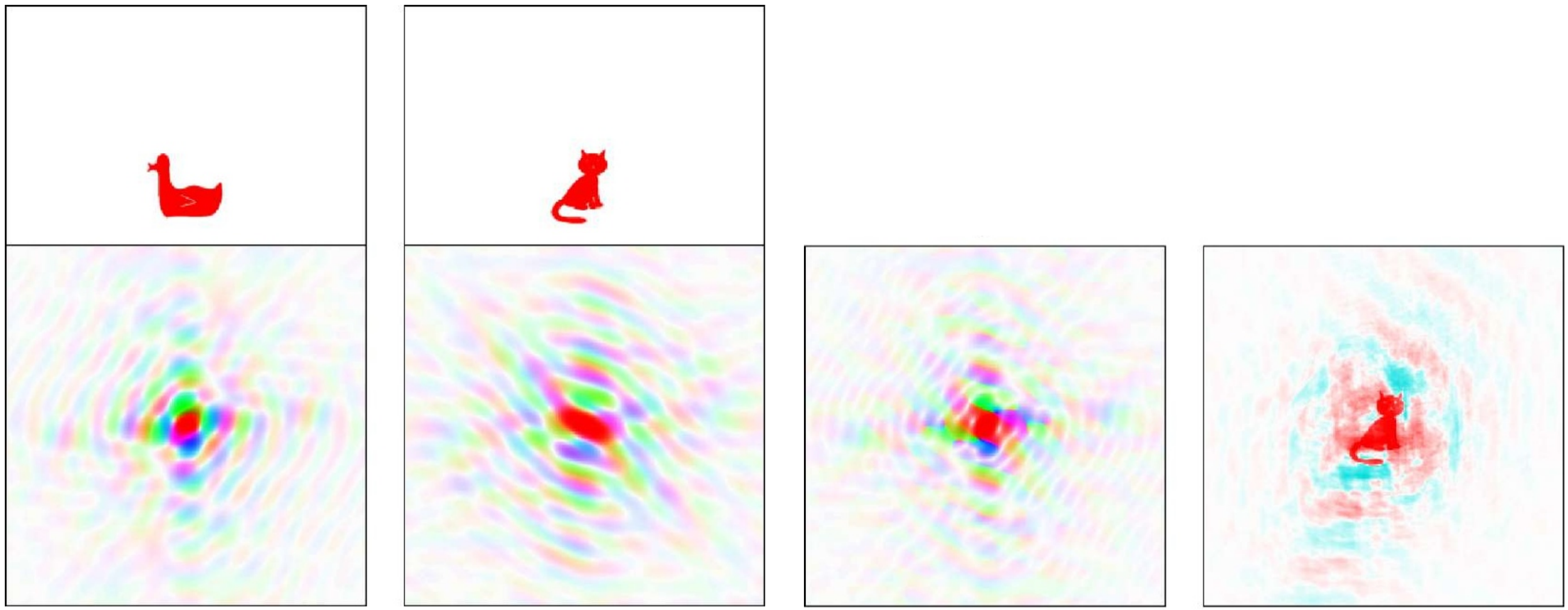
$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl} e^{-2\pi i(hx_j+ky_j+lz_j)}$$

sum over all reflections in diffraction pattern

Structure solution

- How do we plot ρ(*x, y, z*) from diffraction data?
- F_{hkl} is a periodic function and has an amplitude, frequency, and phase
  - amplitude ~ $\sqrt{I_{hkl}}$
  - frequency = $\frac{1}{d_{hkl}}$
  - phase = ??

The phase problem



- Phases contain most of the information necessary for structure determination

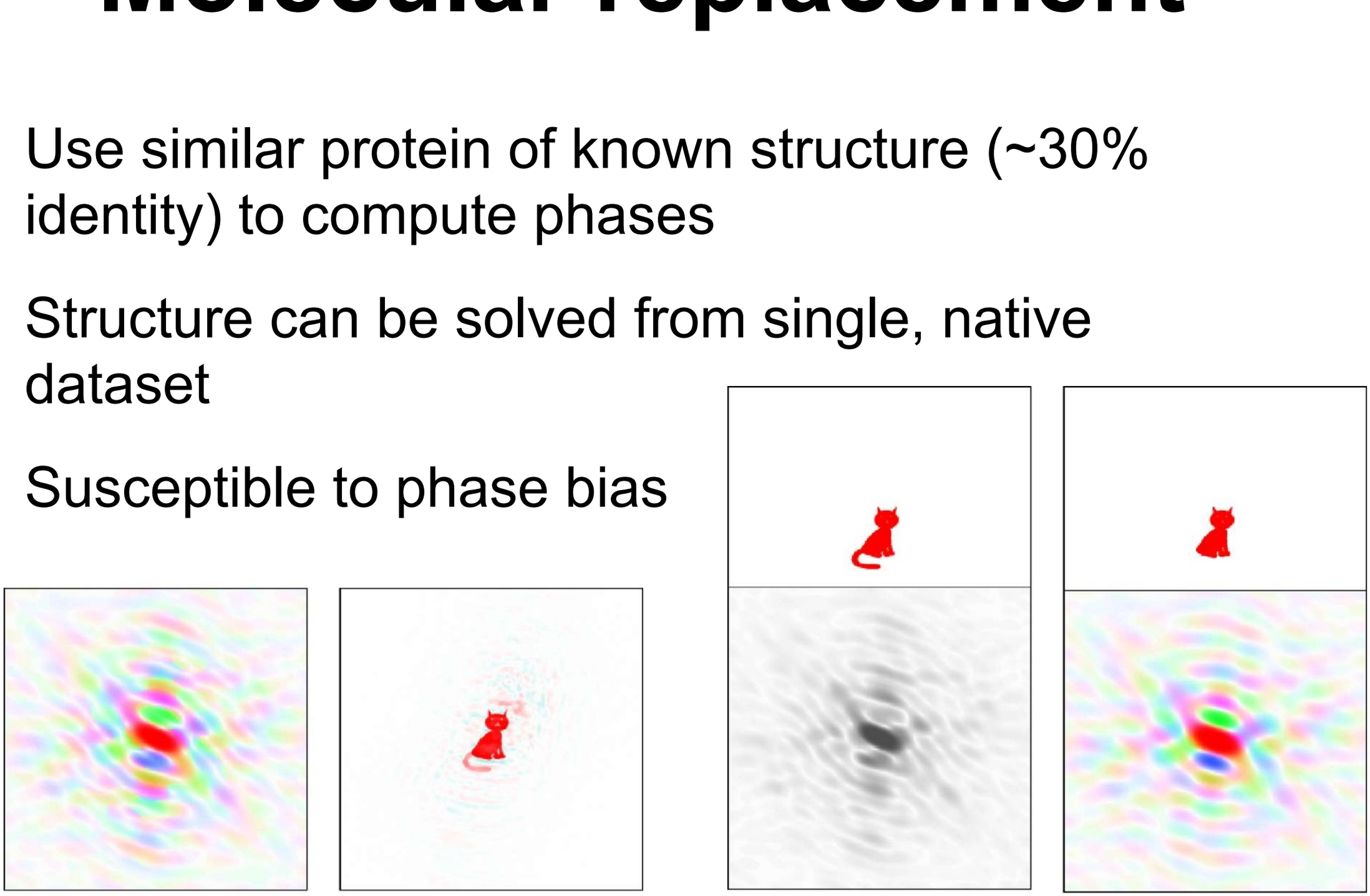
Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Solving the phase problem

- Molecular replacement
- Anomalous diffraction methods
- Multiple isomorphous replacement

Molecular replacement

- Use similar protein of known structure (~30% identity) to compute phases
- Structure can be solved from single, native dataset
- Susceptible to phase bias



Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

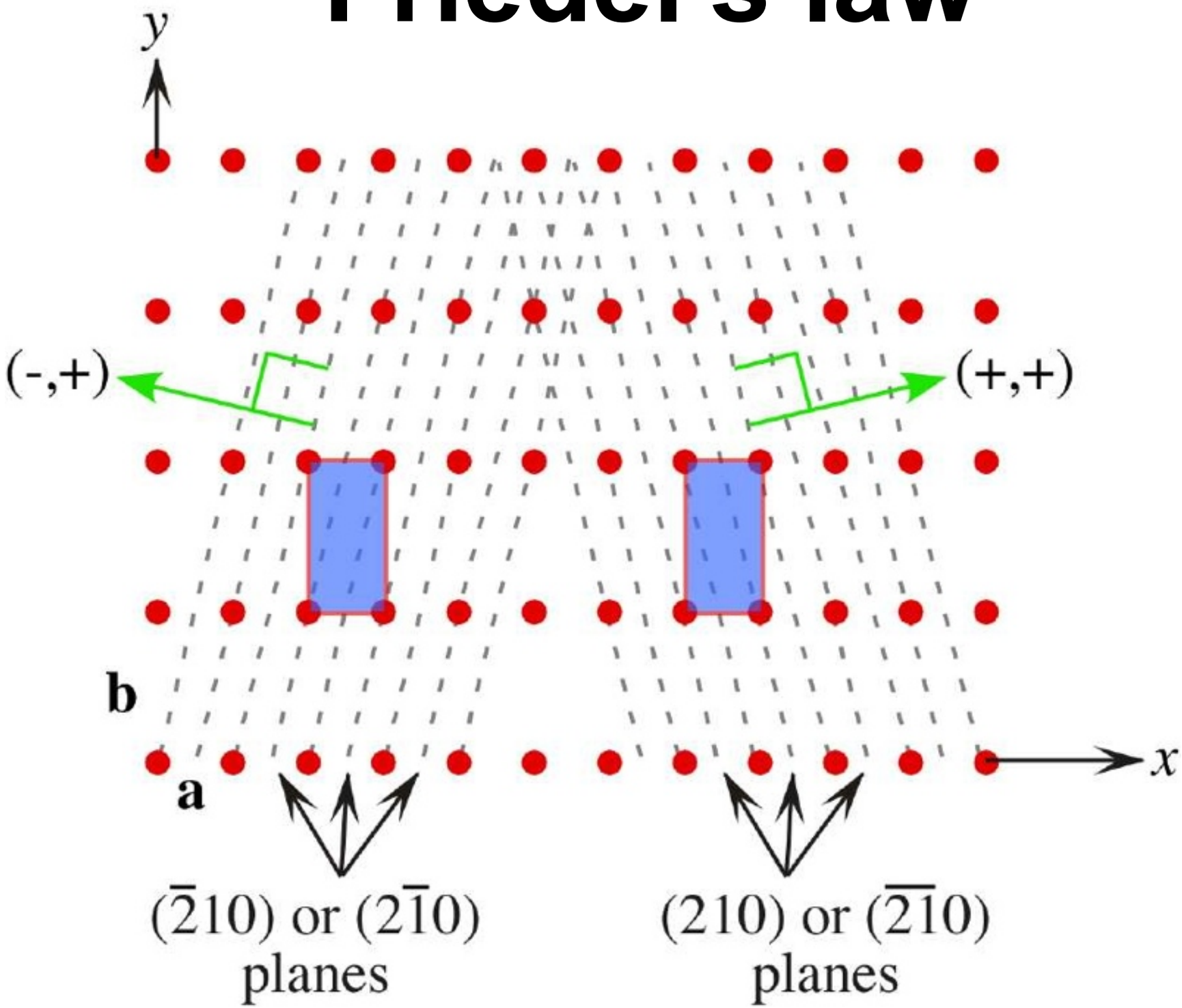
Experimental phasing

- Use heavy atom sites (transition metals or halogens) in crystals (derivative or native)
- Determine position of heavy atoms in unit cell
- Use phase information from heavy atoms to compute protein phases
- Typically requires collection of multiple datasets and/or crystals
- Generates electron density map without a model

Anomalous diffraction

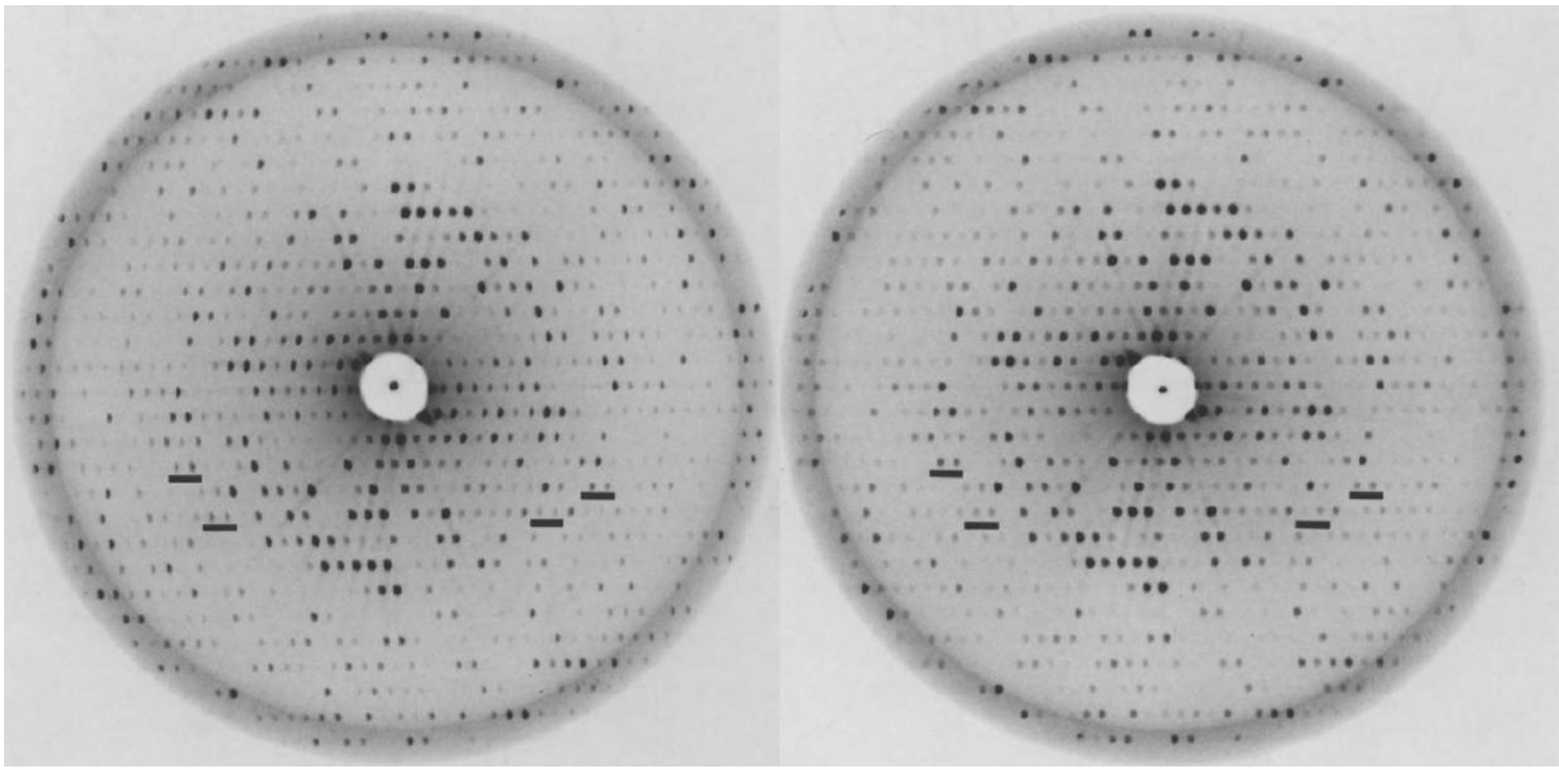
- X-ray absorption by heavy atoms alters diffraction
  - Friedel's law does not hold $I_{hkl} \neq I_{\overline{hkl}}$
- Intensity differences in Friedel pairs allows for heavy atom location in unit cell
- Datasets are collected near heavy atom absorption edge
- SeMet substitution or native metal cofactors used as anomalous scatterers
- Requires a tunable X-ray source

