



COMPUTATIONAL CRYSTALLOGRAPHY INITIATIVE

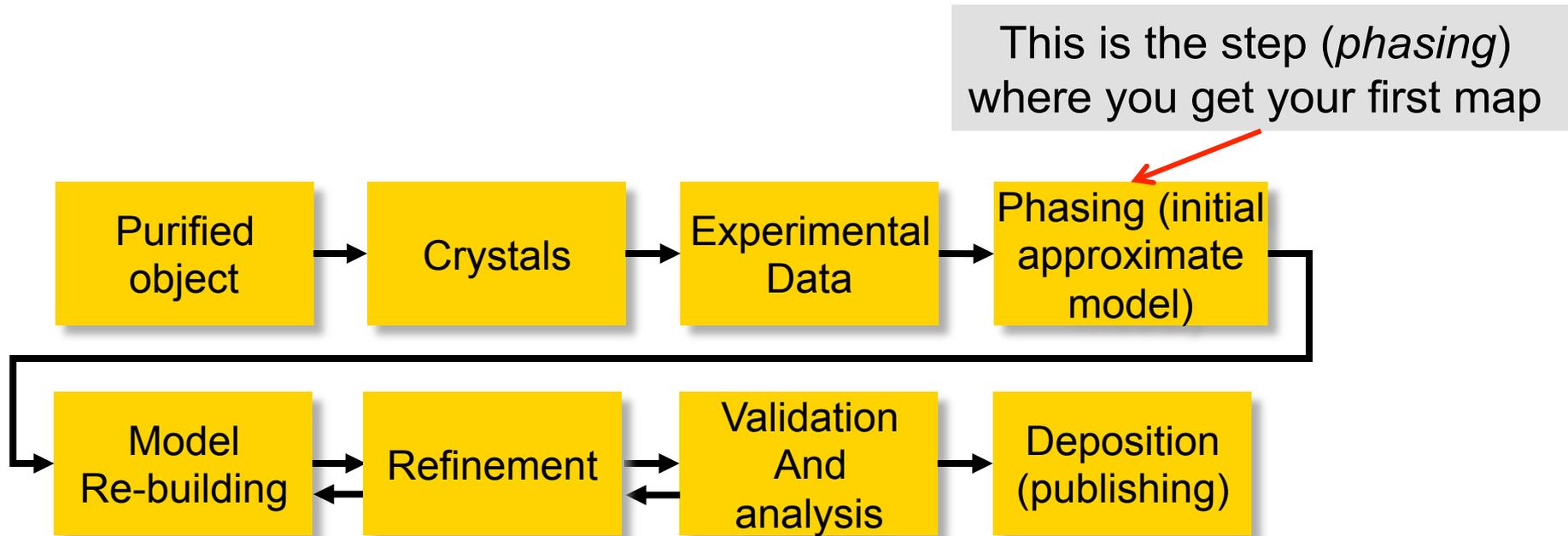
Some Facts About Maps

Pavel Afonine

Computational Crystallography Initiative
Physical Biosciences Division
Lawrence Berkeley National Laboratory, Berkeley CA, USA

First map

Crystallographic structure determination workflow and your first map



When I say *map*...

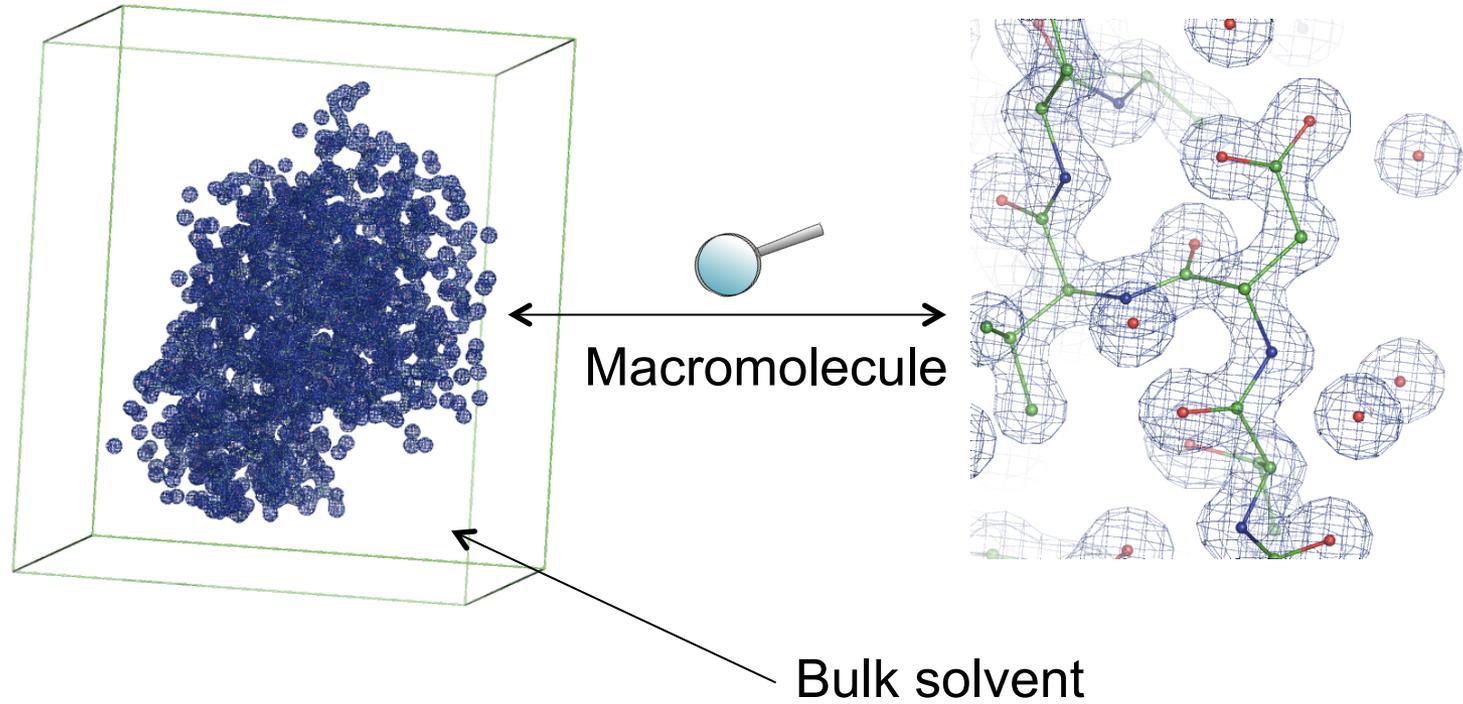
...I mean *Fourier map*, not *electron density map*

- **Electron density**
- **Fourier map (Fourier image or Fourier synthesis)**
- **Map**

Terminology

- **Electron density**
- **Fourier map (Fourier image or Fourier synthesis)**
- **Map**

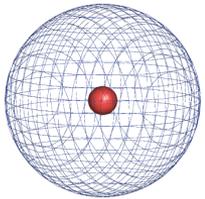
Crystal model: Electron density distribution



$$\rho_{\text{crystal}} = \rho_{\text{atoms}} + \rho_{\text{bulk solvent}}$$

Electron density distribution

- A Gaussian function approximates well electron density of an atom
 - Convenient computationally (Fourier transform of a Gaussian is a Gaussian)
- *Isotropic* distribution of electron density at the point \mathbf{r} of an isolated atom located at position \mathbf{r}_0 and having B-factor B and occupancy q :

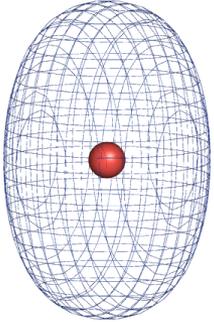


$$\rho_{atom}(\mathbf{r}, \mathbf{r}_0, B, q) = q \sum_{k=1}^5 a_k \left(\frac{4\pi}{b_k + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r} - \mathbf{r}_0|^2}{b_k + B} \right)$$

- Number of terms in the above formula depends on how accurately we want to model an atom
- a_k and b_k are atom-specific, tabulated values

