Tools for Easy and Difficult Problems: Automation of Structure Determination by Macromolecular Crystallography

Indo-US workshop on Macromolecular Structure Determination
Feb. 21-24, 2011

Tom Terwilliger
Los Alamos National Laboratory
A brief introduction to X-ray crystallography

Growing protein crystals

Looking at crystals with X-rays

Getting pictures of proteins from diffraction spots
Growing protein crystals

Protein + alcohol = Crystals

Salts, alcohol, polyethylene glycol...

Los Alamos
Bioscience Division

Innovation for Health and Security
A brief introduction to X-ray crystallography

- Growing protein crystals
- Looking at crystals with X-rays
- Getting pictures of proteins from diffraction spots
A brief introduction to X-ray crystallography

Growing protein crystals

Looking at crystals with X-rays

Getting pictures of proteins from diffraction spots
A brief introduction to X-ray crystallography

Growing protein crystals

Looking at crystals with X-rays

Getting pictures of proteins from diffraction spots
A brief introduction to X-ray crystallography

Growing protein crystals

Looking at crystals with X-rays

Getting pictures of proteins from diffraction spots

Los Alamos
Bioscience Division

Innovation for Health and Security
Mounting crystals in nylon loops and cryo-cooling them
A brief introduction to X-ray crystallography

Growing protein crystals

Looking at crystals with X-rays

Getting pictures of proteins from diffraction spots
A brief introduction to X-ray crystallography

Growing protein crystals

Looking at crystals with X-rays

Getting pictures of proteins from diffraction spots
Getting pictures of proteins from diffraction spots

- Diffraction pattern
- Analysis of diffraction spots
- Picture and model of macromolecule
The intensities of X-ray diffraction spots depend on what is in the crystal.

Electrons in a protein crystal (\( \rho \) is high where there are many electrons)

\[
\rho(x) = \sum_h F_h e^{i\phi_h} e^{-2\pi i h x}
\]

Diffraction pattern (\( I_h \) is intensity of spot “h”)

\[
F_h e^{i\phi_h} = \int \rho(x) e^{2\pi i h x} dx \quad I_h = F_h^2
\]
We can almost calculate a picture of where the atoms are from the diffraction pattern (but are missing the phases of diffraction spots)

\[ \rho(x) \]
(Where the atoms are)

\[ \rho(x) = \sum_h F_h e^{i\phi_h} e^{-2\pi i h x} \]

\( F_h \) is square root of measured intensity of spot \( h \)

We do not know the phase \( \phi_h \).
Estimating crystallographic phases: example with multiwavelength X-ray data

- Measure diffraction \( (I_h, I'_h) \) at two X-ray wavelengths near absorption edge of selenium

- Differences in diffraction are due to changes in scattering from the Se atoms \( (\Delta F_h) \)

First figure out where the Se atoms are located

Then use the Se atoms and the diffraction intensities to draw a picture of all the atoms
**Estimating crystallographic phases with multiwavelength X-ray data**

- Measure diffraction ($I_h$, $I'_h$) at two X-ray wavelengths near absorption edge of selenium.

- Differences in diffraction are due to changes in scattering from the Se atoms ($\Delta F_h$).

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>Scattering density</th>
<th>Structure Factor</th>
<th>Intensity of diffraction spot</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>$\rho(x)$</td>
<td>$F_h = F_i e^{i\phi_h} = \int \rho(x)e^{2\pi i hx} dx$</td>
<td>$I_h =</td>
</tr>
</tbody>
</table>
Estimating crystallographic phases with multiwavelength X-ray data

- Measure diffraction \((I_h, I'_h)\) at two X-ray wavelengths near absorption edge of selenium

- Differences in diffraction are due to changes in scattering from the Se atoms \((\Delta F_h)\)

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>Scattering density</th>
<th>Structure Factor</th>
<th>Intensity of diffraction spot</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_1)</td>
<td>(\rho(x))</td>
<td>(F_h = F_h e^{i\phi_h} = \int \rho(x) e^{2\pi ihx} , dx)</td>
<td>(I_h =</td>
</tr>
<tr>
<td>(\lambda_2)</td>
<td>(\rho'(x) = \rho(x) + \Delta \rho(x))</td>
<td>(F'_h = F_h + \Delta F_h)</td>
<td>(I'_h =</td>
</tr>
</tbody>
</table>
Estimating crystallographic phases with multiwavelength X-ray data