Friedel's law



Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

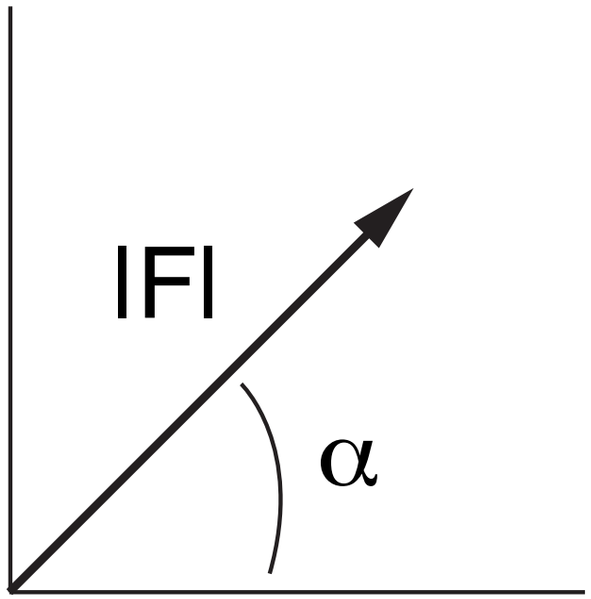
Isomorphous replacement

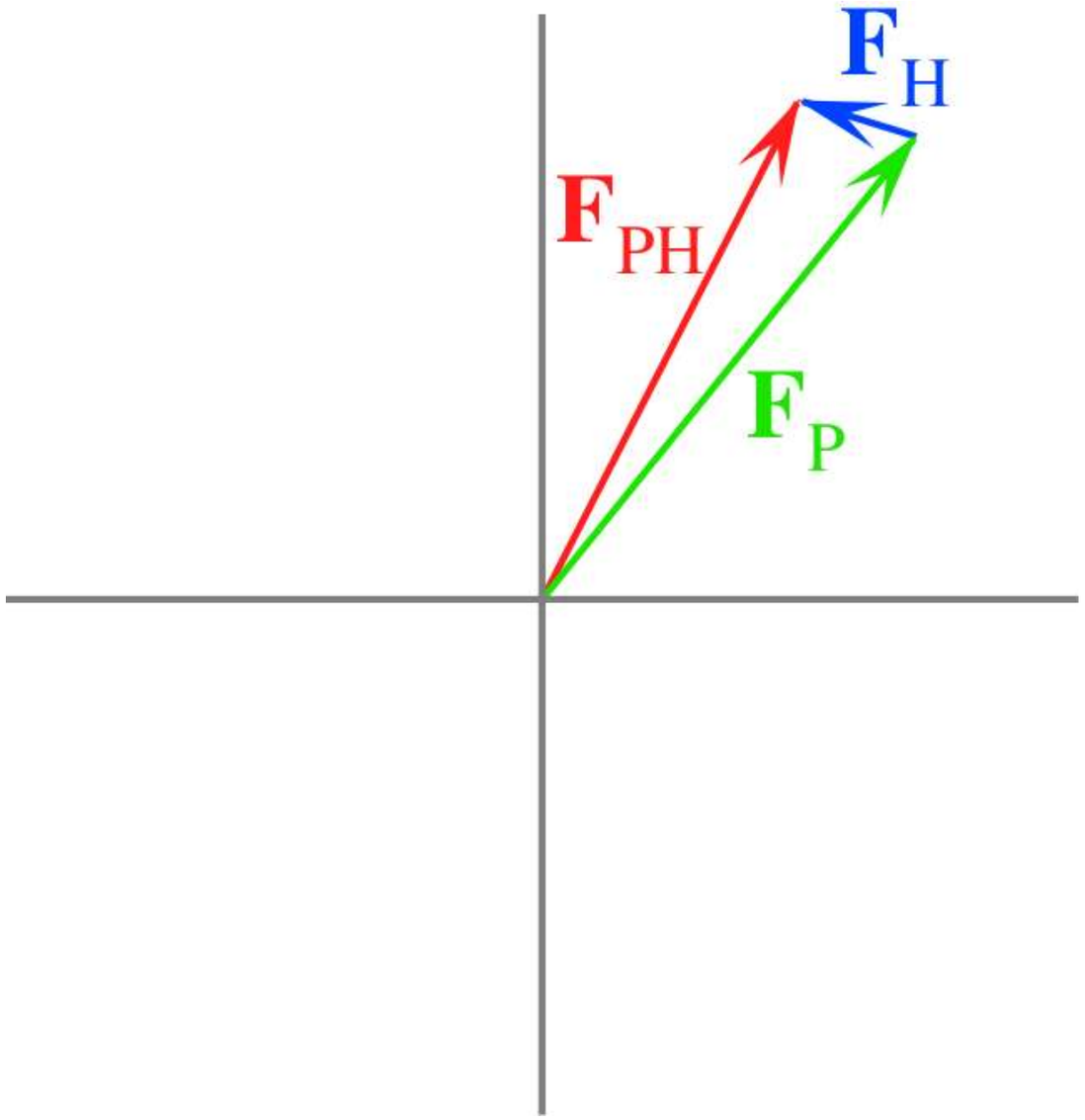


- Requires many crystals
- Heavy atoms must bind to same site
- Crystals must be isomorphous