Atom density

- More accurate approximation assumes atoms moving anisotropically

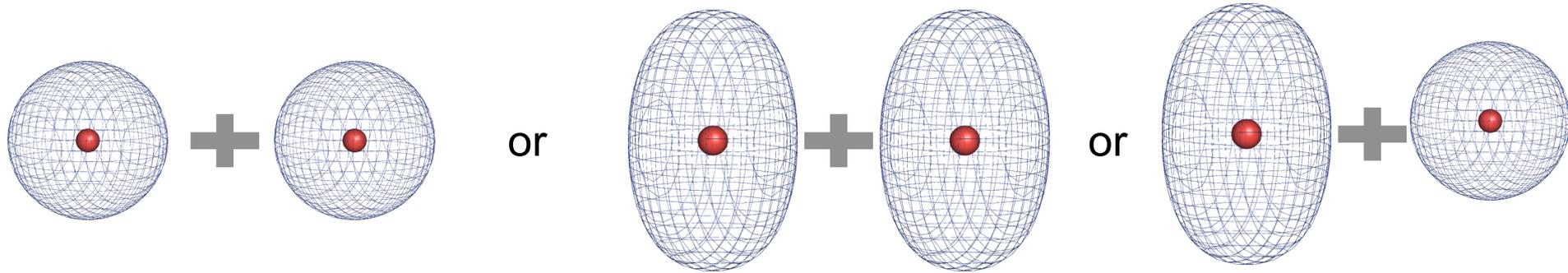


$$\rho_{atom}(\mathbf{r}, \mathbf{U}, q) = q \sum_{j=1}^5 \frac{q a_j (4\pi)^{3/2}}{|8\pi^2 \mathbf{U}_{cart} + b_j \mathbf{I}|^{1/2}} \exp\left(-4\pi^2 (\mathbf{r} - \mathbf{r}_0)^T \mathbf{A}^T [8\pi^2 \mathbf{U}_{cart} + b_j \mathbf{I}]^{-1} \mathbf{A} (\mathbf{r} - \mathbf{r}_0)\right)$$

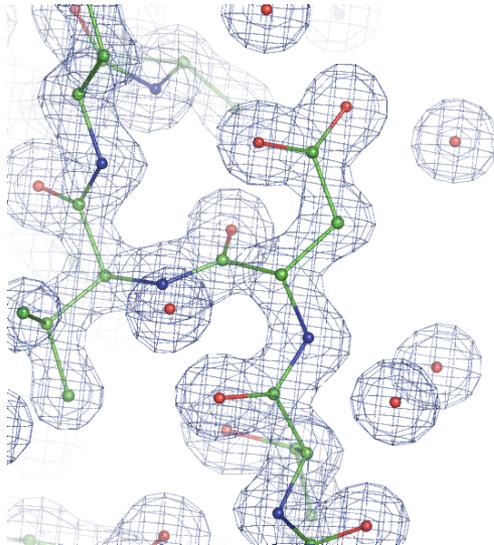
- \mathbf{U}_{cart} – anisotropic atomic displacement parameters (3*3 symmetric matrix).
- \mathbf{U}_{cart} is what is in ANISOU records of PDB files

Atom density

- Electron density of whole molecule is a sum of electron densities of individual atoms (isotropic or anisotropic or mixed)



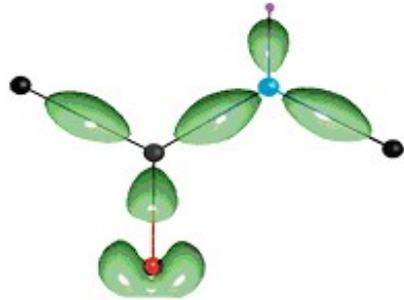
- Bonding effects are ignored (atoms isolated): IAM – individual atom model



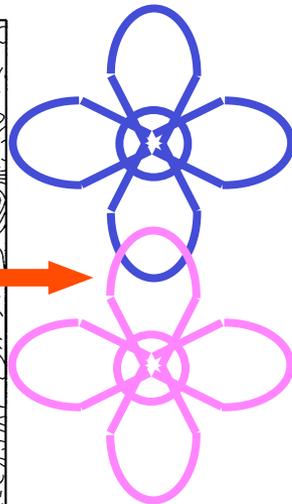
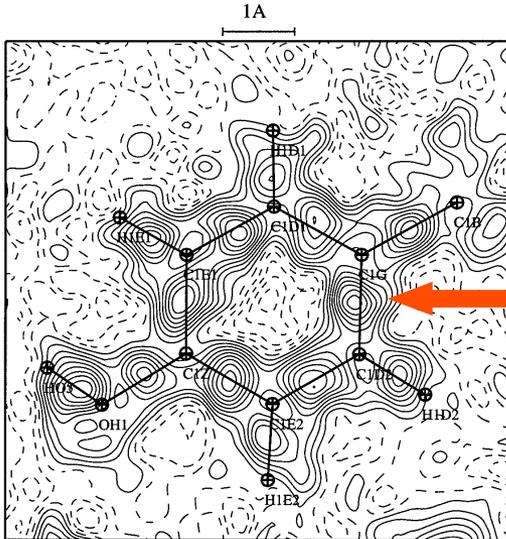
$$\rho_{crystal}(\mathbf{r}) = \sum_{i=1}^{N_{atoms}} \rho_{atoms}(\mathbf{r})$$

Multipolar density model

- Even more more accurate approximation assumes atoms are bonded: multipolar model (Dawson, 1967; Stewart, 1969; Hansen & Coppens, 1978)



$$\rho_{\text{atom}}(\mathbf{r}) = \rho_{\text{core}}(\mathbf{r}) + P_{\text{val}} \kappa^3 \rho_{\text{val}}(\kappa \mathbf{r}) + \sum_{l=0}^{l_{\text{max}}} \kappa'^3 R_l(\kappa' \mathbf{r}) \cdot \sum_{m=-l}^l P_{lm} y_{lm}(\theta, \varphi)$$



ρ_{ATOM} = core electrons
+ valence electrons
+ non-spherical part of the valence electron distribution

- This kind of electron density model is only used at ultra-high resolution (1Å and higher)

Electron density – structure factor relationship

$$\rho_{crystal}(\mathbf{r}) = \sum_{i=1}^{N_{atoms}} q_i \sum_{k=1}^5 a_k \left(\frac{4\pi}{b_k + B_i} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r} - \mathbf{r}_{0,i}|^2}{b_k + B_i} \right)$$



Fourier transformation

Set of structure factors $\{\mathbf{F}(\mathbf{s})\}$, where each one is:

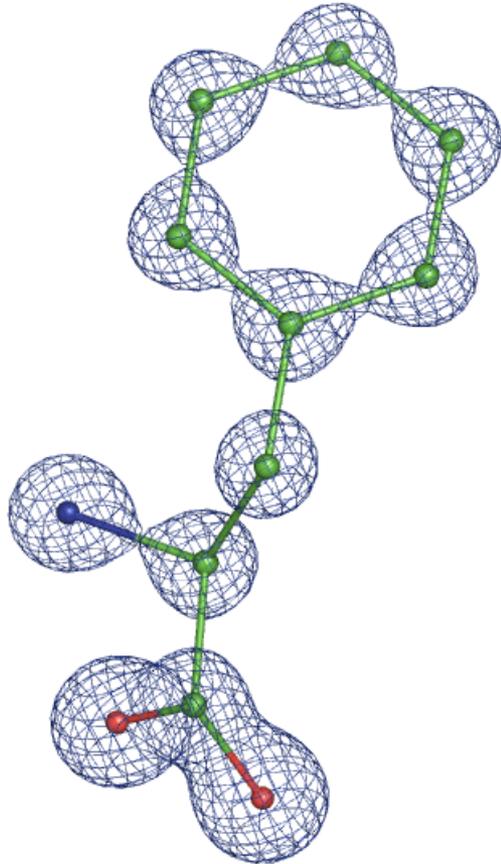
$$\mathbf{F}(\mathbf{s}) = \sum_{i=1}^{N_{atoms}} q_i \sum_{k=1}^5 a_k \exp\left(-\frac{b_k s^2}{4} \right) \exp(-B_i s^2 / 4) \exp(2i\pi \mathbf{r}_i \mathbf{s})$$

- All atoms contribute to each structure factor
- Two very important for computations features of FT:
 - Fourier transform of a Gaussian function is a Gaussian function
 - Fourier transform is a linear operation