• Measure diffraction ($I_h$, $I_h'$) at two X-ray wavelengths near absorption edge of selenium

• Differences in diffraction are due to changes in scattering from the Se atoms ($\Delta F_h$)

How to figure out where the Se atoms are:

Assume that structure factors for Se are similar to changes between wavelengths:

$$|\Delta F_h| \approx |F_h' - F_h|$$

Then use techniques from small-molecule crystallography to find the Se atoms (guess locations, compare calculated and observed $\Delta F_h$, update guess)
Estimating crystallographic phases with multiwavelength X-ray data

• Measure diffraction ($I_h$, $I'_h$) at two X-ray wavelengths near absorption edge of selenium

• Differences in diffraction are due to changes in scattering from the Se atoms ($\Delta F_h$)

<table>
<thead>
<tr>
<th>Wavelength</th>
<th>Scattering density</th>
<th>Structure Factor</th>
<th>Intensity of diffraction spot</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>$\rho(x)$</td>
<td>$F_h = F_he^{i\phi_h} = \int \rho(x)e^{2\pi ix}dx$</td>
<td>$I_h =</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>$\rho'(x) = \rho(x) + \Delta \rho(x)$</td>
<td>$F'_h = F_h + \Delta F_h$</td>
<td>$I'_h =</td>
</tr>
</tbody>
</table>

If we know where the Se atoms are ...

we know: $\Delta \rho(x)$

...so we can calculate: $\Delta F_h$

and the phase ($\phi_h$) must satisfy: $I'_h = (I_h)^{1/2}e^{i\phi_h} + |\Delta F_h|^2$
Many ways to find the phases

<table>
<thead>
<tr>
<th>Method</th>
<th>Source of phasing information</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIR – single isomorphous</td>
<td>A few heavy atoms (e.g., Hg, Au) in “derivative” contribute to differences from “native”</td>
</tr>
<tr>
<td>replacement</td>
<td></td>
</tr>
<tr>
<td>SAD – single-wavelength</td>
<td>A few atoms (e.g., Se, I, Hg atoms) contribute to “anomalous” differences in diffraction between spot $h$ and spot $-h$</td>
</tr>
<tr>
<td>anomalous diffraction</td>
<td></td>
</tr>
<tr>
<td>MAD – multiple-wavelength</td>
<td>A few atoms contribute to anomalous and wavelength-dependent “dispersive” differences</td>
</tr>
<tr>
<td>anomalous diffraction</td>
<td></td>
</tr>
<tr>
<td>SIRAS, MIR</td>
<td>Combinations of SIR and SAD</td>
</tr>
<tr>
<td>Molecular replacement</td>
<td>Molecular location and phases are found using a related molecule as a template</td>
</tr>
<tr>
<td>Direct methods</td>
<td>Guess where atoms are, good guesses match the measured structure factors</td>
</tr>
</tbody>
</table>
Automation of structure determination

Automation…

makes straightforward cases accessible to a wider group of structural biologists

makes difficult cases more feasible for experts

can speed up the process

can help reduce errors

Automation also allows you to…

try more possibilities

estimate uncertainties
Requirements for automation of structure determination of macromolecules by X-ray crystallography

1. Software carrying out individual steps
2. Seamless connections between steps
3. A way to decide what is good
4. Strategies for structure determination and decision-making
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
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
Histogram of electron density values has a positive “skew”

Low density: Points between atoms and in solvent region

High density: Points on top of atoms

Typical histogram of electron density

Histogram skewed to the right
Skew of electron density for poor and good maps

Poor map: nearly-perfect Gaussian

Good map: slight positive skew
### Evaluating electron density maps

<table>
<thead>
<tr>
<th>Basis</th>
<th>Good map</th>
<th>Random map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skew of density (Podjarny, 1977)</td>
<td>Highly skewed (very positive at positions of atoms, zero elsewhere)</td>
<td>Gaussian histogram</td>
</tr>
<tr>
<td>Connectivity of regions of high density</td>
<td>A few connected regions can trace entire molecule</td>
<td>Many very short connected regions</td>
</tr>
<tr>
<td>(Baker, Krukowski, &amp; Agard, 1993)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation of local rms densities</td>
<td>Neighboring regions in map have similar rms densities</td>
<td>Map has uniform rms density</td>
</tr>
<tr>
<td>(Terwilliger, 1999)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-factor in 1st cycle of density modification (Cowtan, 1996)</td>
<td>Low R-factor</td>
<td>High R-factor</td>
</tr>
</tbody>
</table>
Which scoring criteria best reflect the quality of a map?