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Experimental phasing



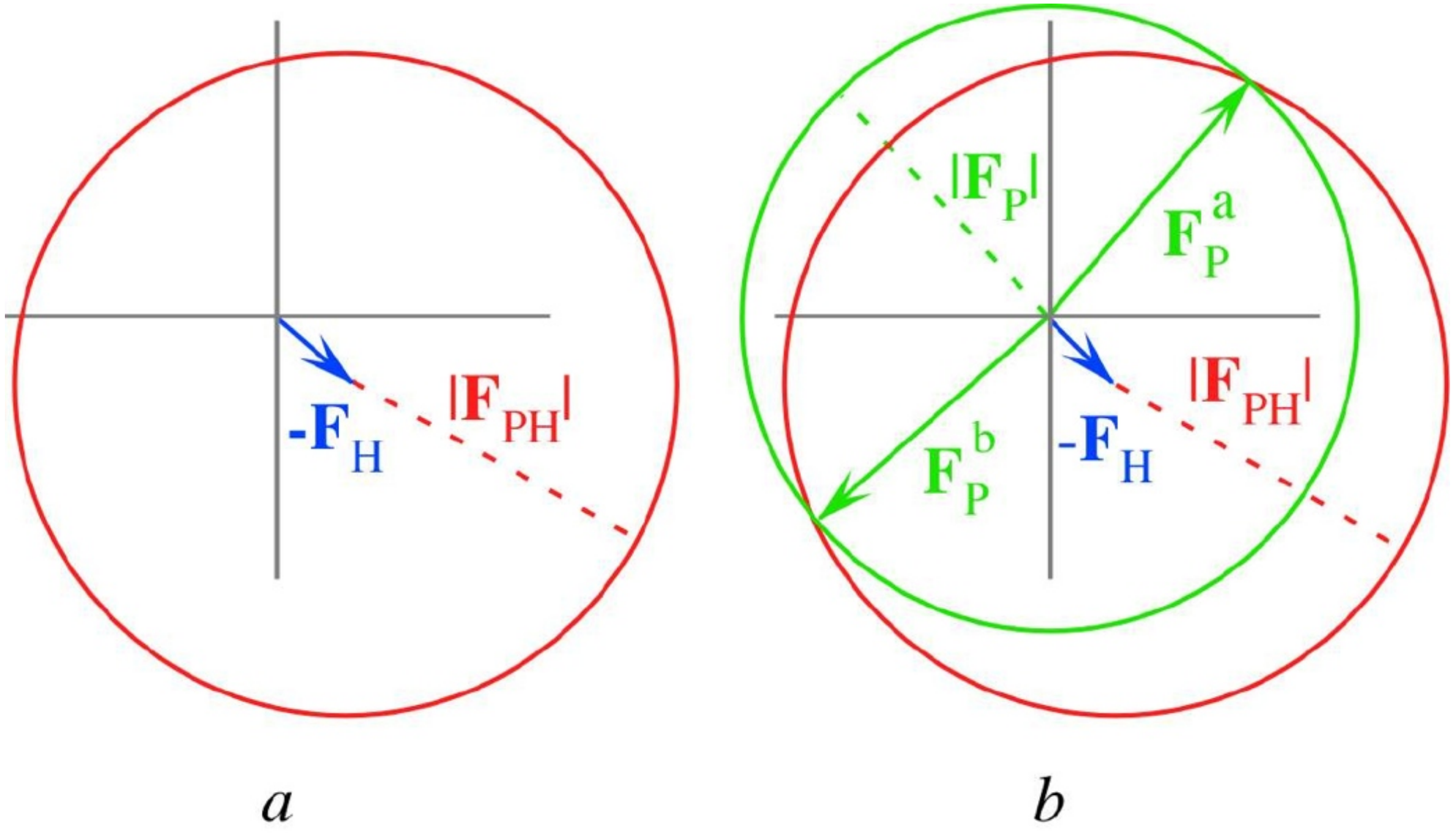


Structure factors can be represented as vectors

$$\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$$

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

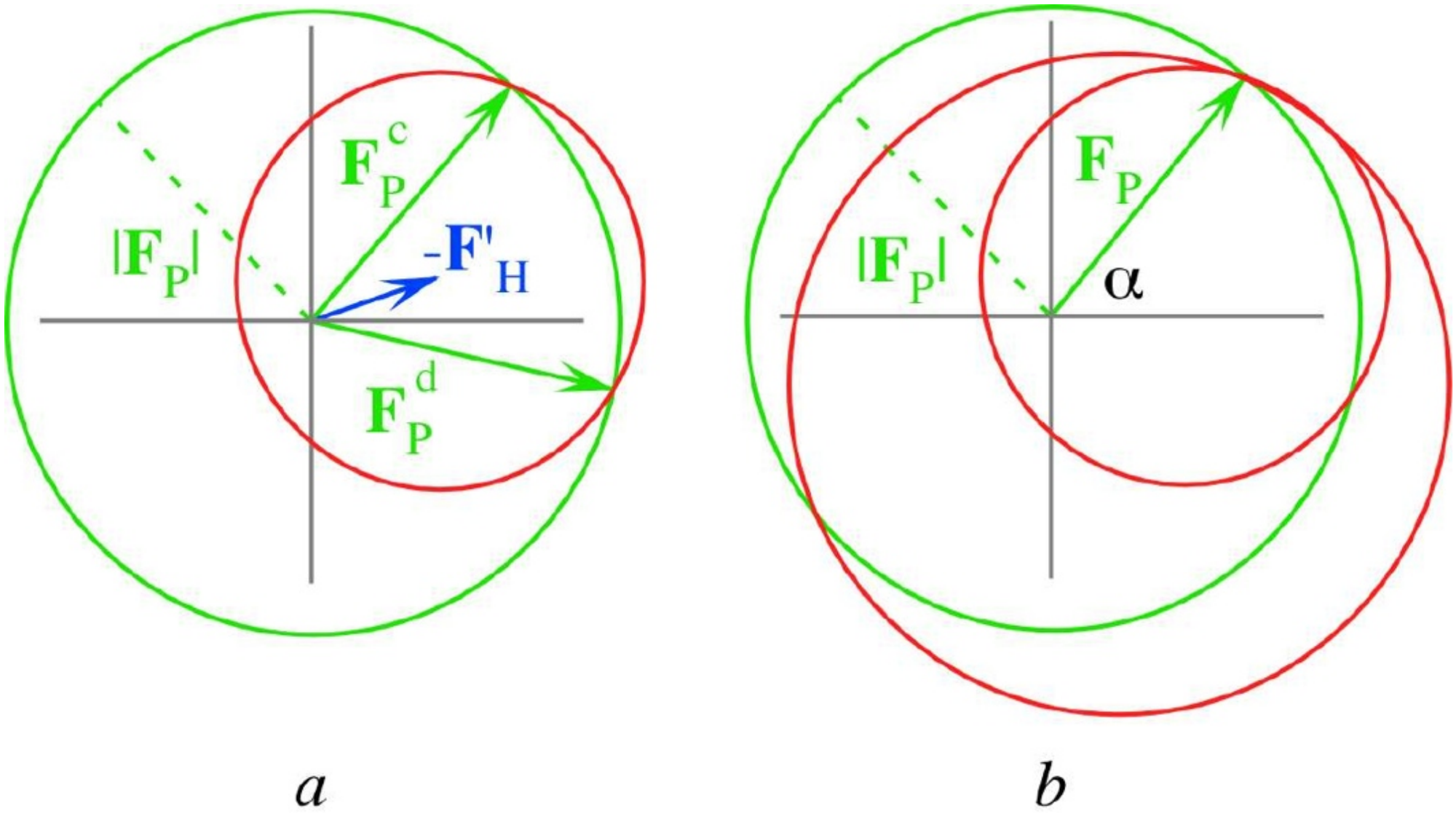
Experimental phasing



Harker diagrams provide vector solutions for $\mathbf{F_P}$

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Experimental phasing



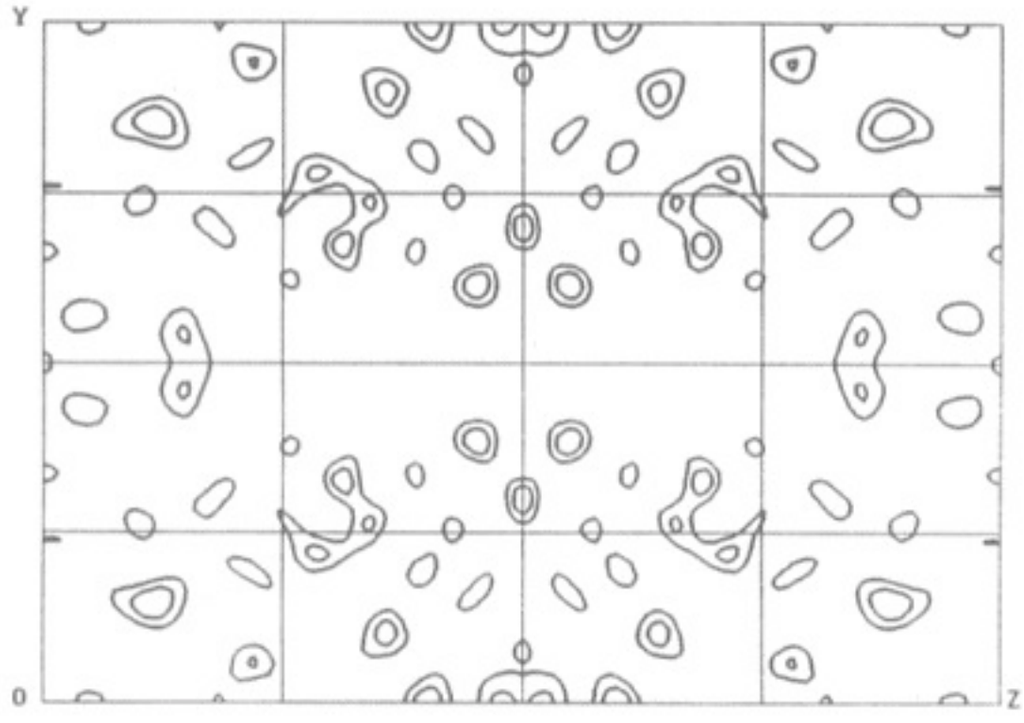
Phase ambiguity resolved by additional derivatives

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Heavy atom location

$$P(u,v,w) = \frac{1}{V}\sum_h\sum_k\sum_l |\mathbf{F}_{hkl}|^2 e^{-2\pi i(hu+kv+lw)}$$

Contour map of Patterson function *P*(*u*, *v*, *w*) shows peaks corresponding to vectors between atoms

