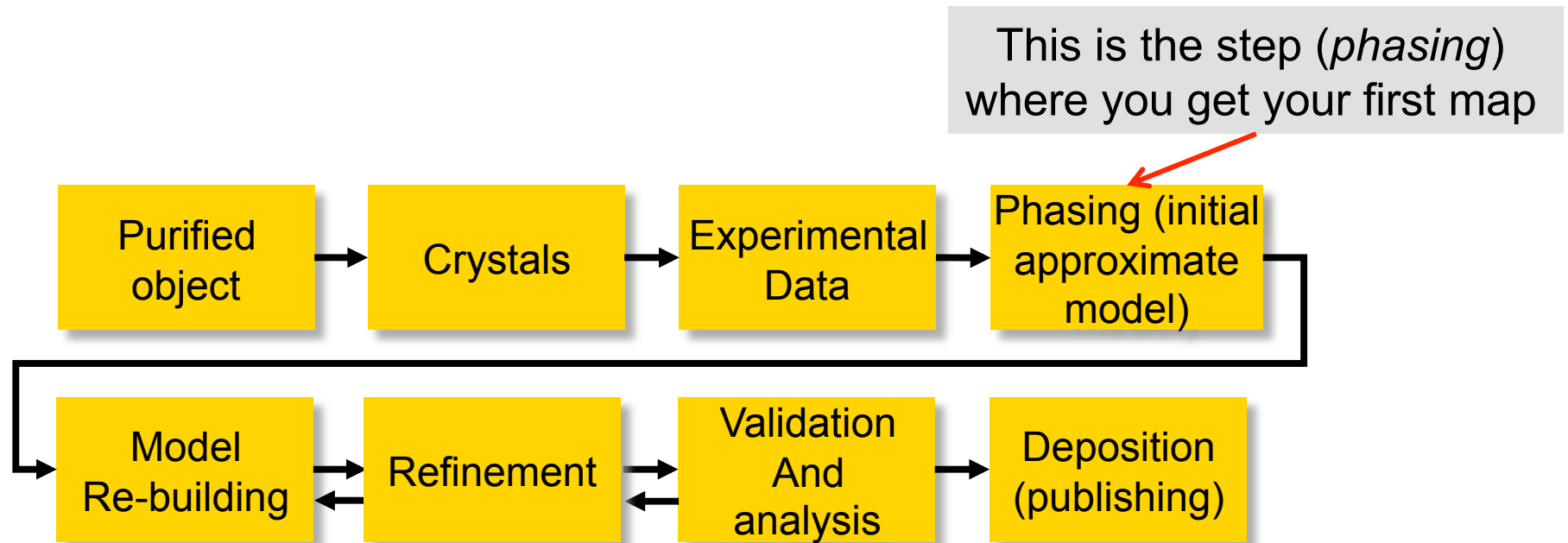# Facts About Maps

Pavel Afonine

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

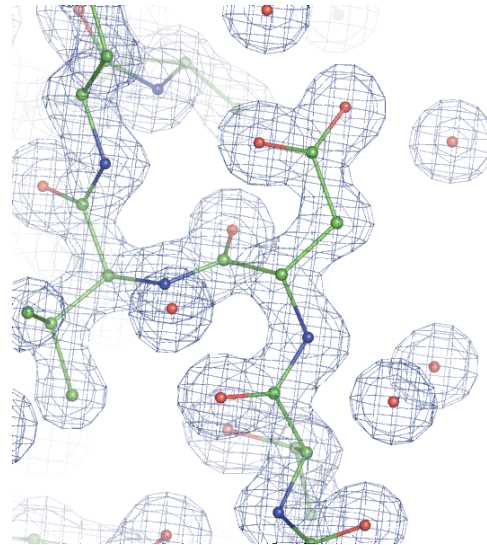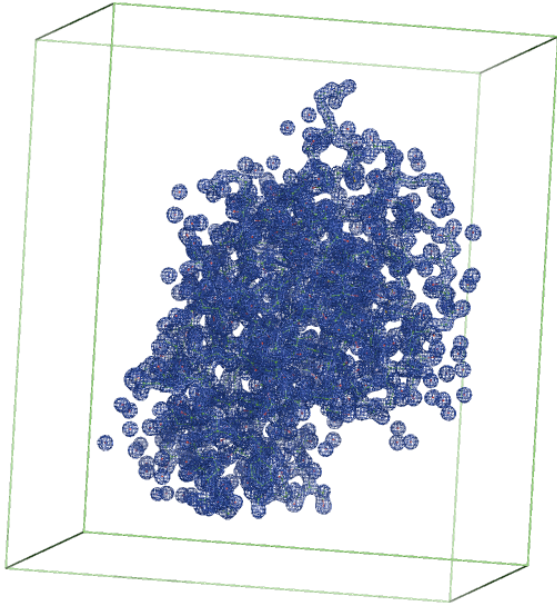Australasian Crystallography school
17th-24th July, 2010

- **Crystallographic structure determination workflow and your first map**

This is the step (*phasing*) where you get your first map

```
┌──────────┐   ┌──────────┐   ┌──────────────┐   ┌──────────────┐
│ Purified │ → │ Crystals │ → │ Experimental │ → │ Phasing      │
│ object   │   │          │   │ Data         │   │ (initial     │
│          │   │          │   │              │   │ approximate  │
│          │   │          │   │              │   │ model)       │
└──────────┘   └──────────┘   └──────────────┘   └──────────────┘

┌──────────┐   ┌──────────┐   ┌──────────────┐   ┌──────────────┐
│ Model    │ ⇄ │Refinement│ ⇄ │ Validation   │ → │ Deposition   │
│Re-building│   │          │   │ And          │   │ (publishing) │
│          │   │          │   │ analysis     │   │              │
└──────────┘   └──────────┘   └──────────────┘   └──────────────┘
```
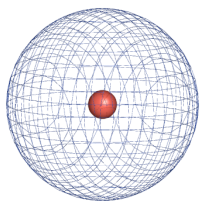
Electron density map

Density of individual atoms in high resolution map appears like a sphere or ellipsoid

- Computationally it is very beneficial to approximate the electron density arising from each atom as a Gaussian function
  - Electron density at the point **r** of an atom located at position **r₀** and having B-factor *B* and occupancy *q*:
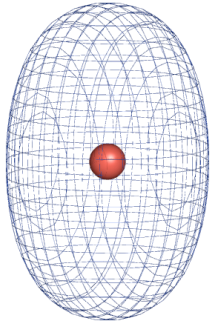

$$\rho_{atom}(\mathbf{r},\mathbf{r}_0,B,q) = q \sum_{k=1}^{5} a_k \left(\frac{4\pi}{b_k + B} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r} - \mathbf{r}_0|^2}{b_k + B} \right)$$

  - Number of terms in the above formula depends on how accurately we want to model an atom
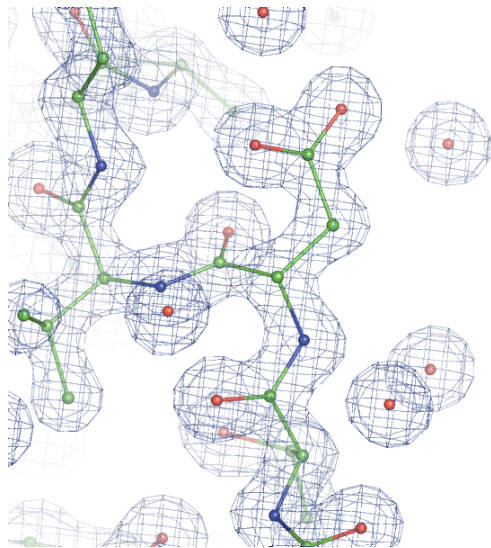
Atom density