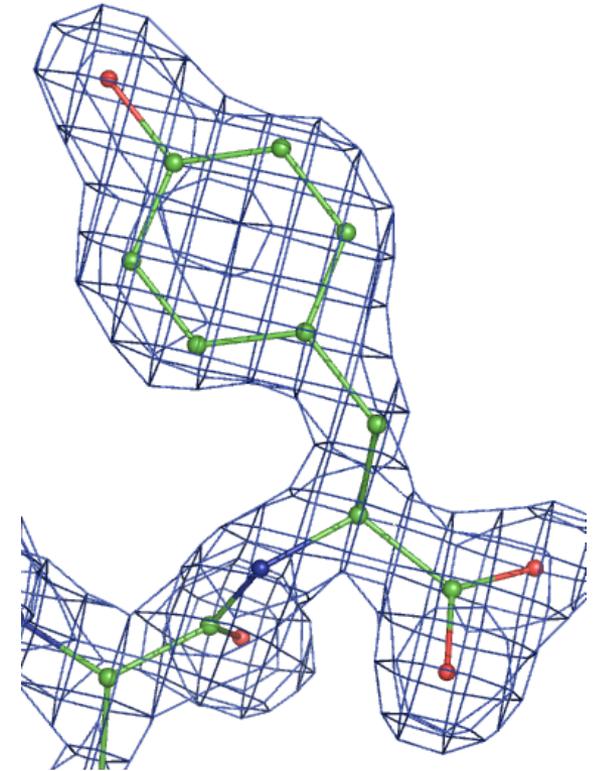
Electron density – structure factor relationship

Electron density computed by formula

$$\rho_{crystal}(\mathbf{r}) = \sum_{i=1}^{N_{atoms}} q_i \sum_{k=1}^5 a_k \left(\frac{4\pi}{b_k + B_i} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r} - \mathbf{r}_{0,i}|^2}{b_k + B_i} \right)$$



In reality we see densities more like

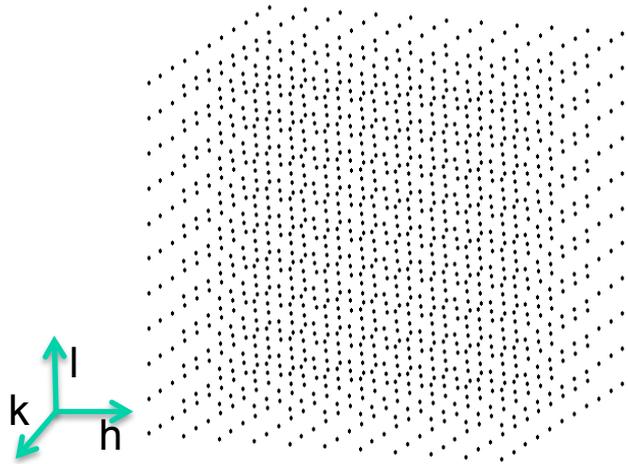


Terminology

- **Electron density**
- **Fourier map (Fourier image or Fourier synthesis)**
- **Map**

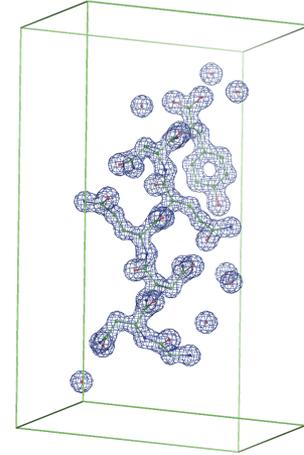
Fourier image

- Experimentally we measure structure factors: $F(\mathbf{s})$



- What we want to study is electron density distribution in unit cell:

ρ_{crystal}



- Exact relationship between ρ and F is only when all terms in the summation are present

$$F(\mathbf{s}) = \int_{V_{\text{cell}}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{s} \cdot \mathbf{r}) dV$$



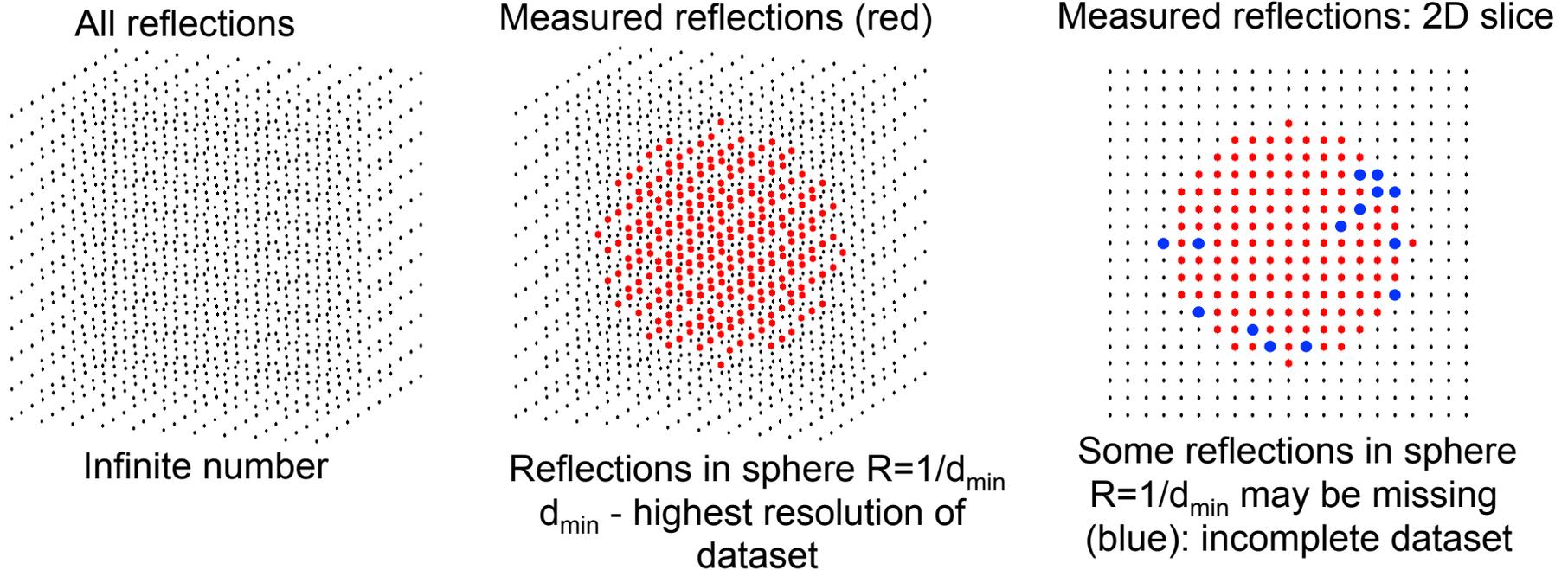
$$\rho(\mathbf{r}) = \frac{1}{V_{\text{cell}}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r})$$

Fourier image

- Exact relationship between ρ and F is only when all terms in the summation are present

$$\rho(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r}) \quad \longleftrightarrow \quad F(\mathbf{s}) = \int_{V_{cell}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{s} \cdot \mathbf{r}) dV$$

- In reality only a subset of all F is measured

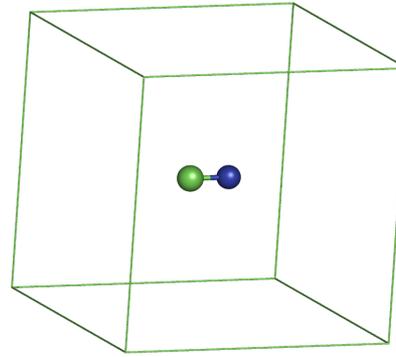


- Incomplete hkl set means density is not accurate anymore: Fourier image of finite resolution

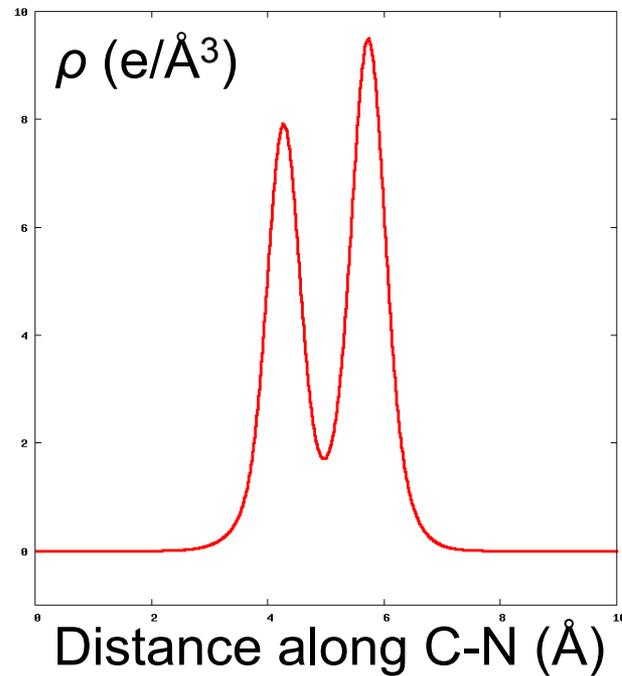
$$\rho(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r}) \quad \longrightarrow \quad \rho_{image}(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h_{min}}^{h_{max}} \sum_{k_{min}}^{k_{max}} \sum_{l_{min}}^{l_{max}} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r})$$