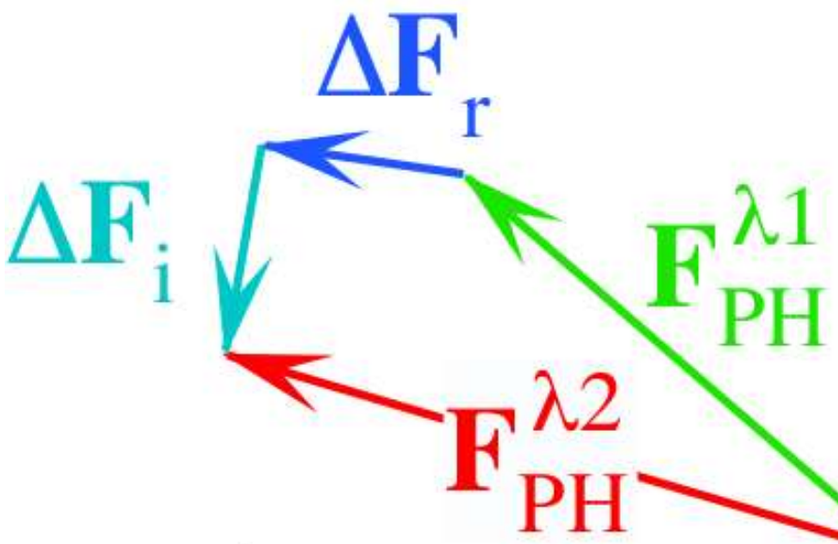


Cell symmetry simplifies 3D search for Patterson atoms

Harker sections contain Patterson vectors for symmetry-related atoms

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Anomalous diffraction



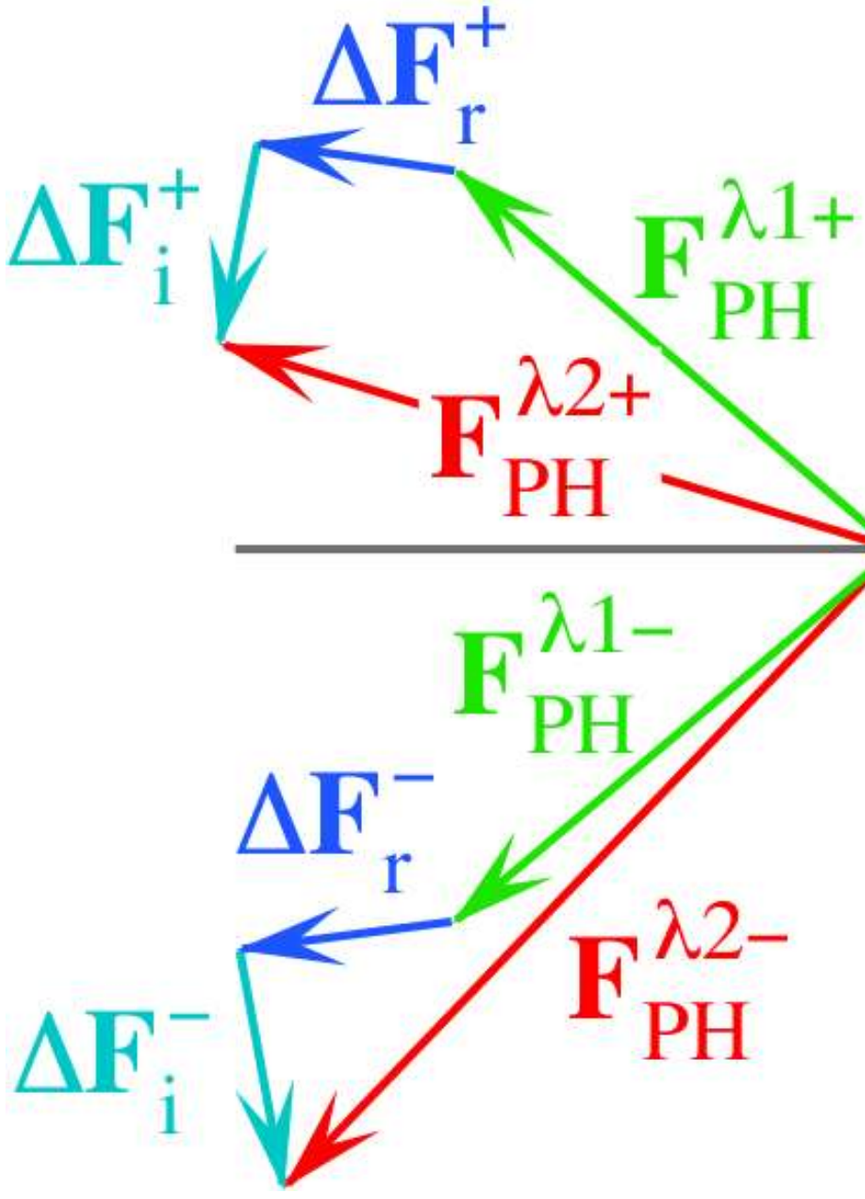
$$\mathbf{F}_{PH}^{\lambda 2} = \mathbf{F}_{PH}^{\lambda 1} + \Delta\mathbf{F}_r + \Delta\mathbf{F}_i$$

$F_{PH}^{\lambda 2}$ w/ anomalous scattering

$F_{PH}^{\lambda 1}$ w/o anomalous scattering

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Anomalous diffraction



$F_{PH}^{\lambda 1}$ Friedel pairs are equal in intensity and with phase opposite in sign

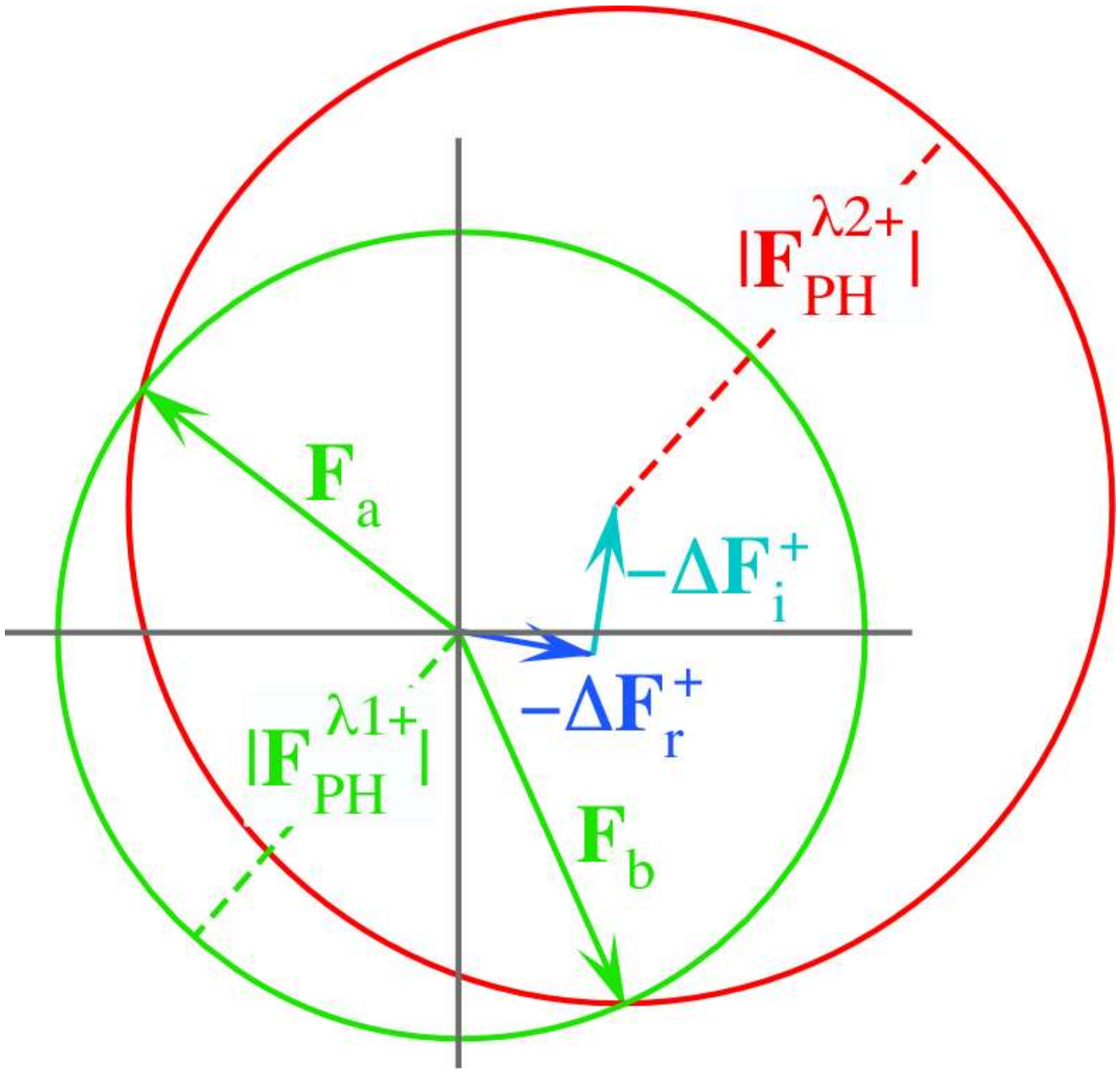
$F_{PH}^{\lambda 2}$ Friedel pairs have different intensity due to change in imaginary scattering factor

Magnitude of anomalous scattering contributions are known quantities

Phases depend only on atom positions

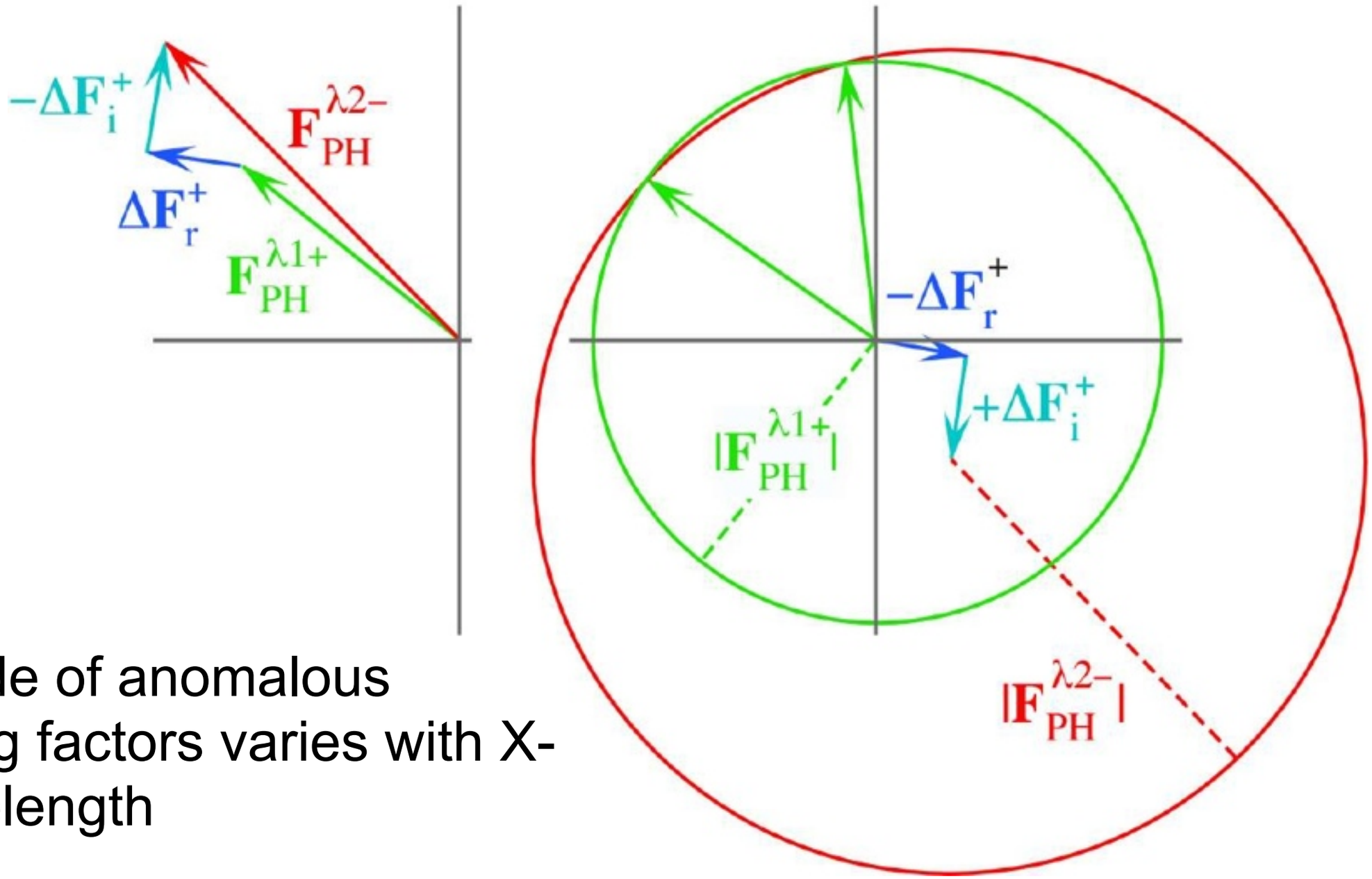
Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Anomalous diffraction



Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

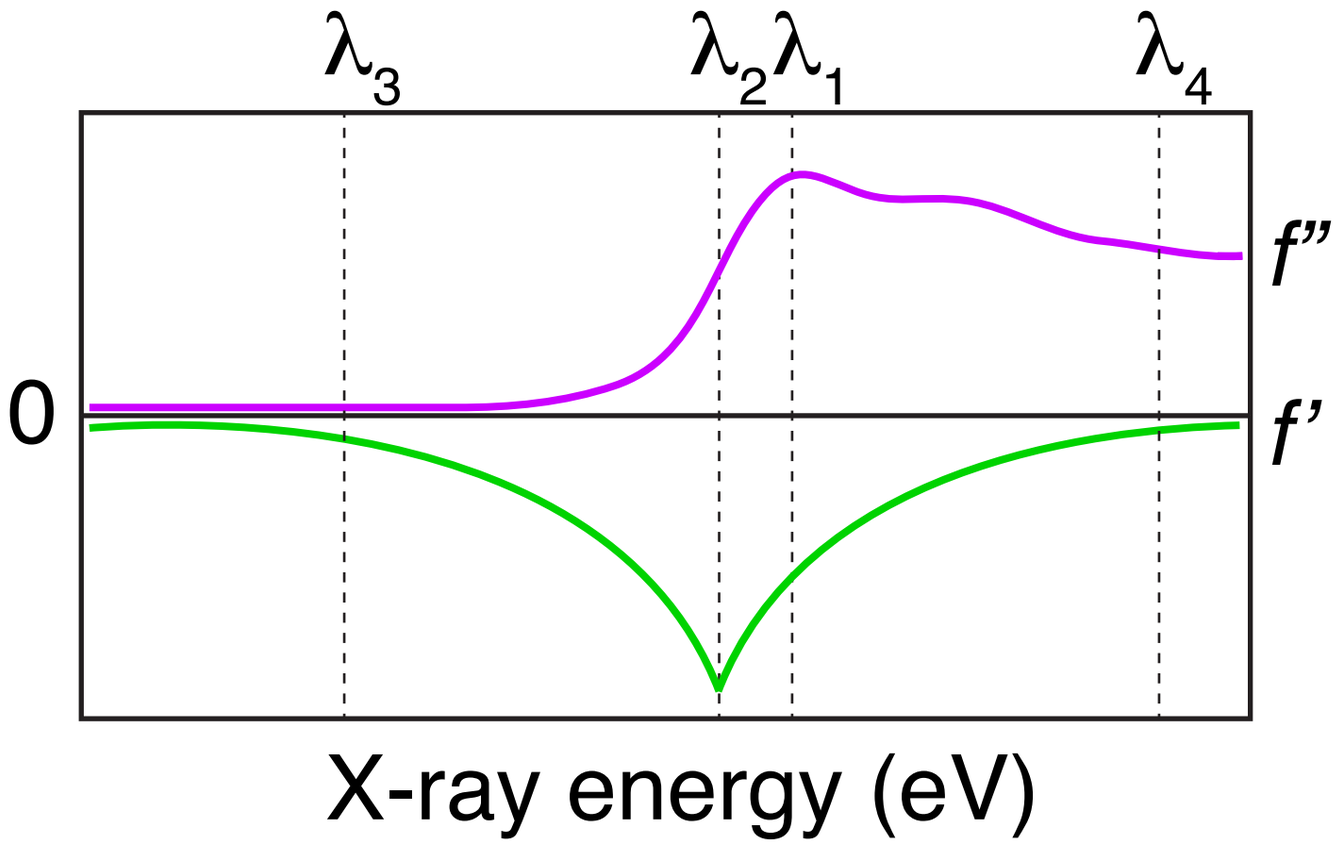
Anomalous diffraction



Magnitude of anomalous scattering factors varies with X-ray wavelength

Rhodes, G. (2006) Crystallography Made Crystal Clear, 3rd Ed. Elsevier Burlington MA.

Wavelength selection



λ_1 at maximum *f”* has largest anomalous signal (peak)

λ_2 chosen at maximum *f’* (inflection)

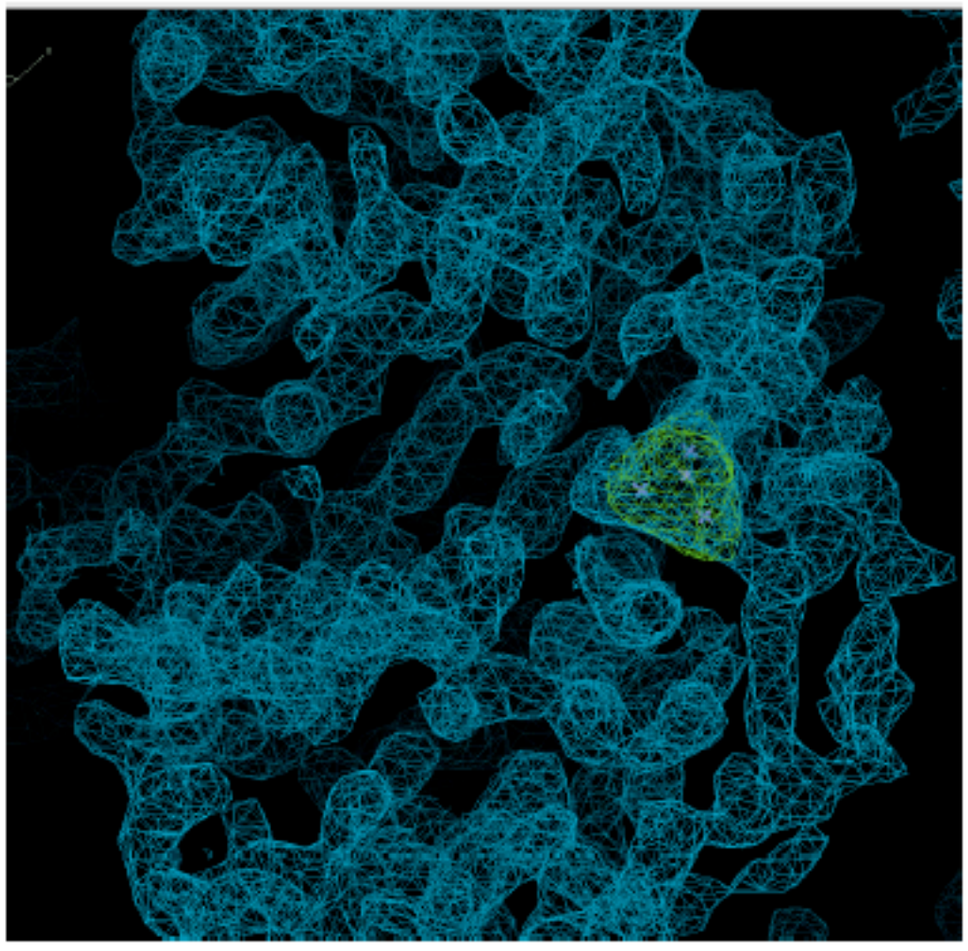
λ_3, λ_4 range between 100-1000 eV of edge (remote)