- More accurate approximation assumes atoms moving anisotropically

$$\rho_{atom}(\mathbf{r}, \mathbf{U}, q) = q \sum_{j=1}^{5} \frac{q\, a_j\, (4\pi)^{3/2}}{\left| 8\pi^2 \mathbf{U}_{cart} + b_j \mathbf{I} \right|^{1/2}} \exp\!\left(-4\pi^2 (\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \left[8\pi^2 \mathbf{U}_{cart} + b_j \mathbf{I} \right]^{-1} \mathbf{A} (\mathbf{r} - \mathbf{r}_0) \right)$$
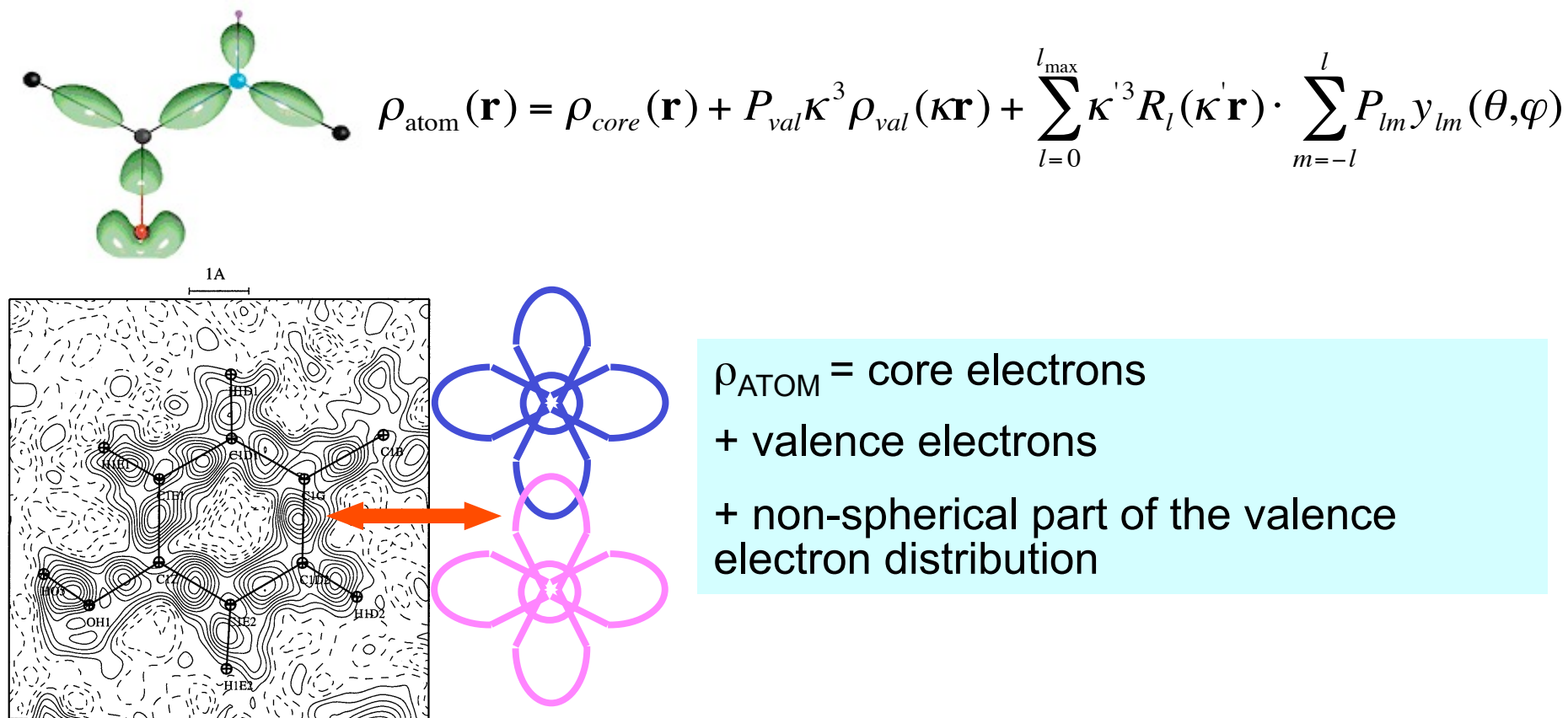
- When combining densities from individual atoms into whole molecule we assume atoms are individual and isolated (not bonded)

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

Mutipolar density model

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

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

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)
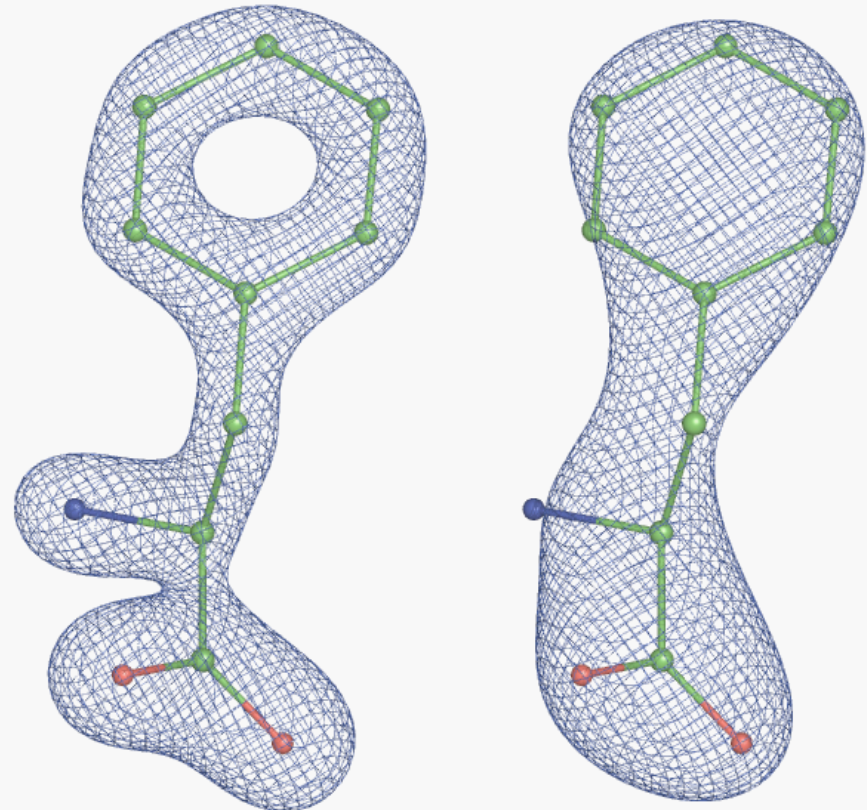
Electron density – structure factor relationship

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

Fourier transformation

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

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

- All atoms contribute to each structure factor

- Two extremely 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

$$\rho_{crystal}(\mathbf{r}) = \sum_{i=1}^{Natoms} q_i \sum_{k=1}^{5} a_k \left(\frac{4\pi}{b_k + B_i} \right)^{3/2} \exp\left(-\frac{4\pi^2 \left| \mathbf{r} - \mathbf{r}_{0,i} \right|^2}{b_k + B_i} \right)$$
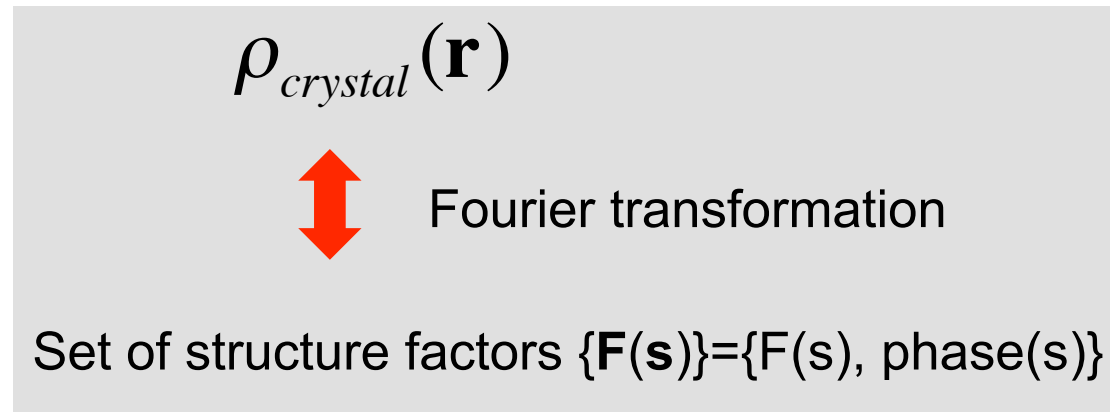
looks very nice:
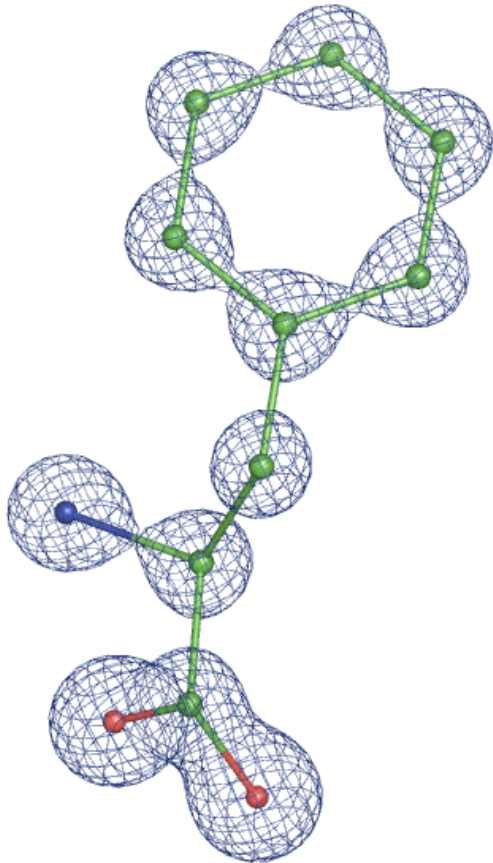
However in practice we see densities more like:

Electron density – structure factor relationship: Fourier image

- This is because…

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

↕ Fourier transformation

Set of structure factors {**F**(**s**)}={F(s), phase(s)}

…in diffraction experiment we do not measure all structure factors, but only a subset of {**F**(**s**)} is measured up to a certain resolution limit and it is typically not complete (some reflections may be missing within measured resolution range)

…there are some other sources of errors.

"Electron density" obtained as FT of limited set of {**F**(**s**)} is not actually $\rho_{crystal}(\mathbf{r})$ but its Fourier image, which may be significantly differ from $\rho_{crystal}(\mathbf{r})$

Electron density – structure factor relationship: model bias

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

$\updownarrow$ Fourier transformation

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

If *phase(s)* comes from the model – fundamental problem: model bias.

▪ **Model bias**

Model bias results from using an atomic model to calculate crystallographic phases.

The resulting electron density map will tend to have the 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).

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

- **Removing bias**

  – There are many methods 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 (it is strange to expect that omitting a piece of model would 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 residual map (mFo-DFc) and expect it unbiased

  - in fact, such a map may serve as a nice example of model bias!

Maps (1)

- **Direct maps**

  - $\{F_{OBS}, \text{phase}\}$

  - They are calculated from amplitudes and phases which have been inferred from the diffraction of the crystal or from a model.