- Create real maps
- Score the maps with each criteria
- Compare the scores with the actual quality of the maps
Creating real maps

247 MAD, SAD, MIR datasets with final model available (PHENIX library and JCSG publicly-available data)

Run AutoSol Wizard on each dataset.

Calculate maps for each solution considered (opposing hands, additional sites, including various derivatives for MIR)
Score maps based on each criteria

Calculate map correlation coefficient (CC) to model map (no density modification, shift origin if necessary)

Model map  
1VQB, 2.6 Å, SG C2

SOLVE MAD map  
CC=0.62

Inverse-hand map  
CC=0.55
Skew of electron density – positive skew of density values
Using scoring criteria to estimate the quality of a map

Skew depends on CC

Estimate CC from skew

Skew = 0.4

CC = 0.6 - 0.7
How accurate are estimates of map quality?

Bayesian estimates of CC using Skew

Actual quality

Estimated quality

Cross-validated estimates of quality
### Estimated map quality in practice

Evaluating solutions to a 2-wavelength MAD experiment  
*(JCSG Tm3681, 1VPM, SeMet 1.6 Å data)*

<table>
<thead>
<tr>
<th>Data for HYSS</th>
<th>Sites</th>
<th>Estimated CC ± 2SD</th>
<th>Actual CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak</td>
<td>12</td>
<td>0.73 ± 0.04</td>
<td>0.72</td>
</tr>
<tr>
<td>Peak (inverse hand)</td>
<td>12</td>
<td>0.11 ± 0.43</td>
<td>0.04</td>
</tr>
<tr>
<td>$F_A$</td>
<td>12</td>
<td>0.73 ± 0.03</td>
<td>0.72</td>
</tr>
<tr>
<td>$F_A$ (inverse)</td>
<td>12</td>
<td>0.11 ± 0.42</td>
<td>0.04</td>
</tr>
<tr>
<td>Sites from diff Fourier</td>
<td>9</td>
<td>0.70 ± 0.17</td>
<td>0.69</td>
</tr>
</tbody>
</table>
Statistical density modification (RESOLVE)

• Principle: phase probability information from probability of the map and from experiment:

\[ P(\phi) = P_{\text{map probability}}(\phi) P_{\text{experiment}}(\phi) \]

• “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

• Method:
  • calculate how map probability varies with electron density \( \rho \)
  • deduce how map probability varies with phase \( \phi \)
  • combine with experimental phase information
Map probability phasing: Getting a new probability distribution for each 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 map with expectations
5. Phase probability of reflection k from map is independent of starting phase probability because reflection k is omitted from the map

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. $P(\phi)$ is high for phase that leads to flat region
A map-probability function – allowing different weighting of information from different parts of the map

Log-probability of the map is sum over all points in map of local log-probability

\[ LL^{MAP}(\{F_h\}) \approx \frac{N_{\text{REF}}}{V} \int_{V} LL(\rho(x, \{F_h\})) d^{3}x \]

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

\[ LL(\rho(x, \{F_h\})) = \ln[p(\rho(x)|\text{PROT})p_{\text{PROT}}(x) + p(\rho(x)|\text{SOLV})p_{\text{SOLV}}(x)] \]

If the point is in the PROTEIN region, most values of electron density \( \rho(x) \) are believable

If the point is in the SOLVENT region, only values of electron density near zero are believable
Statistical density modification

(nsf-N SAD map, 2Å, no NCS, 50% solvent)

Phaser SAD map
(CC=0.43)

Phaser +RESOLVE
(CC=0.79)
Statistical density modification with cross-crystal averaging

Cell receptor at 3.5/3.7 Å. Data courtesy of J. Zhu
Structure solution with phenix.autosol

Experimental data, sequence, anomalously-scattering atom, wavelength(s)

Find heavy-atom sites with direct methods (HYSS)

Calculate phases (Phaser/Solve)

Improve phases, find NCS, build model (phase_and_build)

Decisions to be made

Multiple solutions, different derivatives or wavelengths

Alternative hands of space-group and substructure
AutoSol – fully automatic tests with structure library
(MAD datasets, HYSS search, SOLVE)
RESOLVE/ phase_and_build maps