http://skuld.bmsc.washington.edu/scatter/

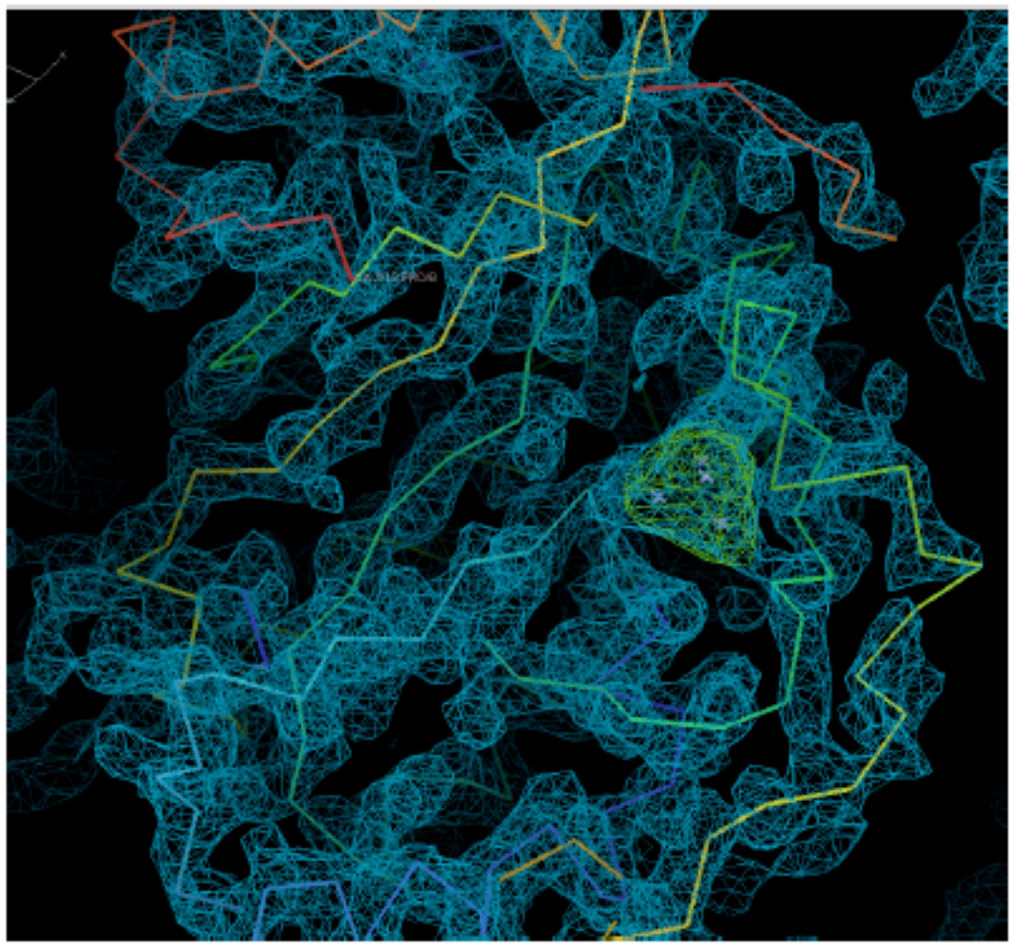
The phase problem

- Often a trial-and-error process
- Phasing methods can be combined
- Phase information is iteratively improved during model building and refinement
- How do we assess the quality of phase determination?

The phase problem



$2F_o$-F_c map (blue) = 1.4σ
Fe anomalous difference map (green) = 5.0σ



$2F_o$-F_c map (blue) = 1.4σ
Fe anomalous difference map (green) = 5.0σ

- Positive outcome is an interpretable electron density map

Model building and validation

- 3D plot of ρ(*x, y, z*) function produces electron density map
- Map interpretation
- Refinement of coordinates
- Validation of final model

phases

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l |F_{hkl}| e^{-2\pi i(hx_j+ky_j+lz_j-\alpha'_{hkl})}$$

amplitudes (from intensities)

Electron density maps

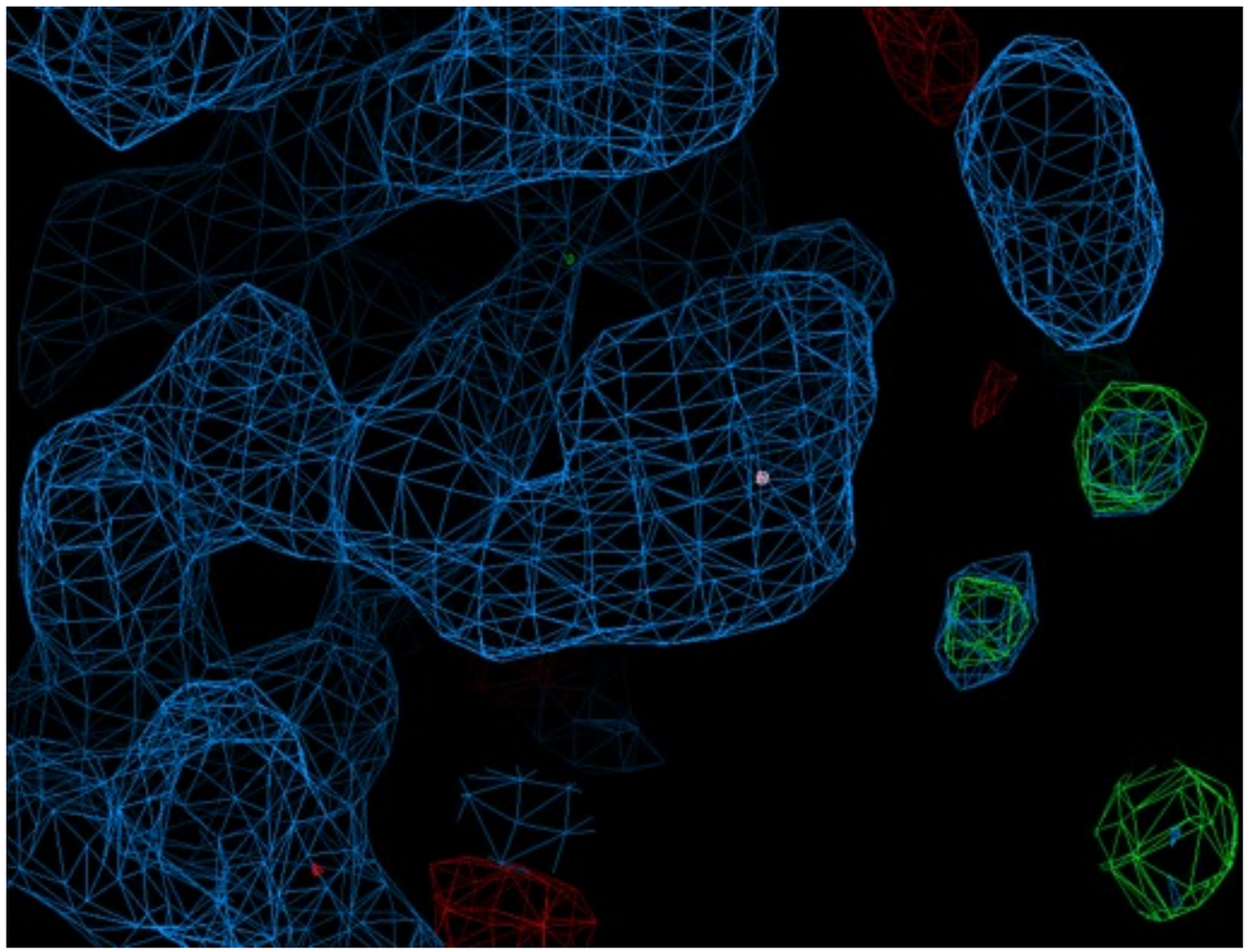
$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l |F_{hkl}|e^{-2\pi i(hx_j+ky_j+lz_j-\alpha'_{hkl})}$$

F_{obs} from diffraction data

F_{calc} from molecular model

- $2F_o$-F_c electron density map
  - Single color molecular surface map
- F_o-F_c electron density map
  - Difference map highlighting errors in the model

