**The PHENIX project**

Crystallographic software for automated structure determination

**Structured Determination by MAD/SAD/MIR in PHENIX**

PROVIDED THAT:
- Data files (H K L Fobs Sigma)
- Crystal information (Space group, Cell)
- Scattering factors (for MAD)

Data are strong, accurate, < 3 Å
- Strong anomalous signal
- Little decay
- Space group is correct
- Scattering factors close (for MAD data)

You are willing to wait a little while...
(10 minutes to hours, depending on size)

Heavy-atom coordinates
- Non-crystallographic symmetry
- Electron-density map
- Atomic model

Model is 50-95% complete
(leading on resolution)
- Model is (mostly) compatible with the data...but is not completely correct
- Model requires manual rebuilding
- Model requires validation and error analysis

**Major needs in automated structure solution**

MAD/SAD/MIR
- Robust structure determination procedures
- Best possible electron density maps to build most complete model
- Decision-making about best path for structure solution

Molecular Replacement
- Use of distant models
- Preventing model bias

All structures
- Model completion/Ligand fitting
- Error analysis
- Decision-making for what data to use and what path to follow
- How to incorporate vast experience of crystallographic community

**Best possible electron density maps to build the most complete model**

Statistical density modification
- Local patterns of density
- ID of fragments

Iterative model-building and refinement

FULL-OMIT density modification and model-building

**Why we need good measures of the quality of an electron-density map:**

Which solution is best?
Are we on the right track?
If map is good:
- It is easy

**Statistical density modification**

(A framework that separates map information from experimental information and builds on density modification procedures developed by Wang, Bricogne and others)

- Principle: phase probability information from probability of the map and from experiment:
  \[
  P(\phi) = P_{\text{map}}(\phi) / P_{\text{experiment}}(\phi)
  \]
- “Phases that lead to a believable map are more probable than those that do not”
- A believable map is a map that has...
  - a relatively flat solvent region
  - NCS (if appropriate)
  - A distribution of densities like those of model proteins

- Calculating map probability (\phi):
  - calculate how map probability varies with electron density \( \rho \)
  - Use chain rule to deduce how map probability varies with phase (equations of Bricogne, 1992).
Map probability phasing:
Getting a new probability distribution for a single phase given estimates of all others

1. Identify expected features of map (flat far from center)
2. Calculate map with current estimates of all structure factors except one (k)
3. Test all possible phases $\phi$ for structure factor k (for each phase, calculate new map including k)
4. Probability of phase $\phi$ estimated from agreement of maps with expectations

A map-probability function

$$L_{n}^{\text{SP}}(\{\phi_{k}\}) = \frac{N_{\text{SP}}}{N} \int_{-h}^{h} g_{\phi}(x) d\phi$$

A function that is relatively flat far from the origin

Function calculated from estimates of all structure factors but one (k)

Test each possible phase of structure factor k. $\phi_{k}$ is high for phase that leads to flat region

Local log-probability is believability of the value of electron density ($\rho(x)$) found at this point

Solve map (52 Se)

Resolve map

Composite omit map with statistical density modification

Statistical density modification allows a separate probability distribution for electron density at each point in the map: can specify that “missing” density is within molecular boundary

Electron density maps of proteins have many features in common

- Connected density
- Preferred distances for spacing between regions of high density
- Preferred shapes of density

Image enhancement using local feature recognition

Approach:

- Use the pattern of density near a point $x$ to estimate the value of density at $x$
- Combine new estimate of density with previous one to improve the overall image
  *“Local NCS averaging”*
Image enhancement using local feature recognition

Approach:

• Create N templates of local density using model data
  - Examine density near each point x in image (within 2 Å)
  - Exclude region very close to x (about 1 Å)
  - Cluster and average local patterns of density (after rotation to maximize CC)
• Identify relationship between finding pattern k of density near x, and density at x
  - Find all locations in the image where template k best matches the local density near x
  - Calculate average value of density at x for these cases = ρmean(k)
• Identify pattern near each point in actual map and use it to estimate density at that point
  - For each point x in the image, identify which template k best matches the local density near x
  - Use ρmean(k) as estimate of density at x

Image enhancement using local feature recognition

Random map at 2.6 Å
Recovered image derived from random map
CC to original random map = 0.01

Iterative procedure for image enhancement using local feature recognition

- Iterative steps:
  - pcurrent(φ)
  - Combined phases
  - pmapped(φ)
  - Density-modified phases
  - ρdm(x)
  - Density-modified map
  - ρimage(x)
  - Recovered image
  - pimage(φ)
  - Image-based phases

Image recovery from a good map...
- gives an image that has (mostly) correct features
- errors are (almost) uncorrelated with original errors

Image recovery from a random map...
- gives an uncorrelated image
Image enhancement using local feature recognition (nusA protein structure)

Starting map
CC=0.65

Cycle 1
CC=0.75

Cycle 3
CC=0.84

Cycle 5
CC=0.85

Removing model bias with prime-and-switch phasing

The problem:
Atomic model used to calculate phases -> map looks like the model

Best current solution: $\sigma_A$-weighted phases

Prime-and-switch phasing

A solution:
Start with $\sigma_A$-weighted map
Identify solvent region (or other features of map)
Adjust the phases to maximize the probability of the map = without biasing towards the model phases

Prime-and-switch phasing

Why it should work...