MAD datasets

Density-modified map correlation

pdz  gene-5  synapsin  gpase  gepe  vmp  cp-synthase  tryparedoxin  mei-kinase  psd-95  aeo-transaminase  rab3a  s-hydrodase  cyanase  l-cysteine  armadillo  epsin  p32  lysozyme  p9  nsf-d2  mbp
AutoSol – fully automatic tests with structure library
(SAD datasets, HYSS, Phaser)
RESOLVE/phase_and_build maps

SAD datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Density-modified map correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Adams/CNS</td>
<td></td>
</tr>
<tr>
<td>AutoSol</td>
<td></td>
</tr>
<tr>
<td>phase_and_build</td>
<td></td>
</tr>
</tbody>
</table>

- sec17
- fusion-complex
- 1102B
- 1071B
- 1063B
- 1038B
- nsf-n
- myoglobin
- rna-se-p
- ut-synthese
- 1029B
- 1167B
- cobs
- calmodulin
AutoSol – fully automatic tests with structure library
(SAD datasets, HYSS, Phaser)
RESOLVE/ phase_and_build maps

MIR datasets

- Paul Adams/CNS
- AutoSol
- phase_and_build

Density-modified map correlation

- hn-rnp
- rop
- rnase-s
- flr
- synaptotagmin
- rh-dehalogenase
- penicillinopepsin
- granulocyte
- qaprtase
- groEL
Iterative density modification, model-building and refinement with phenix.autobuild

Experimental data, sequence, phase information or starting model

Density modification

Model-building and refinement

Resolve building
Secondary-structure only
Connect chains
Fit loops
Build outside model
RESOLVE model-building at moderate resolution

- FFT-based identification of helices and strands
- Extension with tripeptide libraries
- Probabilistic sequence alignment
- Automatic molecular assembly
Initial model-building – strand fragments
Chain extension
(result: many overlapping fragments)
Main-chain as a series of fragments
(choosing the best fragment at each location)
Side-chain template matching to identify sequence alignment to map (IF5A data)
Relative probability for each amino acid at each position
(Correct amino acids in bold)

<table>
<thead>
<tr>
<th></th>
<th>G A S V I L M C F Y K R W H E D Q N P T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6 5 4 18 18 6 1 1 1 1 2 6 2 2 1 9 6 1 0 1 4</td>
</tr>
<tr>
<td>2</td>
<td>4 11 14 37 5 2 0 2 0 0 2 3 0 0 1 2 0 0 0 6</td>
</tr>
<tr>
<td>3</td>
<td>11 23 5 12 5 3 2 0 1 3 7 3 1 0 5 3 2 0 2 2</td>
</tr>
<tr>
<td>4</td>
<td>7 9 6 16 8 5 2 0 1 3 8 4 1 0 7 6 2 0 3 4</td>
</tr>
<tr>
<td>5</td>
<td>31 7 3 7 4 2 1 0 1 3 5 4 1 0 6 2 2 0 11 1</td>
</tr>
<tr>
<td>6</td>
<td>1 3 3 41 14 8 0 0 0 0 0 2 1 0 0 2 4 0 0 1 9</td>
</tr>
<tr>
<td>7</td>
<td>0 0 0 0 0 0 0 0 0 15 63 1 0 17 1 0 0 0 0 0</td>
</tr>
<tr>
<td>8</td>
<td>2 3 6 23 10 6 2 1 0 1 4 3 0 0 5 16 1 0 1 6</td>
</tr>
<tr>
<td>9</td>
<td>96 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>
Addition of side-chains to fixed main-chain positions
AutoBuild – tests with structure library
Fully automated iterative model-building, final R/Rfree

![Graph showing the relationship between AutoBuild R and FreeR values across different resolutions.](image-url)
Rapid building of models for regions containing regular secondary-structure

Helices:

Identification: rods of density at low resolution

Strands:

Identification: \(\beta\) structure as nearly-parallel pairs of tubes

Any protein chains (trace_chain):

Identification: \(C\alpha\) positions consistent with density and geometry of protein chains

RNA/DNA:

Identification: match of density to averaged A or B-form template
Model $\alpha$-helix; 3 Å map
Model $\alpha$-helix; 7 Å map
Trace main-chain with ideal helix, allowing curvature

2 Å radius, 5.4 Å/turn ideal helix
Identify direction and C\(\alpha\) position from overlap with 4 Å radius helices offset +/- 1 Å from main-chain