Fourier image

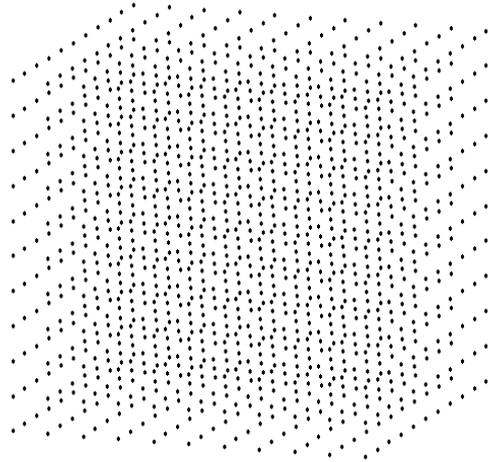
Toy example: C-N in 10*10*10Å P1 box



Electron density distribution along C-N bond vector

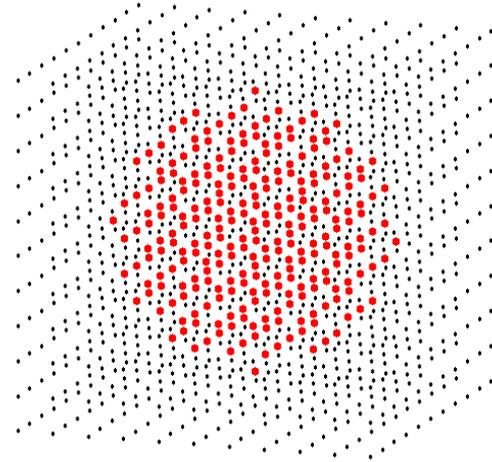


Fourier image



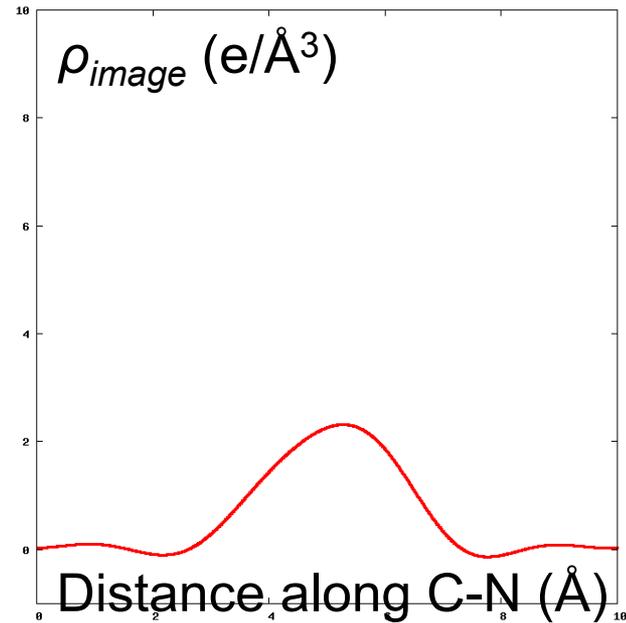
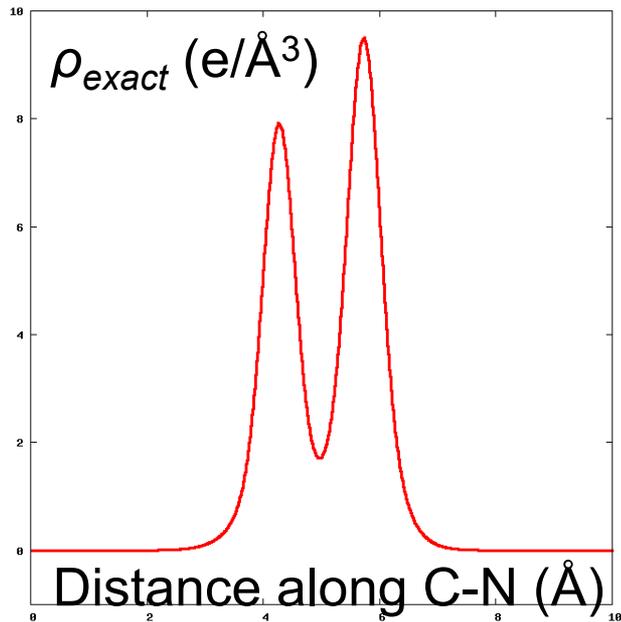
Exact density

$$\rho(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r})$$



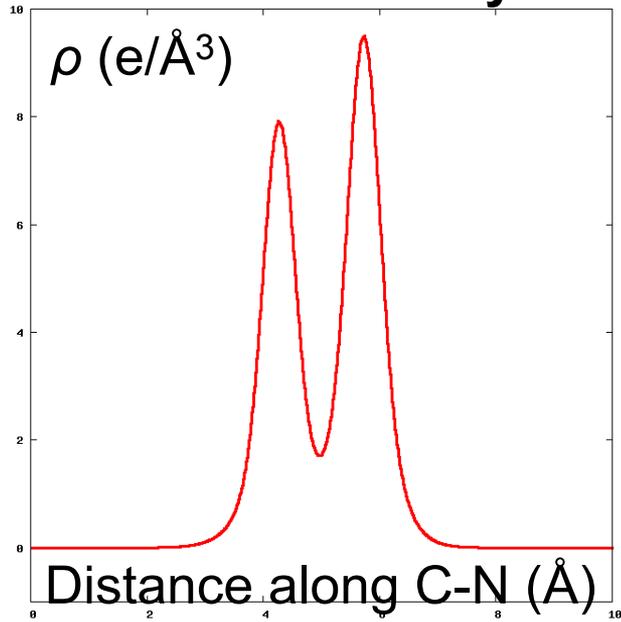
2 Å resolution Fourier image

$$\rho_{image}(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h_{min}}^{h_{max}} \sum_{k_{min}}^{k_{max}} \sum_{l_{min}}^{l_{max}} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r})$$

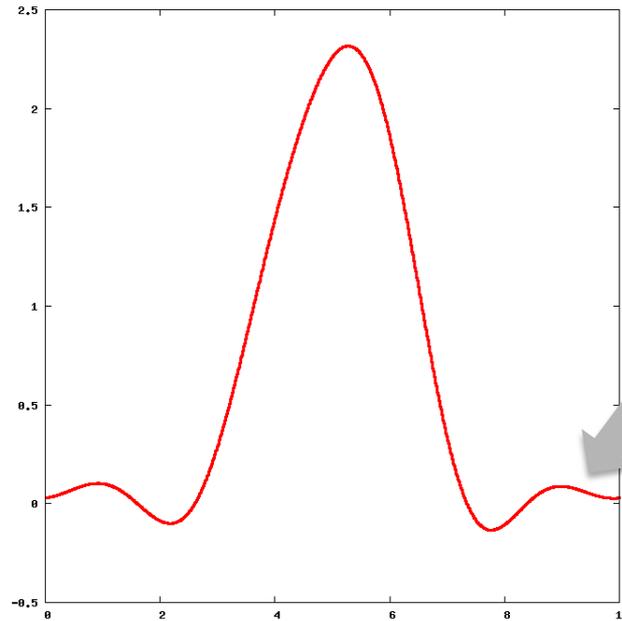
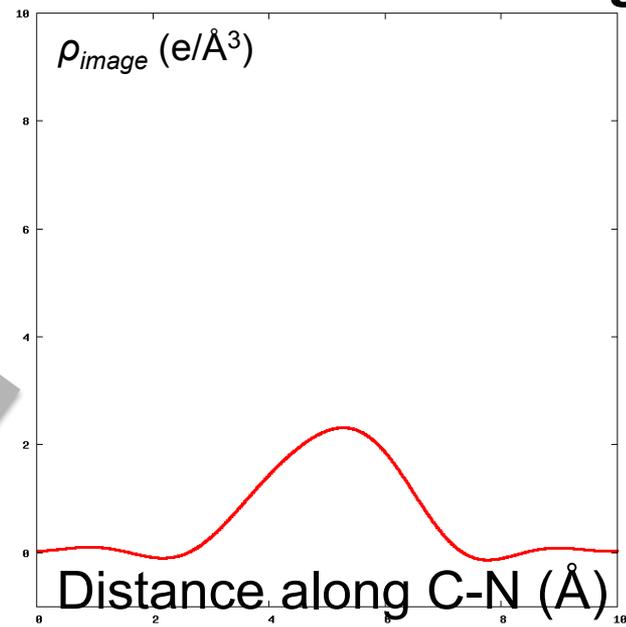