  - Show unmodelled parts at half their density

  - Model biased
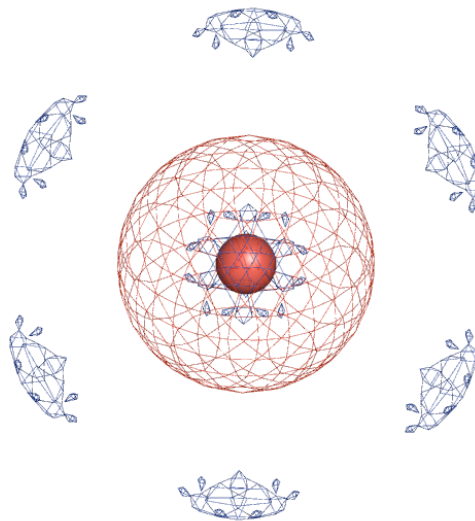
- **Balanced difference maps**

  - {F_{OBS}-F_{MODEL}, phase}
  - Difference maps are used to highlight errors in your model.
  - Positive density - model needs more electrons (additional atom, lower B-factor, or higher occupancy)
  - Negative density - model needs fewer electrons (removal of an atom, higher B-factor, or lower occupancy)
  - Positive and negative density close together - need to shift atom
  - Beware of noise !

$$\mathbf{F}_{MODEL} = k_{OVERALL} e^{-\mathbf{s}\mathbf{U}_{CRYSTAL}\,\mathbf{s}^t} \left(\mathbf{F}_{CALC\_ATOMS} + k_{SOL} e^{-\frac{B_{SOL}\,s^2}{4}} \mathbf{F}_{MASK} \right)$$

- **Unbalanced difference maps (Composite maps)**

  - Combination of a Direct Map and a Difference Map:
    $\{i*F_{OBS}-j*F_{MODEL}, phase\}$, where $i \neq j$

  - Commonly used

    $\{2F_{OBS}-F_{MODEL}, phase\}$, $\{3F_{OBS}-2F_{MODEL}, phase\}$,…

  - show modelled and unmodelled parts at full density

    $2F_{OBS}-F_{MODEL} = F_{OBS} + (F_{OBS}-F_{MODEL})$

- **Demonstration maps**

  – Used to prove something, $\{F_{OBS}-F_{OBS}, phase\}$

  – Bias-free

- **Manipulated maps:**

  – Kick maps

  – B-factor sharpened maps

  – Maps arising from density modification

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

2Fo-Fc
Fo: complete model, Fc and phase: Lys omit
3σ
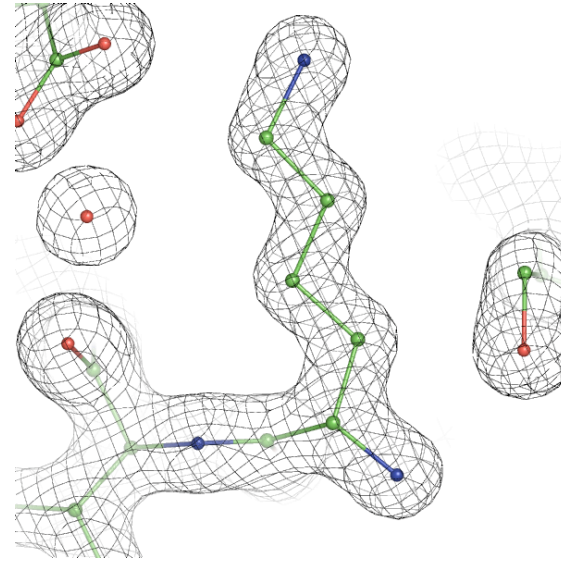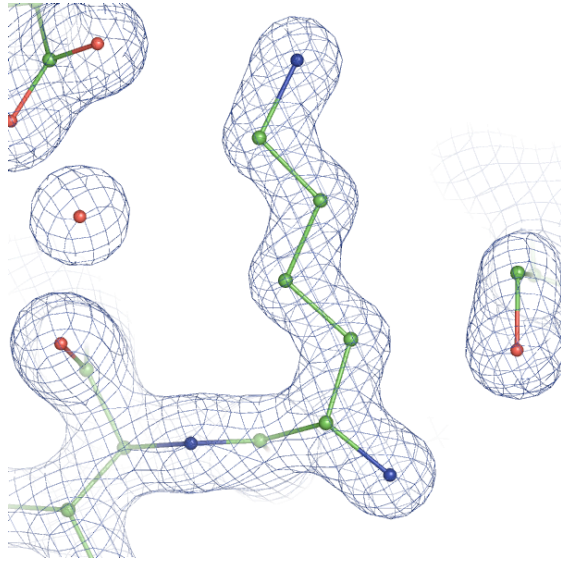
Fc-map, complete model
3σ

Fo, Fo: complete model, phase: Lys omit
3σ
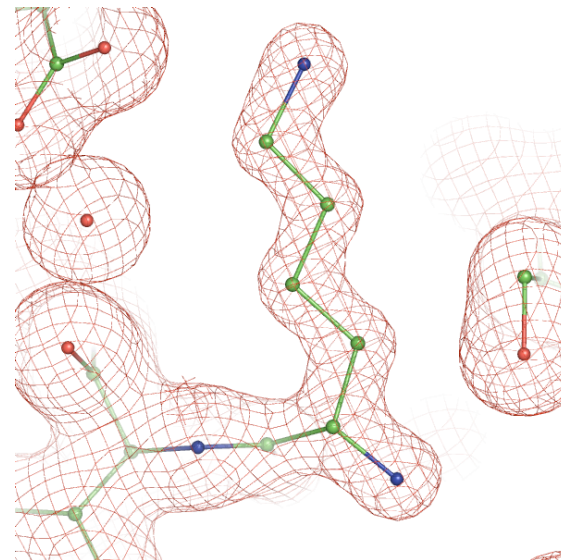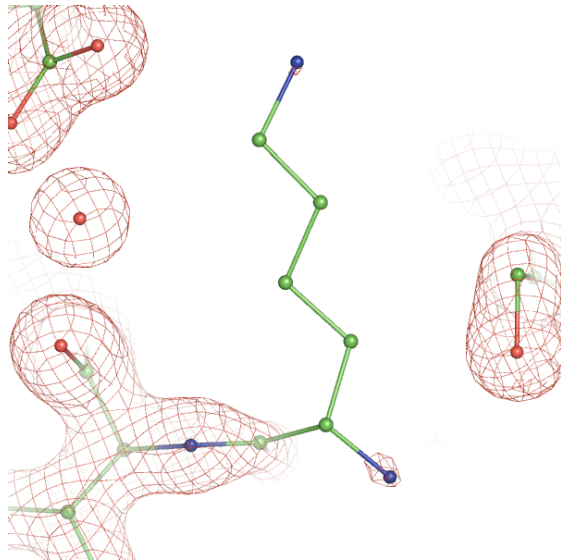
Fo, Fo: complete model, phase: Lys omit
1.5σ

- **Balanced difference maps**

  - $\{F_{OBS}-F_{MODEL}, \text{phase}\}$
  - Difference maps are used to highlight errors in your model.
  - Positive density - model needs more electrons (additional atom, lower B-factor, or higher occupancy)
  - Negative density - model needs fewer electrons (removal of an atom, higher B-factor, or lower occupancy)
  - Positive and negative density close together - need to shift atom
  - Beware of noise !

Balanced difference maps: model errors

Error in position
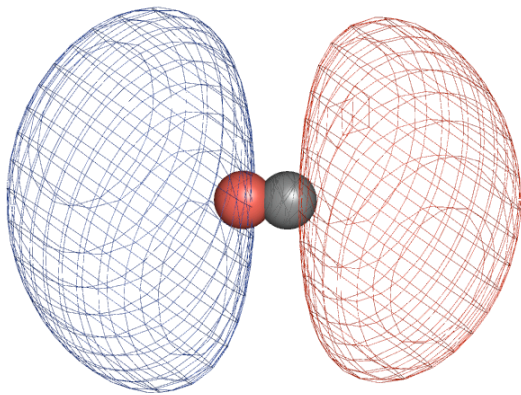
Error in occupancy

Error in B-factor

Model anisotropic atom with isotropic
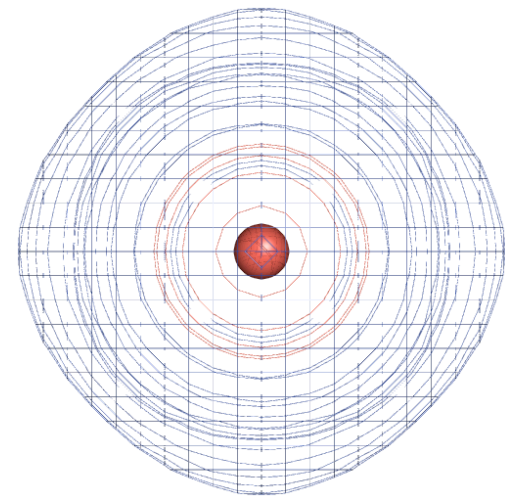
Balanced difference maps: model errors