Signal: peak height at correct atomic positions
Bias: peak height at incorrect atoms in starting model

Prime-and-switch example:
(IF5A, T. Peat)

Blue: model used to calculate phases
Orange: correct model

Prime-and-switch example:
(Gene V protein, Matt Skinner)

Bottom:
Prime-and-switch phases starting from incorrect model

Top:
SigmaA phases from incorrect model
**PHENIX AutoBuild wizard standard sequence**  
(Following ideas from Lamzin & Perrakis)

- Fp, phases, HL coefficients
- Density modify (with NCS, density histograms, solvent flattening, fragment ID, local pattern ID)
- Build and score models  
  Refine with phenix.refine
- Density modify including model information
- Evaluate final model

**SAD data at 2.6 A**  
gene 5 protein

- Solve SAD map

**Cycle 50 of iterative model-building, density modification and refinement**

- Density-modified SAD map

**Why iterative model building, density modification, and refinement can improve a map** (following ideas of Perrakis & Lamzin):

1. New information is introduced: flat solvent, density distributions, stereochemically reasonable geometry and atomic shapes
2. Model rebuilding removes correlations of errors in atomic positions introduced by refinement
3. Improvement of density in one part of map improves density everywhere.
Iterative model-building and refinement is very powerful but isn't perfect...

Model-based information is introduced in exactly the same place that we will want to look for details of electron density

→ How can we be sure that the density is not biased due to our model information?
   (Will density be higher just because we put an atom there?)
   (Will solvent region be flatter than it really is because we flattened it?)
   (Will we underestimate errors in electron density from a density-modified map?)
   (Are we losing some types of information by requiring the map to match partially incorrect prior knowledge?)

A FULL-OMIT iterative-model-building map: everywhere improved, everywhere unbiased

→ Use prior knowledge about one part of a map to improve density in another

Related methods: "Omit map", "SA-composite omit map", density-modification OMIT methods, "Ping-pong refinement"

Principal new feature:
The benefits of iterative model-building are obtained yet the entire map is unbiased

Requires:
Statistical density modification so that separate probability distributions can be specified for omit regions (allow anything) and modified regions (apply prior knowledge)

Including all regions in density modification comparison with FULL-OMIT

Density-modified  ----------------- Iterative model-building -----------------

Molecular Replacement:
(Mtb superoxide dismutase, 1IDS, Cooper et al, 1994)

Uses:
Unbiased high-quality electron density from experimental phases
High-quality molecular replacement maps with no model bias
Model evaluation

Computation required:
~24 x the computation for standard iterative model-building
FULL-OMIT iterative-model-building maps

Requirement for preventing bias:
Density information must have no long-range correlated errors
(the position of one atom must not have been adjusted to compensate for errors in another)

Starting model (if MR) must be unrefined in this cell

Major needs in automated structure solution

MAD/SAD/MIR
Robust structure determination procedures
Best possible electron density maps to build most complete model

Molecular Replacement
Use of distant models
Preventing model bias

All structures
Model completion/Ligand fitting
Error analysis
Decision-making for what data to use and what path to follow
How to incorporate vast experience of crystallographic community

Acknowledgements

PHENIX: www.phenix-online.org
Computational Crystallography Initiative (LBNL):
Paul Adams, Ralf Grosse-Kunstleve, Nigel Moriarty, Nick Sauter, Pavel Afonine, Peter Zwart
Randy Read, Airlie McCoy, Laurent Storoni, Hamsapriye (Cambridge)
Tom Ioerger, Jim Sacchettini, Kresna Gopal, Lalji Kanbi, Erik McKee Tod Romo, Reetal Pai, Kevin Childs, Vinod Reddy (Texas A&M)
Li-wei Hung, Thiru Radhakannan (Los Alamos)

PHENIX web site:
http://phenixonline.org

SOLVE/RESOLVE web site:
http://solve.LANL.gov

SOLVE/RESOLVE user's group:
solve@LANL.gov

Generous support for PHENIX from the NIGMS Protein Structure Initiative