Fourier image

Exact density



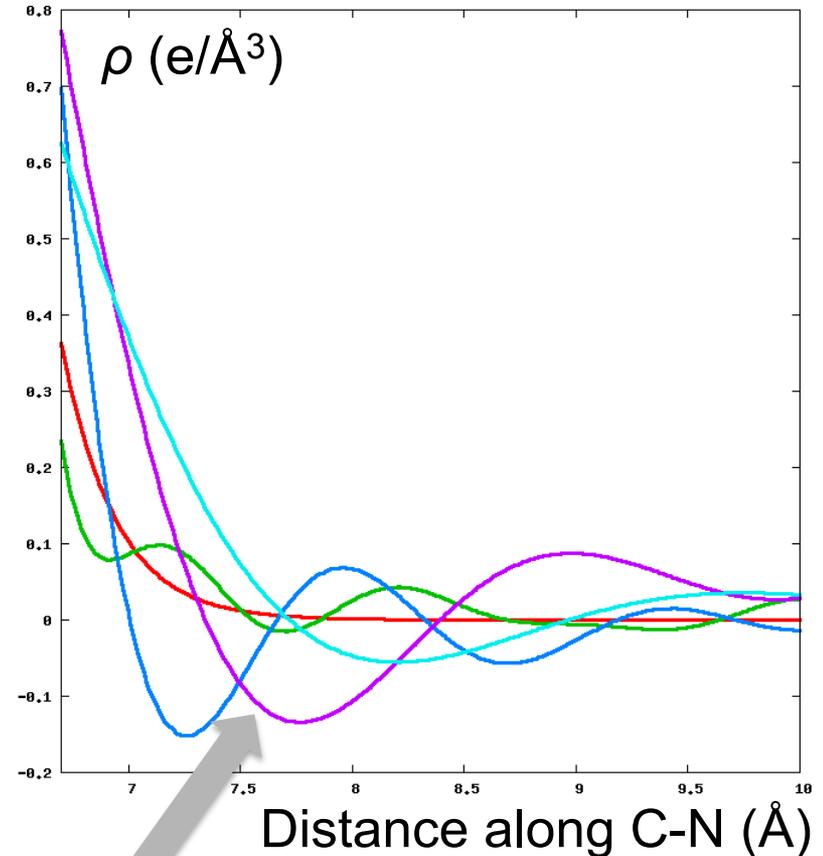
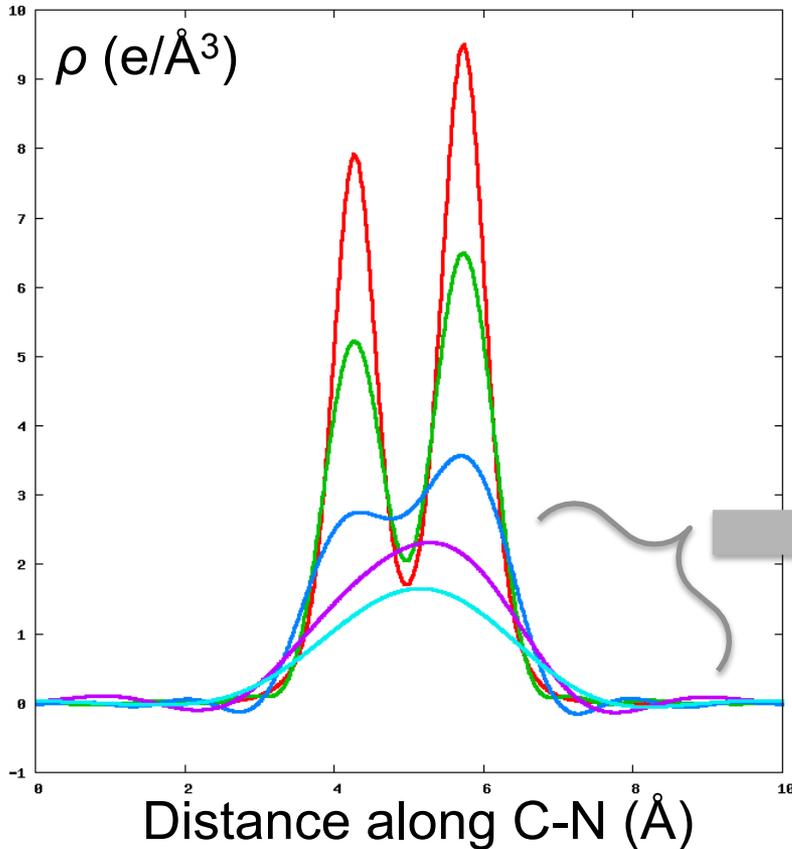
2 Å resolution Fourier image



Positive and negative spurious peaks – Fourier truncation ripples (artifacts)

Fourier image

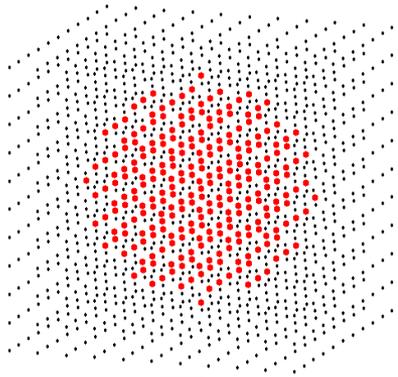
More examples: **exact** density (red) and **1**, **1.5**, **2** and **2.5 Å** resolution Fourier images



Positive and negative spurious peaks – Fourier truncation ripples (artifacts)

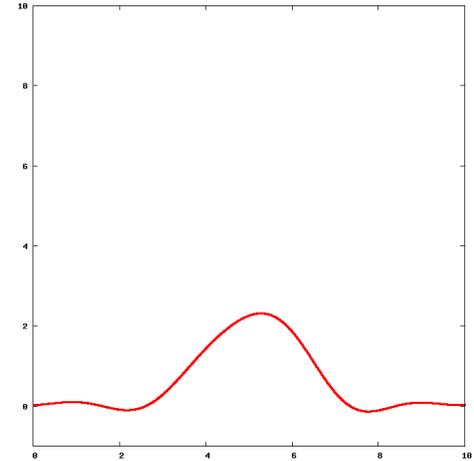
Inverse problem in crystallography

2 Å resolution set of F



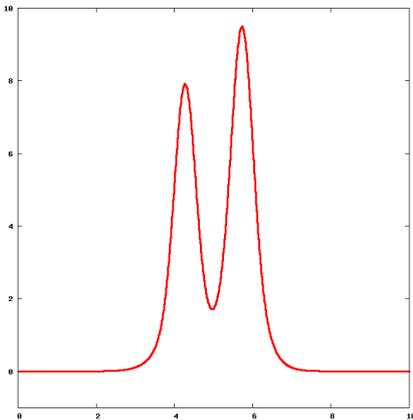
$$\rho_{image}(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h_{min}}^{h_{max}} \sum_{k_{min}}^{k_{max}} \sum_{l_{min}}^{l_{max}} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \mathbf{r})$$

2 Å resolution Fourier image



Exact density

$$\rho(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \mathbf{r})$$



- **Inverse problem:** converting observed measurements into information about a physical object.

- Ill-posed problem in crystallography: we want to reconstruct image damaged due to finite amount of measured data

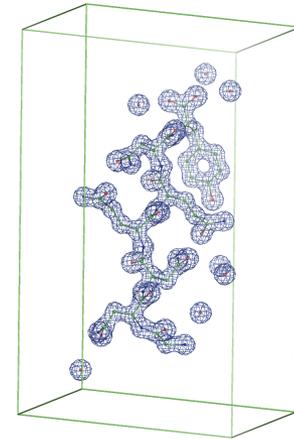
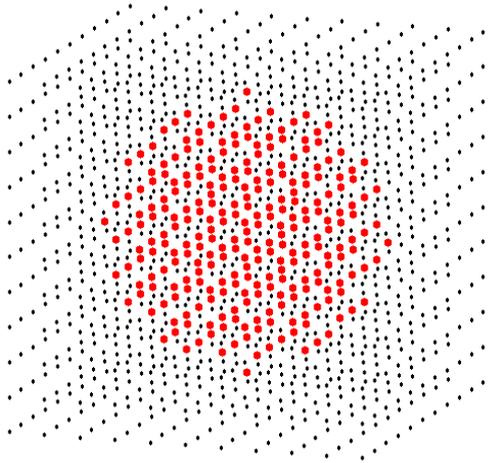
- Regularization involves introducing additional information in order to solve an ill-posed problem

- It involves encoding prior knowledge in terms of constraints on the solution space like positivity or smoothness for example.