Map Interpretation

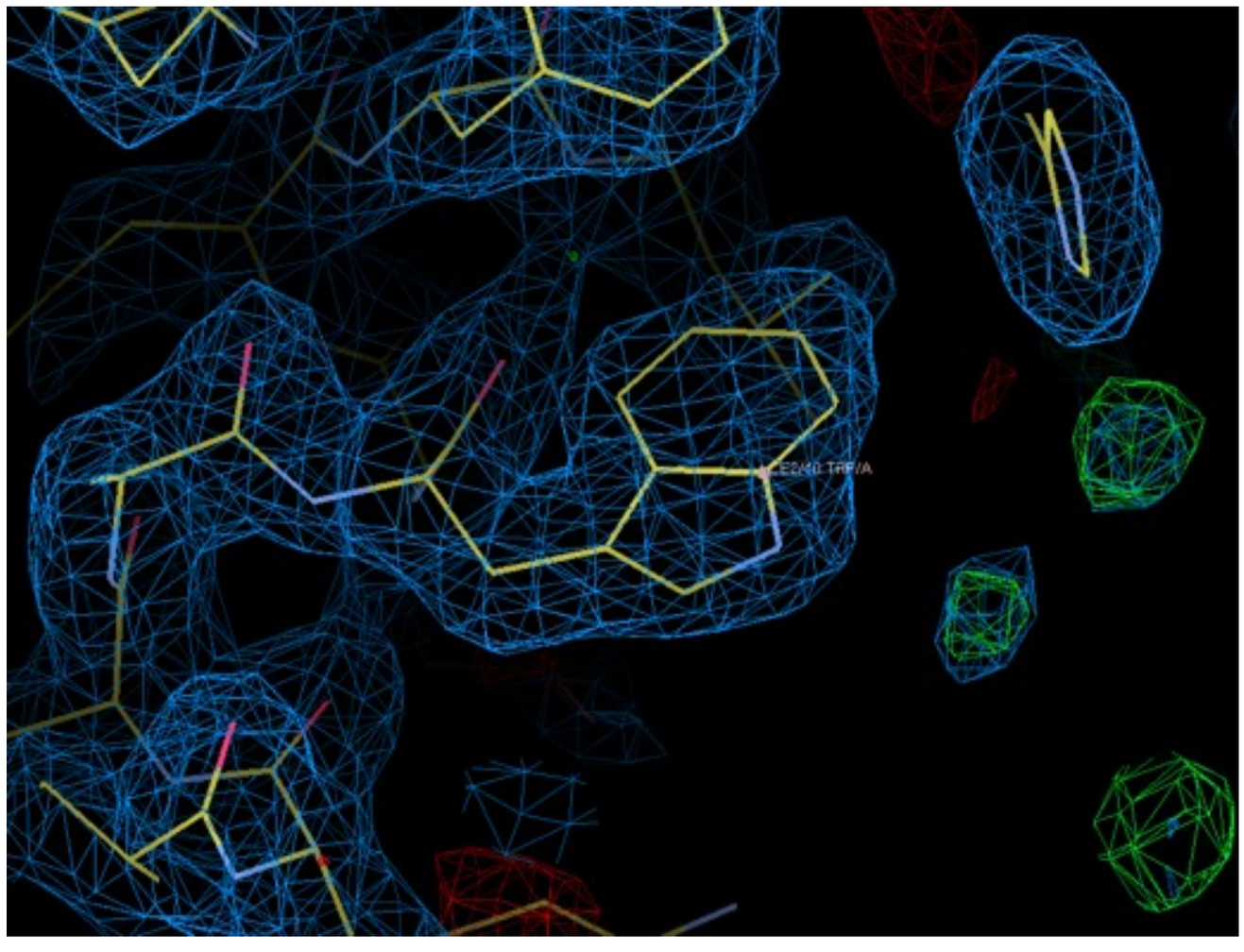


Blue = $2F_o$-F_c map

Red/green = F_o-F_c map

1.8 Å resolution map

Map Interpretation



Blue = $2F_o$-F_c map

Red/Green = F_o-F_c map

1.8 Å resolution map

Crystal structure refinement

- Computer-assisted improvement of the agreement between model and diffraction intensity data
- Least-squares/maximum likelihood methods
- Refine model parameters to improve fit of F_{calc} to F_{obs}
  - Atom positions (x_j, y_j, z_j)
  - Temperature factors (B)
  - Occupancy

Model evaluation

- **How well does the model predict the X-ray data?**
  - *R*-factor
  - R_{free} statistic

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

- **How well does the model conform to known geometric parameters?**
  - RMS deviation (bond lengths, angles)
  - Ramachandran statistics
  - Molprobity server

Model quality is a local property

Refinement statistics

	WT
Data collection statistics	
Maximum resolution (Å)	1.50
Wavelength (Å)	0.9918
Total reflections	98,554
Unique reflections	27,758
Completeness (%)[a]	98.9 (97.1)
$I/\sigma(I)$	7.8 (1.5)
R_{sym} (%)[b]	5.1 (40.4)
Refinement statistics	
Resolution limits (Å)	31.5–1.50
R-factor[c]	0.184
R-free	0.216
Estimated coordinate error (Å)[d]	0.09
RMS deviations from ideal values	
Bond lengths (Å)	0.024
Bond angles (°)	2.192
Dihedral angles (°)	25.11
Improper torsion angles (°)	1.60
Average temperature factor (Å^2)	
Protein	22.1, 20.6
Iron-sulfur	13.0, 12.8
Water	36.8
Zinc	37.1
Sulfate	44.3
Ramachandran plot,[e] residues in	
Most favored regions (%)	90.2
Additional allowed regions (%)	9.1
Generously allowed regions (%)	0.6
Disallowed regions (%)	0.0

→ *R* should be between 0.1-0.3 and within 0.05 of R_{free}

Yeh, A.P., Ambroggio, X.I., Andrade, S.L.A., Einsle, O., Chatelet, C., Meyer, J., and Rees, D.C. (2002) *J. Biol. Chem. 277*, 34499-34507.

Model evaluation

- **How well does the model predict the X-ray data?**
  - *R*-factor
  - R_{free} statistic

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

- **How well does the model conform to known geometric parameters?**
  - RMS deviation (bond lengths, angles)
  - Ramachandran statistics
  - Molprobity metrics and other validation tools

Model quality is a local property

Model validation resources

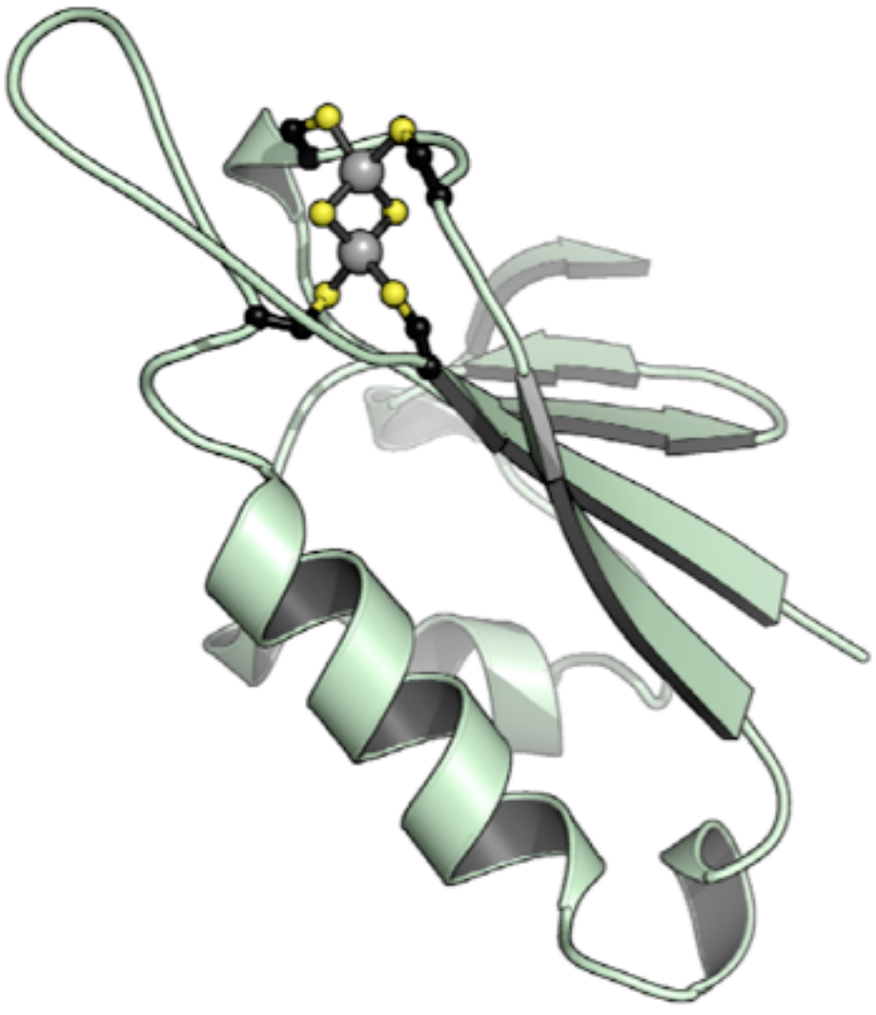
- Molprobity server
  - molprobity.biochem.duke.edu/
- PDB Validation Server
  - validate.rcsb.org/
- Other resources
  - www.rcsb.org/pdb/static.do?p=software/software_links/analysis_and_verification.html

Crystallography workflow

- Biological sample production
- Crystallization
- X-ray diffraction data collection
- Structure solution/phase determination
- Model building and validation

A preview of the tutorial

Complete de novo structure determination of a thermophilic [2Fe-2S] ferredoxin by MAD phasing



Diffraction data processing

Derive heavy atom positions

Use automated phasing pipeline

Model building and refinement

Oliver Einsle, Amie Boal, Andrew Mitchell