Model anisotropic atom with isotropic
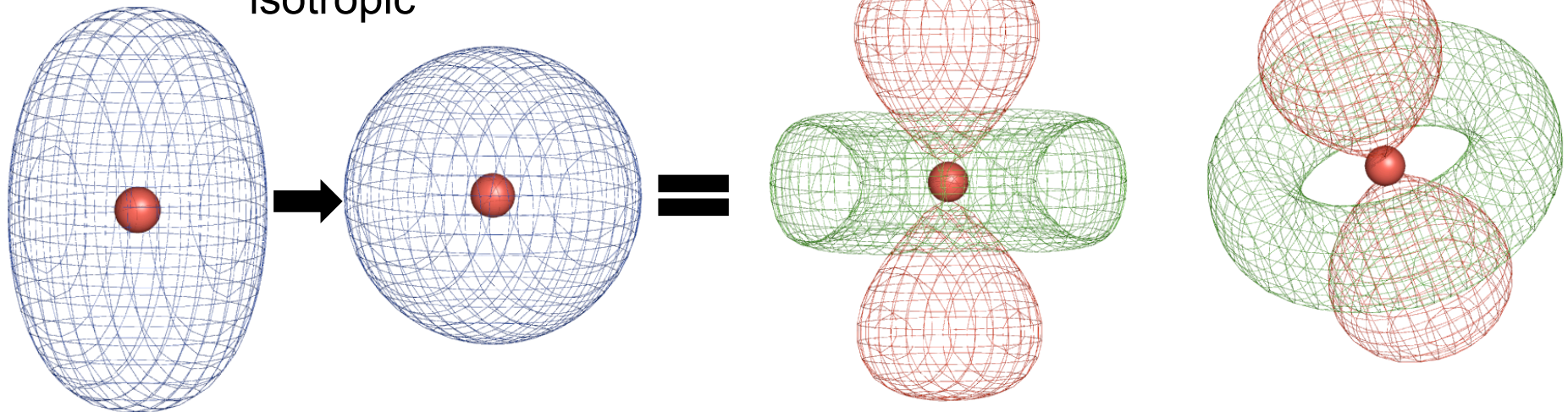
Add positional error
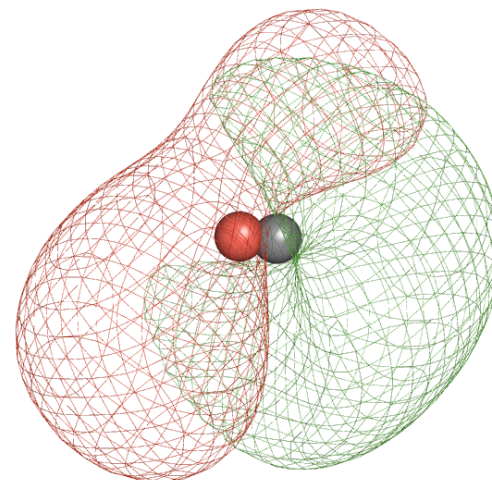
Balanced difference maps

- **Ser residue needs a different rotamer**
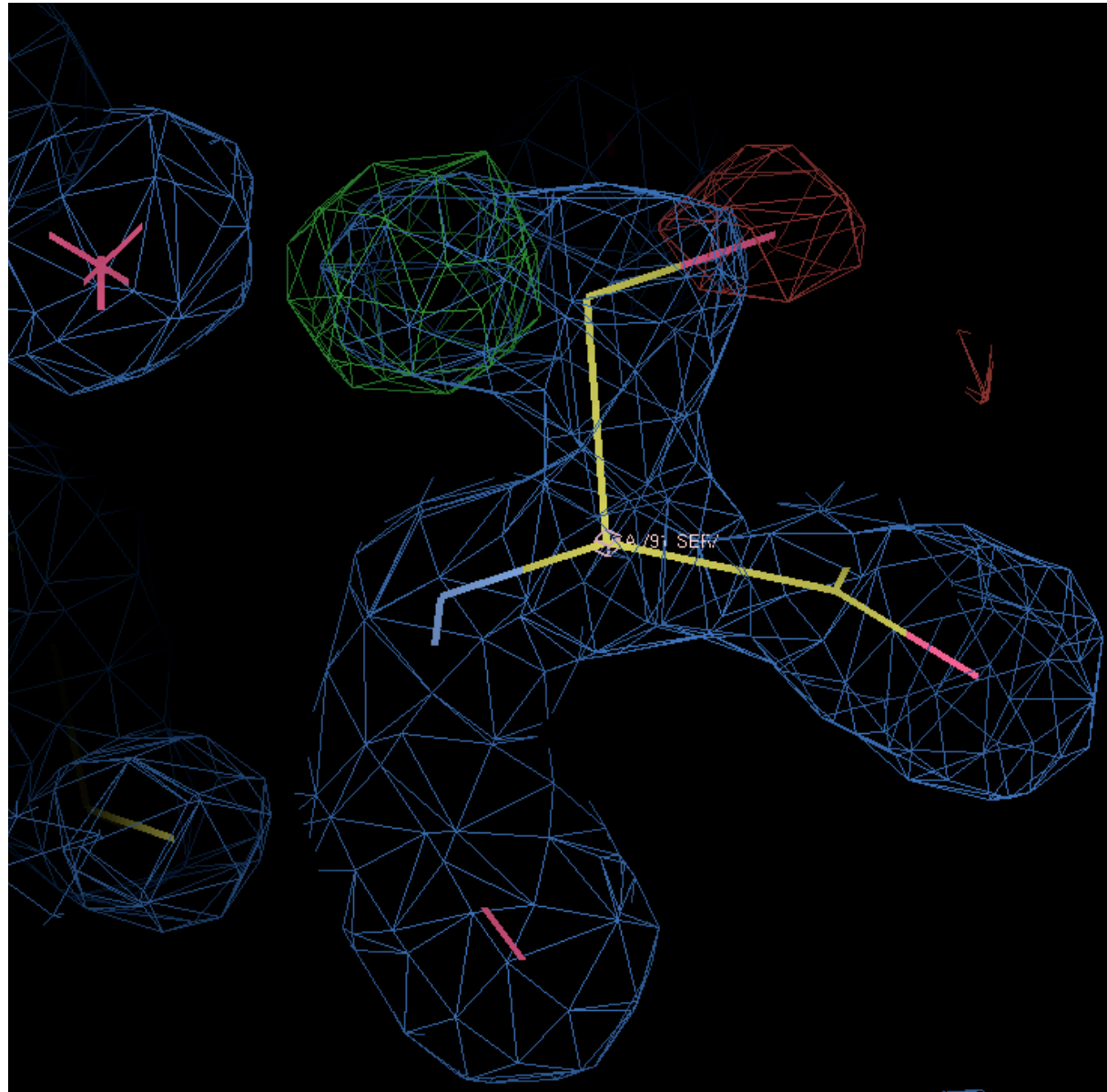
Maps (3)

- **Unbalanced difference maps:** least model biased map (σ_A-map)

    - Although $2F_{OBS}-F_{MODEL}$ is better than F_{OBS} map (show missing parts at full height), it is still model biased.

    - 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 Fo and Fc

        m and D are obtained by minimization of Maximum-Likelihood function w,r,t. these parameters *using test set of reflections* (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)

Other maps

- **Demonstration maps**

  – Used to prove something

  – Bias-free

- **Manipulated maps:**

  – Kick maps

  – B-factor sharpened maps

  – Phase-combined maps

Average Kick Maps

• Proposed by Dusan Turk more than 10 years ago and first implemented in his program MAIN

Original model

- → N models, shake rmsd 0.1 Å ⟶ map
- → N models, shake rmsd 0.2 Å ⟶ map ⟶ One averaged map
- …   …   (Average Kick Map)
- → N models, shake rmsd 0.7 Å ⟶ map
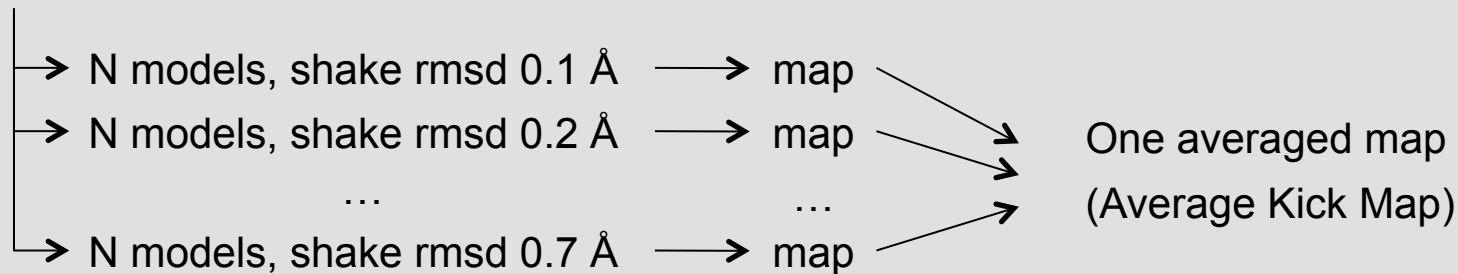
- **N** ~ 50 is enough
- **map** can be mFo-DFc, 2mFo-DFc, omit maps…
- **Average Kick Map** is expected to be
  – Less noisy
  – Less model biased

- Available in *PHENIX*
- J. Praznikar, P. V. Afonine, G. Guncar, P. D. Adams and D. Turk. *Acta Cryst.* (2009). D**65**, 921-931.

People say it's useful…

Subject: Minor question
From: "Roy, Siddhartha" <XXX@XXX.edu>
Date: Thu, 27 Aug 2009 14:21:53 -0600
To: Pavel Afonine <PAfonine@lbl.gov>

Hi Pavel,
*I calculated both 2fofc and fofc maps and also kicked maps for my 1.6A data. **It looks like kicked map is much better than corresponding 2fofc and fofc maps**. I am able to see **more water molecules and clear side chains (looks like less model bias)**. (…)*
Thanks,
Sid

Phase-combined maps

Map amplitudes

$F = F_{OBS} - F_{MODEL}$

or

F_{OBS}

or

$2mF_{OBS} - F_{MODEL}$ or …

Phases:

Model

+

Experimental

Phase-combined map

Phases from model are combined with experimental phases using HL coefficients (and some scary formulas, phase integrals)