- Example: density is positive and total charge F000

Inverse problem in crystallography

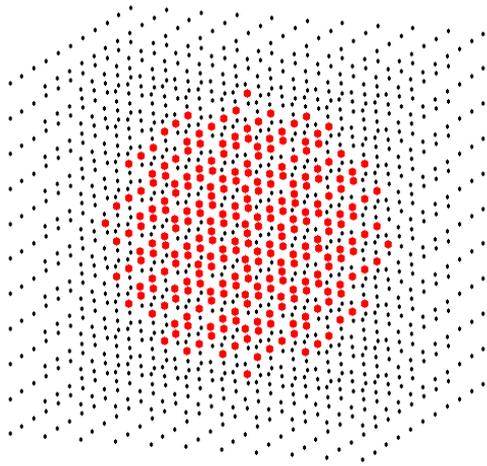
$$\rho(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r})$$



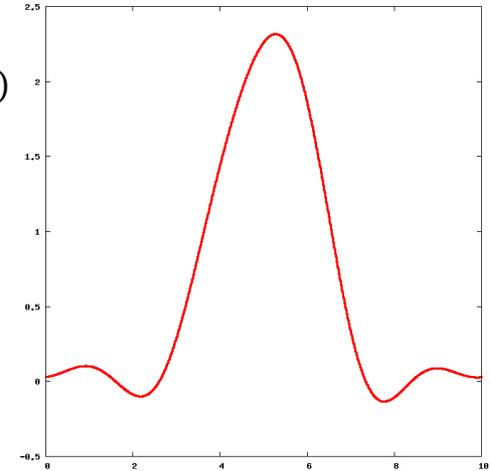
$$F(\mathbf{s}) = \frac{V_{cell}}{N_x N_y N_z} \sum_{j_x=0}^{N_x-1} \sum_{j_y=0}^{N_y-1} \sum_{j_z=0}^{N_z-1} \rho(j_x, j_y, j_z) \exp(2\pi i [h j_x + k j_y + l j_z]) \quad (\text{Sayre, 1951})$$

- One can iterate back and forth any number of times – this will not change F or ρ
- Values of F calculated for Miller indices that were not used in calculation of ρ are always equal to zero

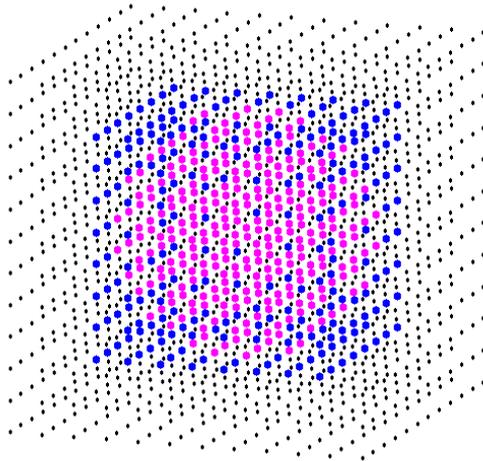
Inverse problem in crystallography



$$\rho(\mathbf{r}) = \frac{1}{V_{cell}} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(\mathbf{s}) \exp(-2\pi i \mathbf{s} \cdot \mathbf{r})$$

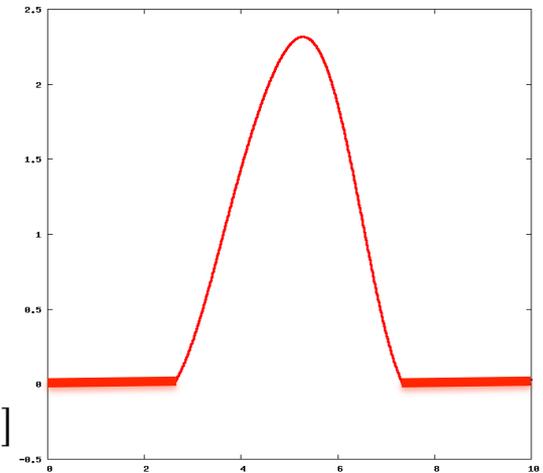


Modify density
in some way



New indices are in box:
 $|h| < N/2, |k| < N/2, |l| < N/2$
N – map gridding

$$F(\mathbf{s}) = \frac{V_{cell}}{N_x N_y N_z} \sum_{j_x=0}^{N_x-1} \sum_{j_y=0}^{N_y-1} \sum_{j_z=0}^{N_z-1} \rho(j_x, j_y, j_z) \exp(2\pi i [h j_x + k j_y + l j_z])$$

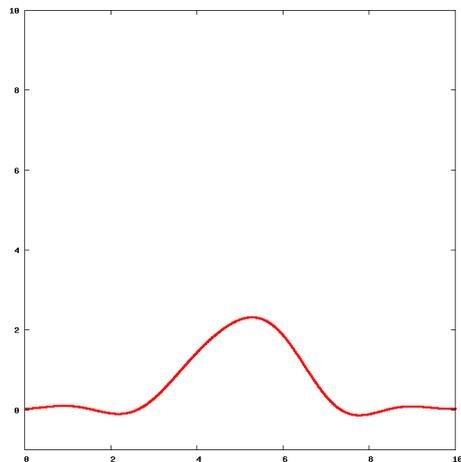
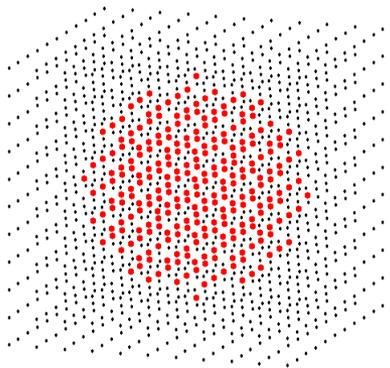


- This is a foundation for regularization methods that in crystallography called Density Modification

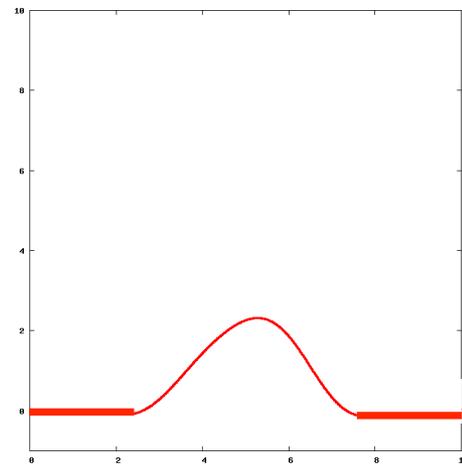
Addressing inverse problem in crystallography: density modification

- Density constraints to improve phases and extend resolution:
 1. Atomicity (Hoppe & Gassmann, 1964)
 2. Positivity (Barrett & Zwick, 1971)
 3. Noncrystallographic symmetry (Bricogne, 1974)
 4. Solvent flatness (Bricogne, 1974)
 5. Map connectivity (continuity) (Bhat & Blow, 1982)
 6. Histogram matching (Lunin, 1988)
 7. MEM (maximum entropy methods) (Collins, 1982)
- Good reviews:
 - Classical density modification:
 - Podjarny, Rees & Urzhumtsev, 1996;
 - Cowtan, 2012;
 - Statistical density modification:
 - A series of papers in Acta Cryst by Tom Terwilliger

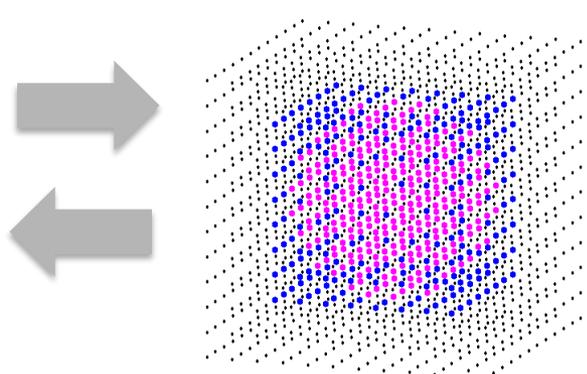
Inverse problem in crystallography



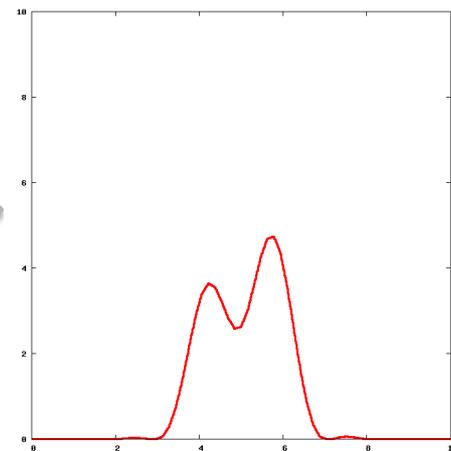
Original map



Modified map

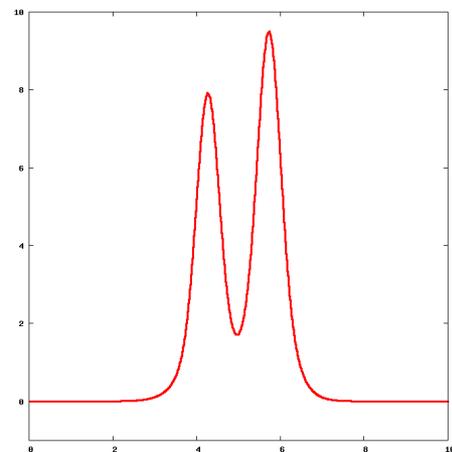


New set of reflections



Improved map

This is may still be far from ideal density



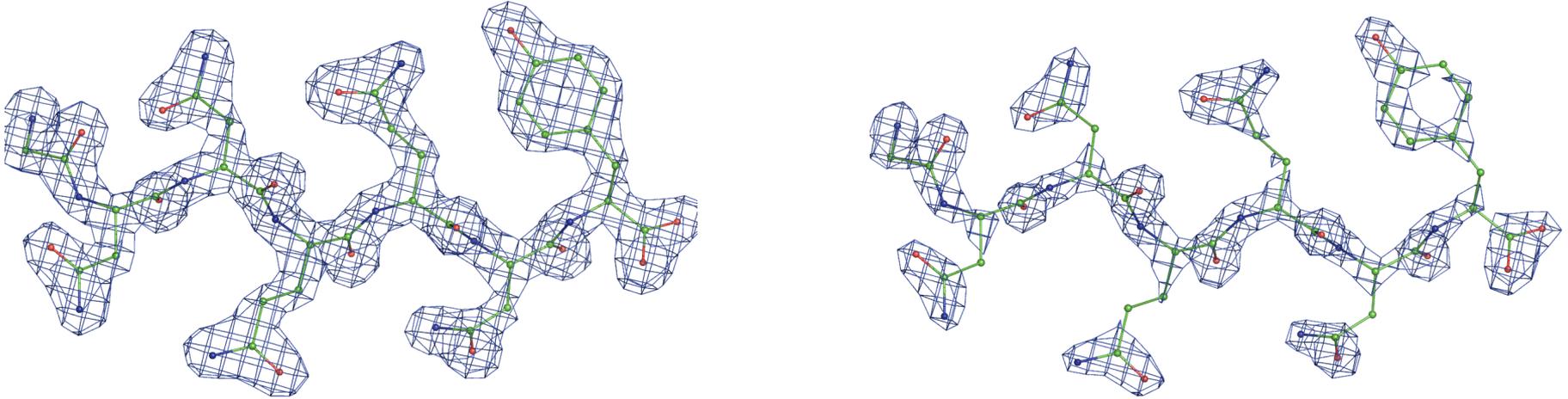
Terminology

- **Electron density**
- **Fourier map (Fourier image or Fourier synthesis)**
- **Map**

Displaying Fourier map

- What we look at are never electron density maps but its Fourier images

Same Fourier maps contoured at 1σ (left) and 2.5σ (right)



- Showing a map does not make much sense unless the following is specified:
 - resolution (and completeness) of corresponding set of miller indices
 - contouring level
 - units

Map scaling

- Fourier transform of structure factors gives map in arbitrary units
- Map needs to be scaled:
 - Absolute scale ($e/\text{\AA}^3$)
 - Divide map values by unit cell volume, and
 - Add $F(0,0,0)$ reflection, which is total charge of the unit cell
 - $F(0,0,0)$ is never measured in diffraction experiment but estimated from unit cell content
 - Also can be estimated from measuring crystal density
 - Convenient as there is a reference point: mean solvent density $0.35e/\text{\AA}^3$
 - Scale by standard deviation (sigma, σ):
 - Calculate map standard deviation and divide map values by it
 - Widely (almost exclusively) used
 - Difficult to compare maps since standard deviation depends on map values

Map output formats

- Actual map (3D function computed on a grid)
 - X-plor/CNS format
 - Text file (typically huge size), longer to load
 - Can be visualized in PyMol
 - CCP4 format
 - Binary file format (smaller size), faster to load
 - Can be used with Coot or PyMol
- File with Fourier map coefficients
 - Typically in MTZ format (other formats exist)
 - Small files, quick to load, may contain several “maps”
 - Graphics programs, like Coot, will do Fourier transform internally in order to produce the map
 - No control over how the map is scaled

Model bias

$$\rho_{crystal}(\mathbf{r})$$



Fourier transformation

Set of structure factors $\{\mathbf{F}(\mathbf{s})\} = \{F(\mathbf{s}), \text{phase}(\mathbf{s})\}$

If *phases* come from the model – fundamental problem: model bias.

Model bias

- **Model bias**

- Results from using atomic model to calculate phases

- Map tends to have features present in the model even if they are not actually present in the structure (Ramachandran & Srinivasan, 1961; Read, 1986; Bhat, 1988; Hodel et al., 1992; Adams et al., 1999; Kleywegt, 2000).

- **Explanation in a nutshell**

Once an atomic model has been refined, the positions and other parameters describing correctly placed atoms are adjusted during refinement in order to compensate for the incorrectly placed atoms.

Consequently, even if the incorrectly placed atoms are removed from the model before the calculation of phases, a memory of their positions can remain and the resulting map can retain incorrect features.

Model bias

- **Methods and tools to reduce model bias**
 - σ_A -map: $2mF_{OBS} - DF_{MODEL}$ (Read, 1986; Urzhumtsev et al., 1996)
 - OMIT map (Bhat, 1988)
 - Simulated-annealing OMIT maps (Hodel et al., 1992; Brunger et al., 1998)
 - ‘kicked’ OMIT maps (Guncar et al., 2000)
 - Model rebuilding with randomization (Zeng et al., 1997; Reddy et al., 2003)
 - Prime-and-switch density modification (Terwilliger, 2004)
 - Carry out the usual model building and refinement avoiding a specific model part, such as ligand
 - ‘ping-pong refinement’ (Hunt & Deisenhofer, 2003)
- **Most of the above methods may or may not remove the bias completely**
- **The map may be of worse overall quality**
 - omitting a piece of model is unlikely to make the model better

Model bias

- **The most efficient method of removing bias**

- “Iterative-build OMIT procedure” removes bias completely and results in good map
- Available in PHENIX only
- May take a while to run

Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. Acta Cryst. (2008). D64, 515–524. Terwilliger et al.

- **Typical mistake:**

- remove a piece of model in question, compute a map and expect it unbiased
- in fact, such a map may serve as a nice example of model bias!

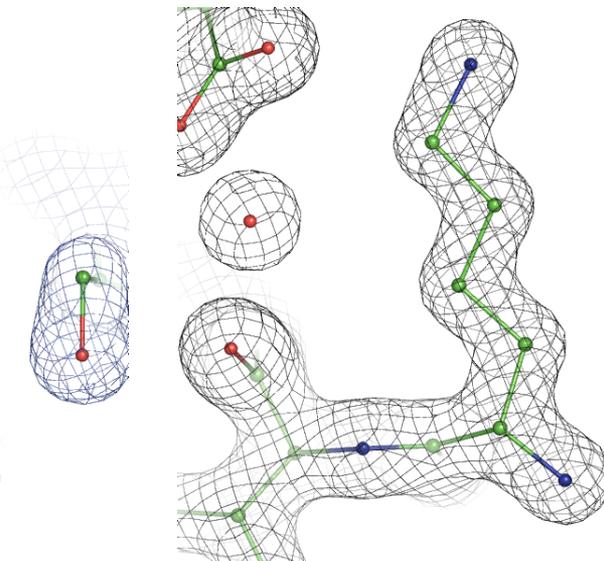
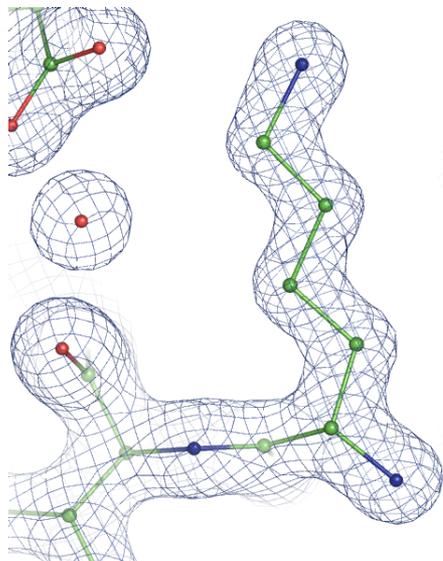
$\{2F_{\text{OBS}} - F_{\text{MODEL}}, \text{phase}\}$ versus $\{F_{\text{OBS}}, \text{phase}\}$