Anomalous Difference Map

Shows positions of anomalous scatterers

$$\frac{|F_{OBS}(\mathbf{h}) - F_{OBS}(-\mathbf{h})|}{2i} \exp(i\varphi) \quad = \quad |F_{OBS}(\mathbf{h}) - F_{OBS}(-\mathbf{h})| \exp(i\varphi - \frac{\pi}{2})$$
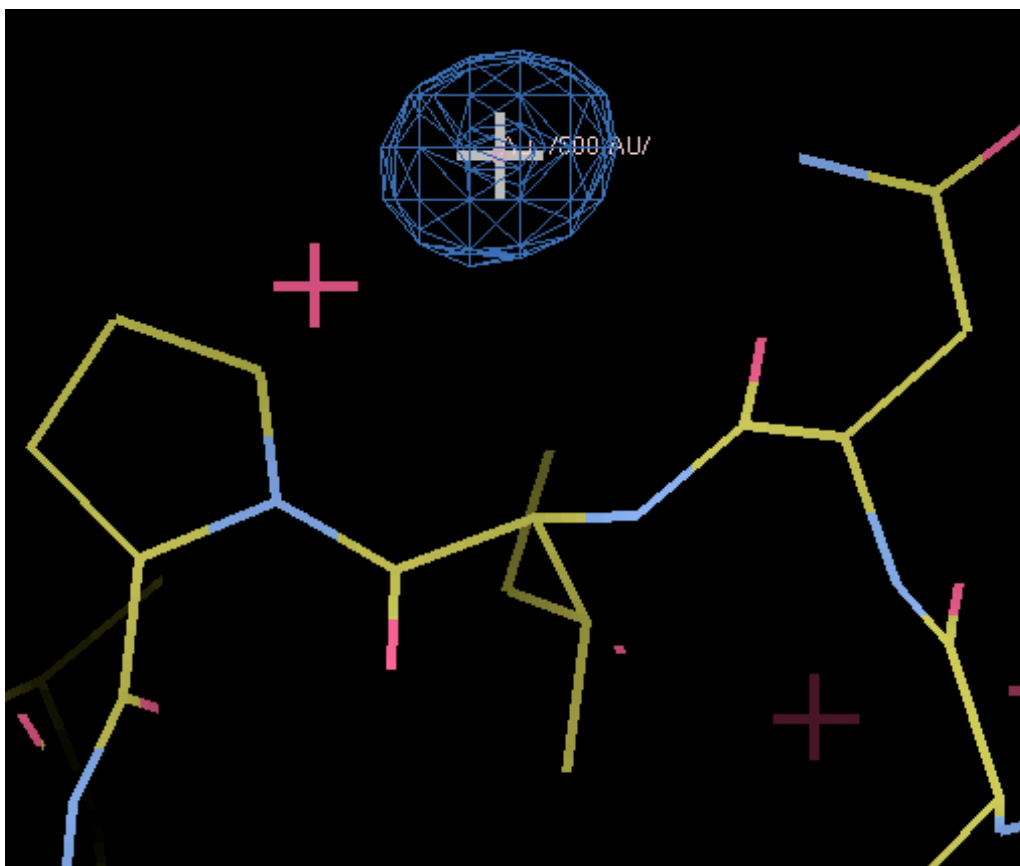
B-factor sharpened maps

- The whole map (Fourier map coefficients) is multiplied by ~exp($-Bs^2$)

- Very efficient at improving low resolution map interpretability
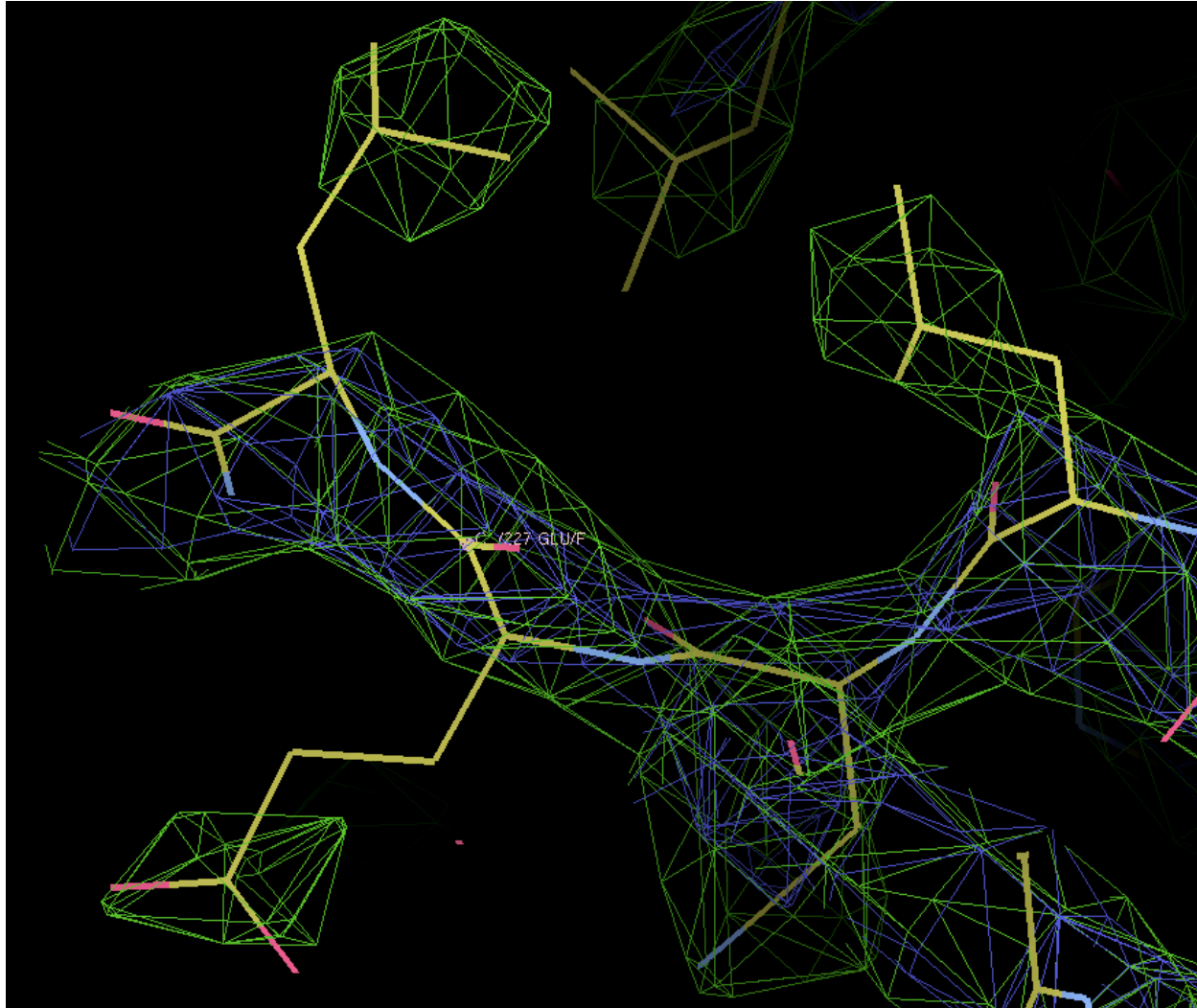
Fourier truncation ripples
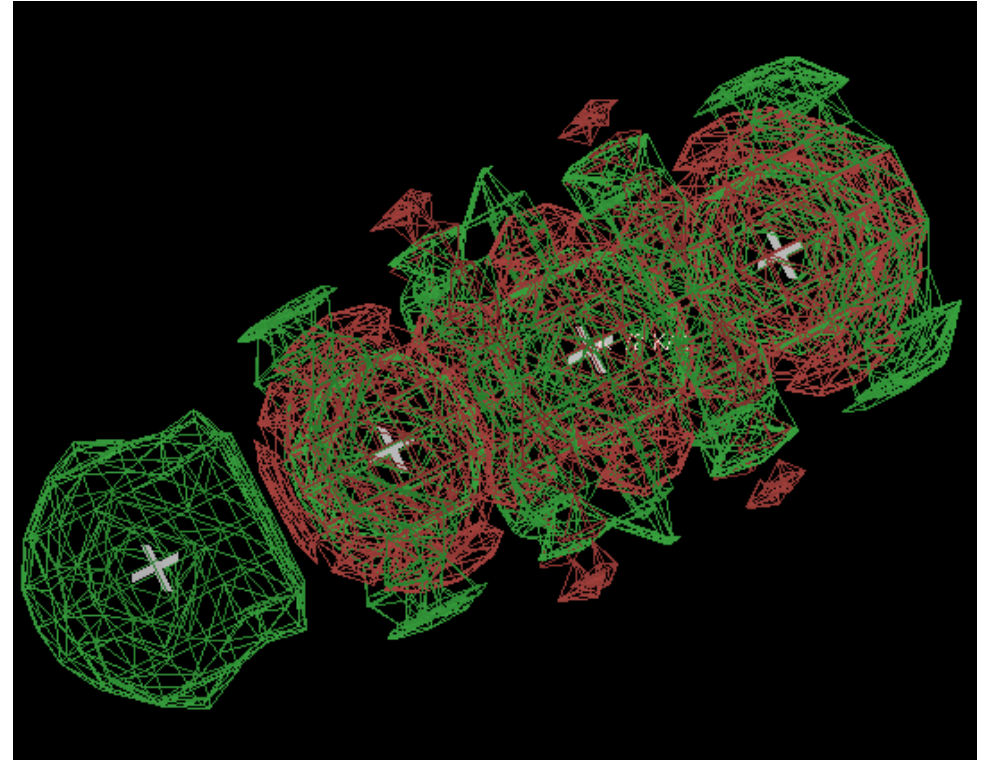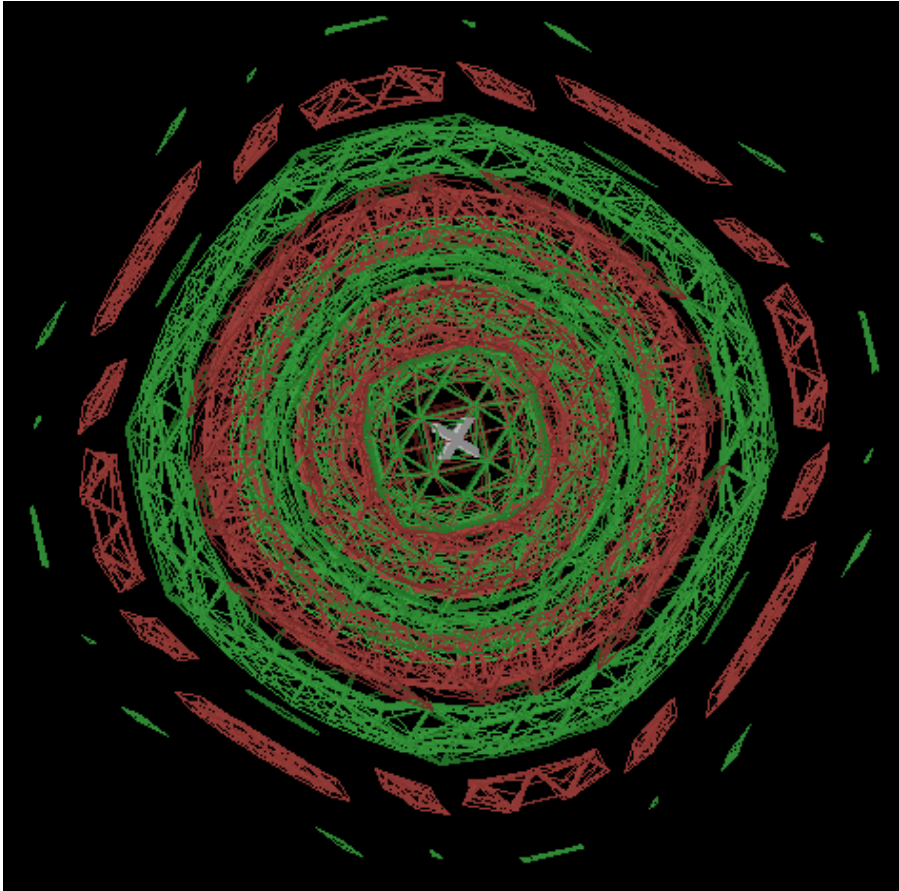
Refinement at subatomic resolution