- **Why $\{2F_{\text{OBS}} - F_{\text{MODEL}}, \text{phase}\}$ and not simply $\{F_{\text{OBS}}, \text{phase}\}$?**
 - $\{F_{\text{OBS}}, \text{phase}\}$ shows missing features in half-size

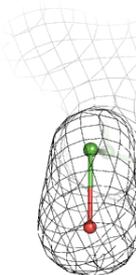
Direct maps $\{F_{OBS}, \text{phase}\}$ - show unmodelled parts at half their density

2Fo-Fc

Fo: complete model, Fc and phase: Lys omit
 3σ

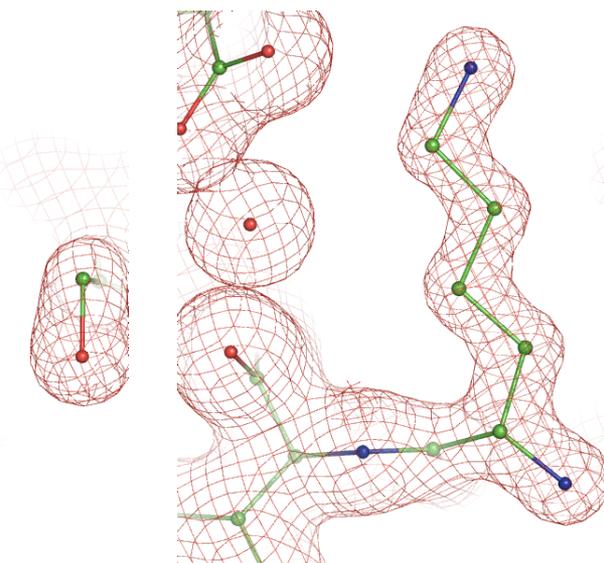
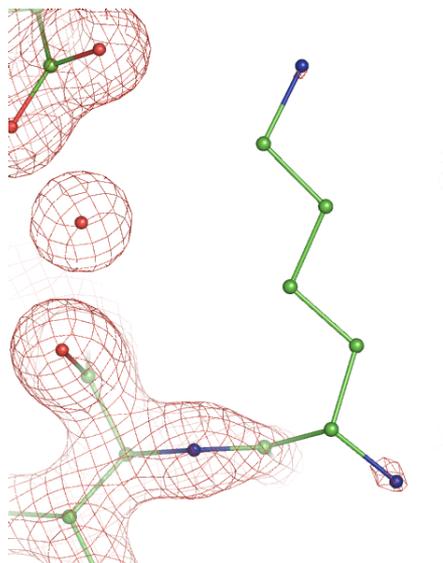


Fc-map, complete model
 3σ

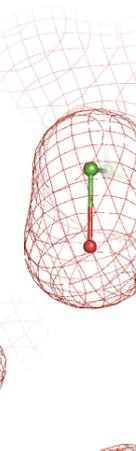


Fo,

Fo: complete model, phase: Lys omit
 3σ



Fo,
Fo: complete model, phase: Lys omit
 1.5σ



σ_A -map (sigma-a map)

- Least model biased map (σ_A -map)
 - Randy Read (1986) showed that if properly weighted $2F_{OBS}-F_{MODEL}$ map can be least model biased (σ_A -map):
 - acentric reflections: $2mFo-DFc$
 - centric reflections (*): mFo
 - m – figure of merit, D – model error and scale between F_o and F_c
 - m and D are obtained by minimization of Maximum-Likelihood function w.r.t. these parameters *using test set of reflections only* (first proposed by Urzhumtsev et al., 1996).

(*) Definition: A reflection (h,k,l) is said to be centric if in the space group there is at least one symmetry operation $g(x)=R_g*x+t_g$ whose rotational part R_g sends the reflection to minus itself.

$$R_g*(h,k,l)=(-h,-k,-l)$$

Phenix

http://www.phenix-online.org/

Apple Yahoo! Google Maps YouTube Wikipedia News (1731) Popular

Phenix

NEW [Development release of PHENIX version 1.4 now available](#)

Python-based Hierarchical ENvironment for Integrated Xtallography

PHENIX is a new software suite for the automated determination of macromolecular structures using X-ray crystallography and other methods.

Citing PHENIX:
 PHENIX: building new software for automated crystallographic structure determination P.D. Adams, R.W. Grosse-Kunstleve, L.-W. Hung, T.R. Ioerger, A.J. McCoy, N.W. Moriarty, R.J. Read, J.C. Sacchettini, N.K. Sauter and T.C. Terwilliger. *Acta Cryst.* D58, 1948-1954 (2002)

Download the latest development release (1.4-3) [First request download password]

Help: [FAQ](#) [Mailing List Subscription](#) [List Archives](#) [Report a Bug](#) [Email for Help](#)

Using PHENIX (release 1.4-3): [Full Documentation](#) [PDF](#)

- Assessing data quality with [phenix.xtriage](#)
- Automated structure solution with [AutoSol](#)
- Automated molecular replacement with [AutoMR](#)
- Automated model building and rebuilding with [AutoBuild](#)
- Automated ligand fitting with [LigandFit](#)
- Structure refinement with [phenix.refine](#)
- Generation of ligand coordinates and restraints with [elbow](#)
- The [PHENIX Graphical User Interface](#)

[Documentation for 1.3-final](#)

The PHENIX system also includes SOLVE/RESOLVE, Phaser, Textal, the CCI Applications (phenix.xtriage, phenix.refine, elbow and many more), components from Molprobit, and the Computational Crystallography Toolbox in a Python framework.

Funding for PHENIX: [Protein Structure Initiative \(NIH General Medical Sciences\)](#)

The PHENIX Industrial Consortium [Information](#)

For-profit groups can obtain access to PHENIX through a Consortium agreement. This provides a license to use PHENIX and research funds to develop new features in PHENIX tailored to the needs of commercial users. [Members](#) [Download](#) [Contact Us](#)

Groups developing PHENIX:

[Paul Adams](#) [Randy Read](#) [Jane & Dave Richardson](#) [Tom Terwilliger](#) [Tom Ioerger & Jim Sacchettini](#)

[Privacy and Security Notice](#) [About this website](#)

This presentation (PDF file) and much more

Acknowledgments

● Lawrence Berkeley Laboratory

- **Paul Adams**, Nat Echols, Jeff Headd, Ralf Grosse-Kunstleve, Nigel Moriarty, Nicholas Sauter, Peter Zwart

● Los Alamos National Laboratory

- Tom Terwilliger, Li-Wei Hung

● Cambridge University

- Randy Read, Airlie McCoy, Laurent Storoni, Gabor Bunkoczi, Robert Oeffner

● Duke University

- Jane Richardson & David Richardson, Ian Davis, Vincent Chen, Jeff Headd, Chris Williams, Bryan Arendall, Laura Murray

● Collaborations

- David Baker and Frank DiMaio
- Andy Karplus and Dale Tronrud
- Rhiju Das and Fang Chou
- Paul Langan and Marat Mustyakimov

● Others

- Garib Murshudov & Alexi Vagin
- Kevin Cowtan, Paul Emsley, Bernhard Lohkamp
- Alexandre Urzhumtsev & Vladimir Lunin
- David Abrahams
- PHENIX Testers & Users: James Fraser, Herb Klei, Warren Delano, William Scott, Joel Bard, Bob Nolte, Frank von Delft, Scott Classen, Ben Eisenbraun, Phil Evans, Felix Frolow, Christine Gee, Miguel Ortiz-Lombardia, Blaine Mooers, Daniil Prigozhin, Miles Pufall, Edward Snell, Eugene Valkov, Erik Vogan, Andre White, and many more

● Funding:

- NIH/NIGMS:
 - *P01GM063210, P50GM062412, P01GM064692, R01GM071939*
- Lawrence Berkeley Laboratory
- PHENIX Industrial Consortium