Aldose Reductase (0.66 Å resolution)
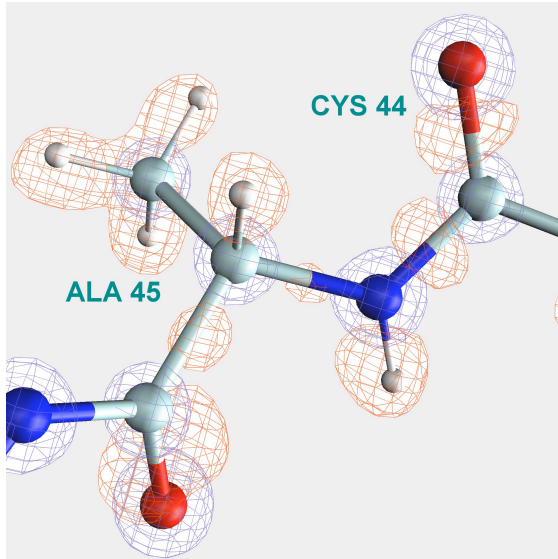
1UCS (0.62Å), Fo-Fc, 1.7σ

Fo-Fc (orange)   2Fo-Fc (blue)

1P9G (0.8Å), Fo-Fc, 1.45σ
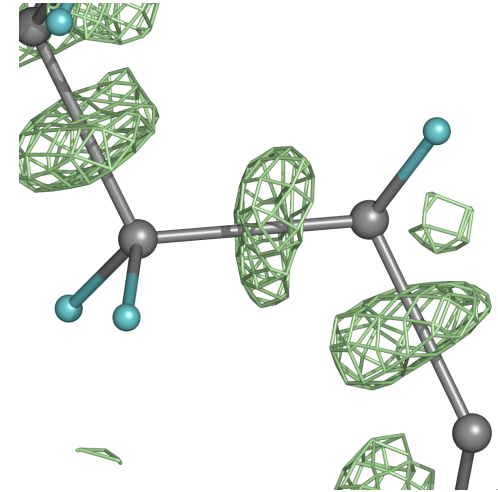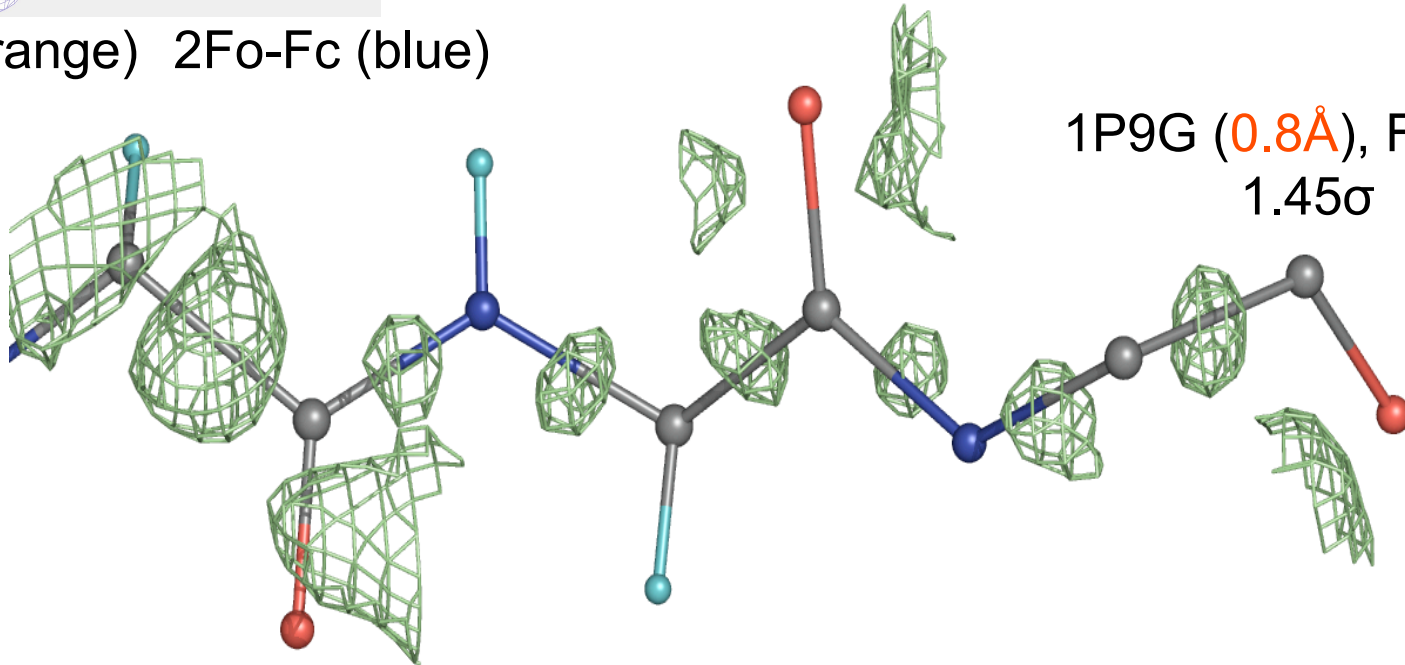
Maps at subatomic resolutions: dangers

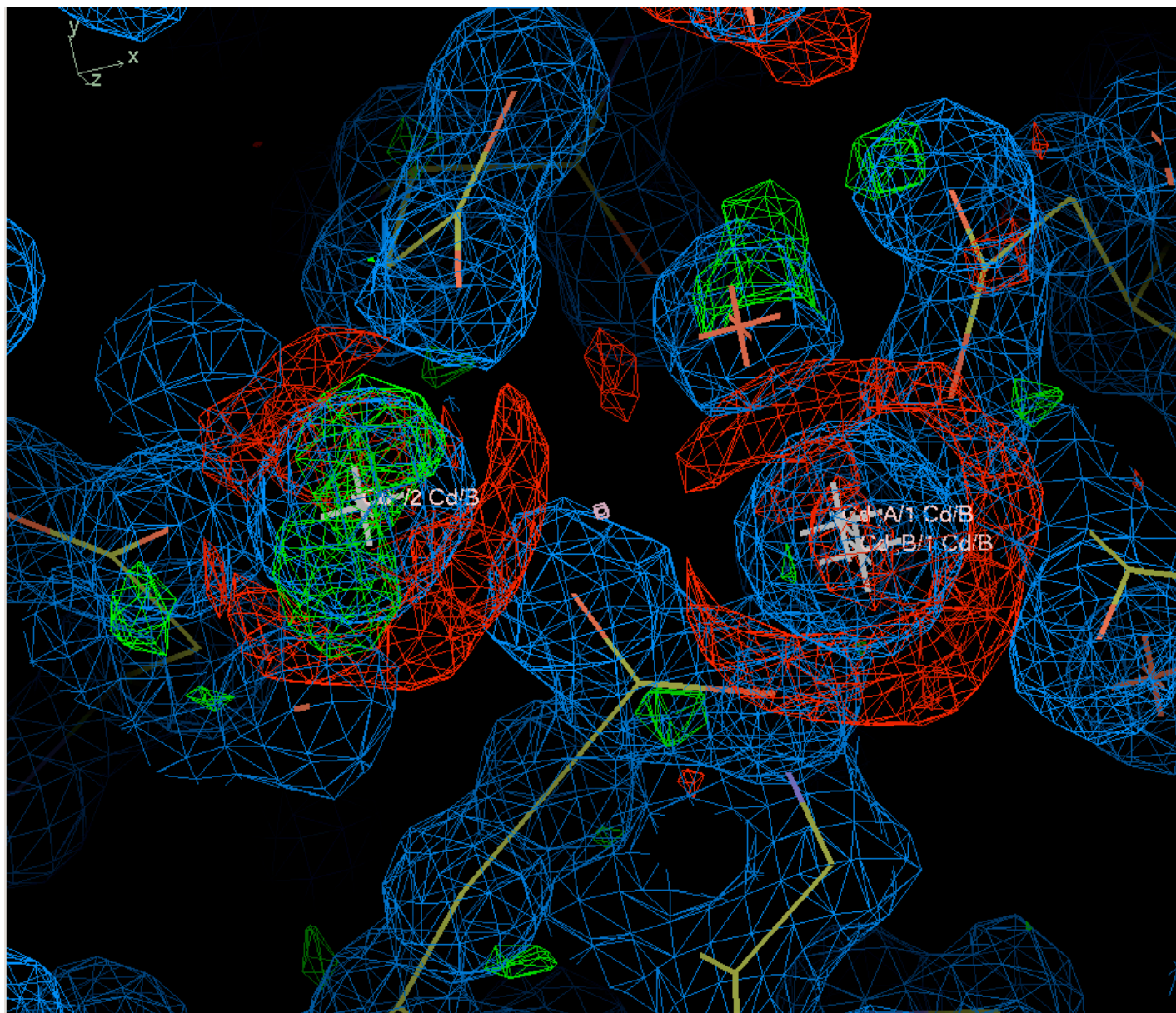▪ Copy from "Experimental Observation of Bonding Electrons in Proteins", JBC, 1999, Vol. 274.

▪ $(F_{CALC}, \varphi_{CALC})$ synthesis at 0.6 Å :
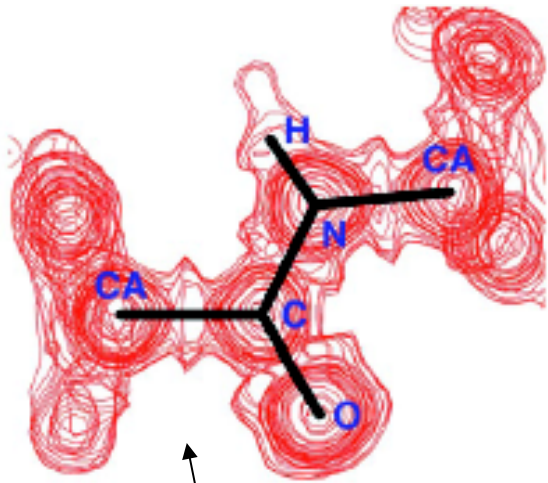
This is not bond electrons…
These are Fourier series truncation ripples !
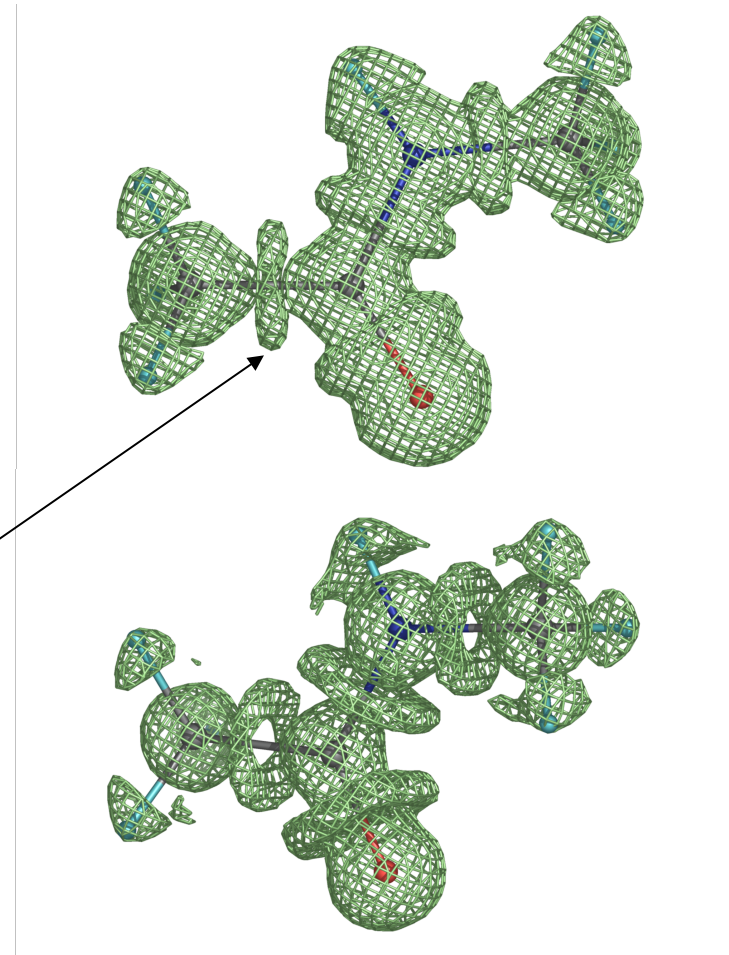
▪ *Unbalanced Fourier syntheses should be interpreted with care.*
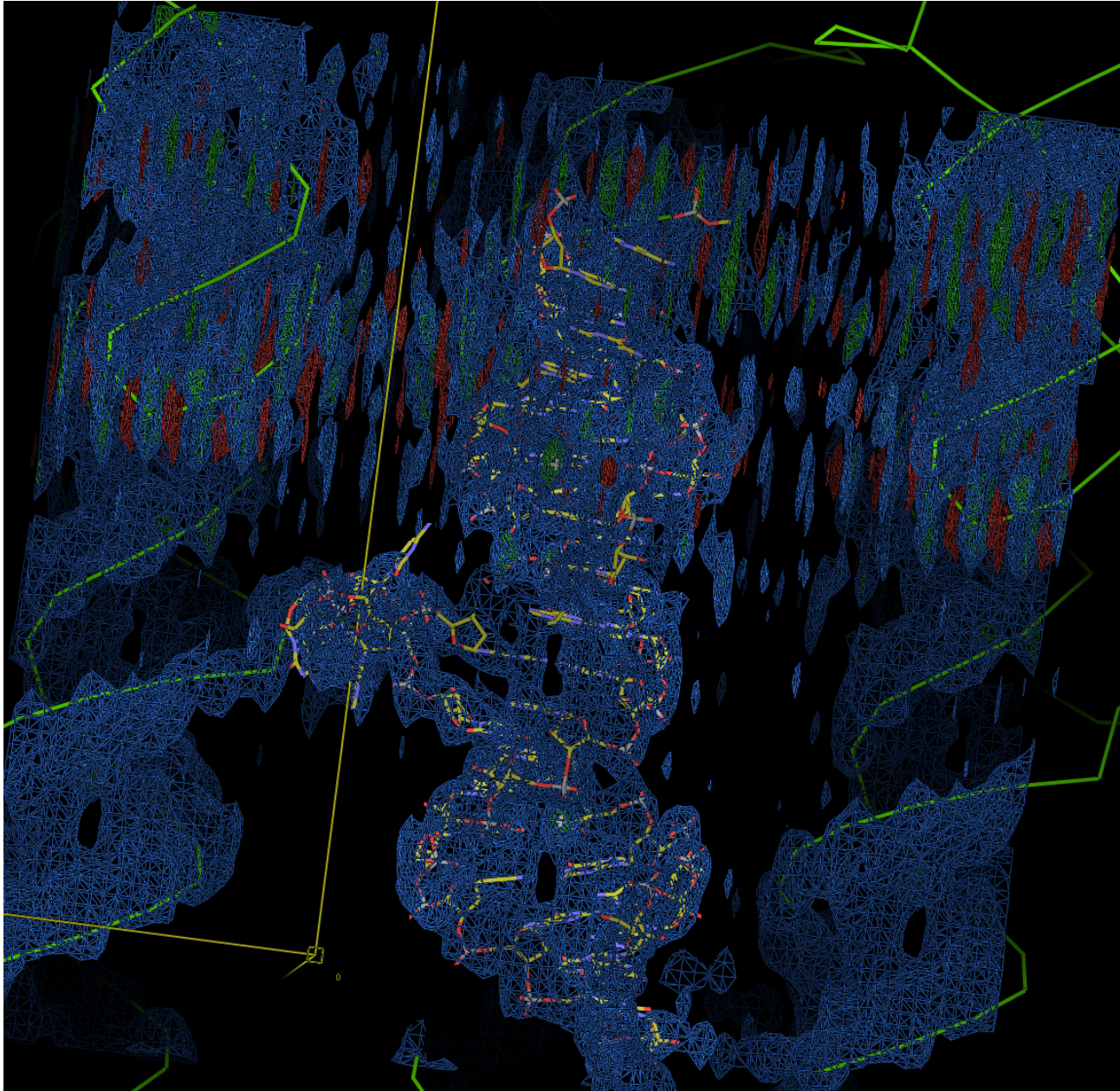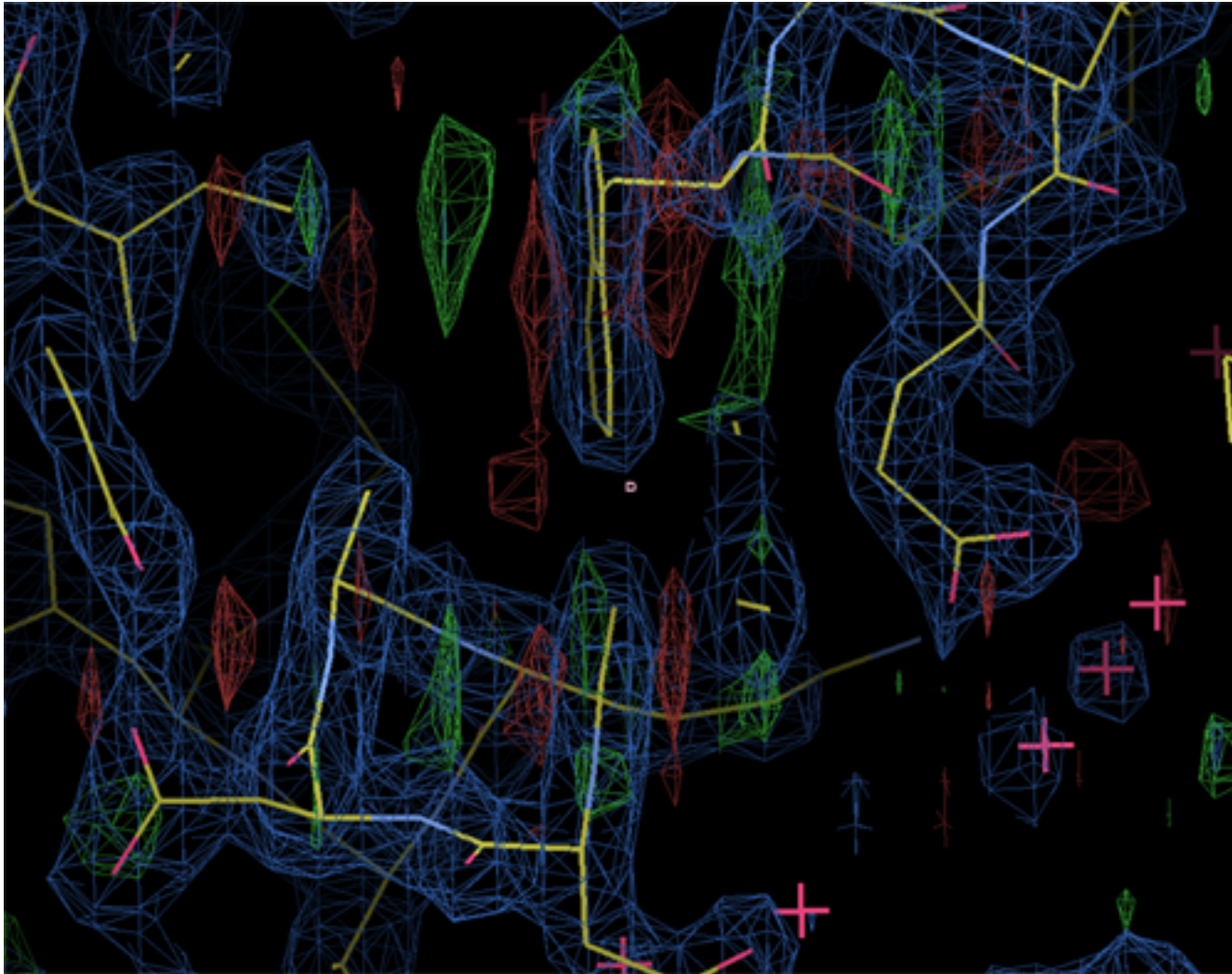
Map output formats

- X-plor/CNS formatted map: a text file with actual map (3D function computed on the grid). This map is used to see in PyMol. Typically huge file, longer to load.

- File with Fourier map coefficients. Typically in MTZ format. Small files, quick to load, may contain several "maps". Graphical programs, like Coot, will do Fourier transform internally in order to produce the map.

- Latest version of PHENIX can output maps in CCP4 format. This is an actual map in binary format. Smaller file. Can be used with Coot or PyMol.

Map units

- Electron density map

$$\{\mathbf{F(s)}\} \xrightarrow{\text{FT}} \rho_{crystal}(\mathbf{r})$$

has some arbitrary units.

- Two ways of bringing a map into some scale:

  – Divide it by standard deviation (map in sigmas)

  – Include reflection F(000) and divide map by the unit cell volume. Model should be complete to estimate F(000). Map in e/Å^3.

What refinement programs write out (typically)

- phenix.refine always outputs three maps:

  2mFo-DFc, 2mFo-DFc (Fobs filled), mFo-DFc

- For anomalous dataset (dataset containing Fobs+ and Fobs-) phenix.refine will output an anomalous difference map

- All the above maps are output in MTZ file as Fourier map coefficient
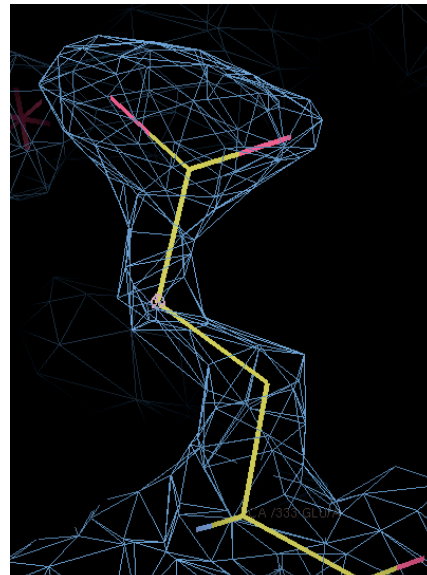
Missing Fobs filled maps

- **Missing Fobs data:**

  Data incompleteness can seriously impact maps and systematically missing reflections may even make invisible whole sections of a molecule (Lunin, 1988; Urzhumtsev *et al*., 1989; Lunin & Skovoroda, 1991; Lunina *et al*., 2002).
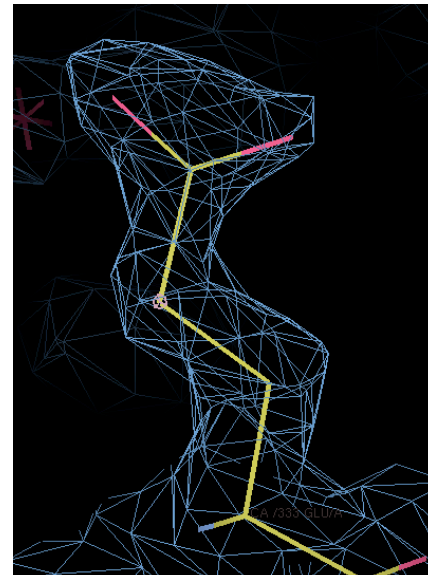
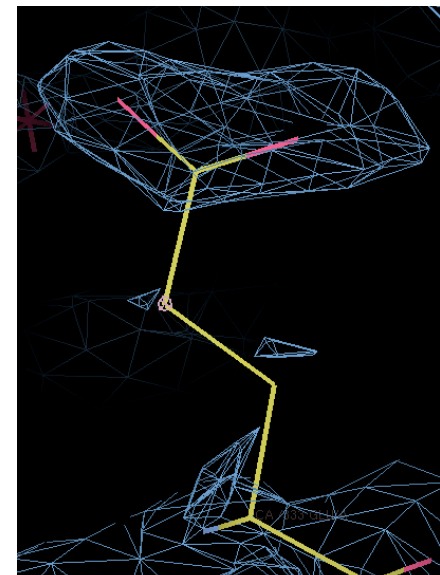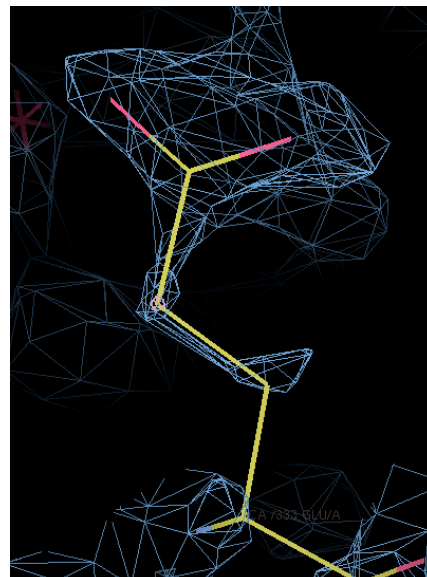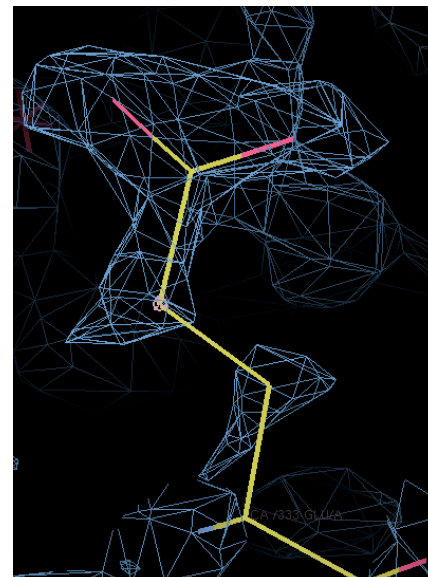# Effect of systematic data incompleteness
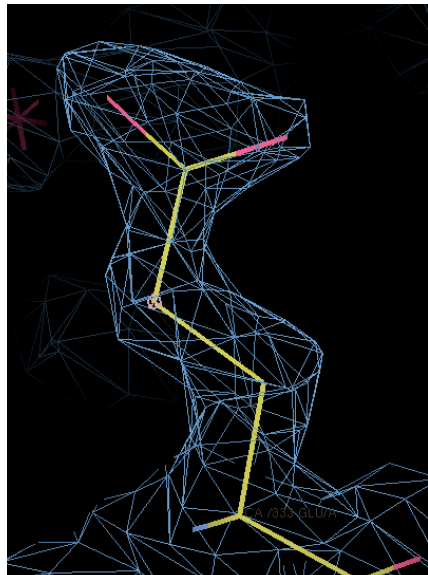
original

random

weak

cone

low

strong_and_weak

strong

Missing Fobs filled maps

• Restoring missing Fobs with DFc (for example, other options exist) in map calculation may eliminate the problem.

• Pros & Cons:
  – maps gets improved;
  – risk of model bias.

• phenix.refine always outputs three maps:
      2mFo-DFc, 2mFo-DFc (Fobs filled), mFo-DFc

• As far as I know, REFMAC outputs 2mFo-DFc (Fobs filled) map by default.