4 Å radius, 5.4 Å /turn ideal helix offset +1 Å along x

4 Å radius, 5.4 Å /turn ideal helix offset -1 Å along x
A real case: 1T5S SAD map (3.1 Å) Data courtesy of P. Nissen
A real case: 1T5S SAD map (7 Å)
Finding helices in 1T5S SAD map (7 Å)
Finding helices in 1T5S SAD map (3.1 Å)
Helices from 1T5S SAD map compared with 1T5S (3.1 Å)
$C_\alpha$ tracing
(s-hydrolase, PDB entry 1A7A)
C$_\alpha$ tracing
(mevalonate kinase, PDB entry 1KKH, 9 sec)
Using secondary structure content to evaluate map quality
Building RNA
Group II intron at 3.5 Å. Data courtesy of J. Doudna
Rapid phase improvement and model-building with phenix.phase_and_build

First improve the map

NCS identification from density
Iterative rapid model-building and density modification

Then build a full model

Model-building and refinement with NCS
Comprehensive sequence assignment
Loop fitting
phenix.phase_and_build – tests with structure library

Final R/Rfree
phase_and_build – tests with structure library
One cycle (approx 500 residues/hour)
What can you do with automated procedures for structure solution and model-building?

If a task is modular and automated…

you can run it many times

…checking different space groups, datasets to use

…checking if your model is biasing your map

…checking if you always get the same model
Iterative-Build OMIT procedure

“Is the density in my map biased by the model?”

2mFo-DFc omit map

After building outside OMIT region 10 cycles

1HP7 molecular replacement with 1AS4

R/Rfree after initial refinement: 0.41/0.48
Iterative-Build OMIT procedure
“Removing model bias”

2mFo-DFc map
Phased with 1zen model
Iterative-Build OMIT procedure
“Removing model bias”

2mFo-DFc omit map
Phased with 1zen model
Iterative-Build OMIT procedure
“Removing model bias”

2mFo-DFc SA-omit map
Phased starting with 1zen model
Iterative-Build OMIT procedure
“Removing model bias”

2mFo-DFc iterative-build omit map
Phased starting with 1zen model
Multiple-model representation of uncertainties

*20 models built for 1CQP, no waters, Dmin=2.6 Å  
\( R=0.19-0.20; R_{\text{free}}=0.26-0.27 \)*

The variation among models is a lower bound on their uncertainty.
Building 20 models for each of 10 structures
The RMSD among models tells us (a lower bound on) the uncertainty in our models

(It is not the RMSD of true structures in the crystal)

Rebuild with 4.5 Å data  
Rebuild with 1.75 Å data
Complementarity of PHENIX and Rosetta model-building
(Randy Read, David Baker, Frank DiMaio)

- **Characteristic**
  - Optimization
  - Model-building approach
  - Fragment libraries
  - Model-building target
  - Refinement target

- **Crystallographic model-building (PHENIX)**
  - Interpretation of patterns of density
  - Density search for regular secondary structure
  - 3-residue fragment library
  - Fit to density
  - Structure-factor likelihood refinement target

- **Structure-modeling (Rosetta)**
  - Creating physically plausible models
  - Ab-initio modeling or homology modeling
  - 3- and 9-residue libraries
  - Rosetta force field with density term
  - Rosetta force field with density term
Combining structure-modeling with crystallographic model-building

20 templates for 1XVQ from PDB (optimally superimposed)
Molecular replacement using distant homology models with PHENIX and Rosetta (phenix.mr_rosetta)

Identify search models with homology server (http://toolkit.tuebingen.mpg.de/hhpred)

Sequence-specific Rosetta fragments files from Robetta server (http://robetta.bakerlab.org/fragmentsubmit.jsp)

Parameters file listing sequence, homology file, fragments files

Automated download and editing of search models and creation of alignment files

- target.hhr
- fragments_3.gz
- fragments_9.gz
- target.eff
- target.ali
- target_mr.pdb
Molecular replacement using distant homology models with PHENIX and Rosetta (phenix.mr_rosetta)

Molecular replacement and model-building

Molecular replacement (Phaser)
Refine (phenix.refine) to create new map

Build 20-2000 models with Rosetta
(including density term in Rosetta energy)

Pick models with highest Phaser LLG from top 10% of Rosetta models

Refine and autobuild (phenix.autobuild) starting from top rosetta models

iterate until convergence

target_1.pdb
refine_map_coeffs.mtz

targ.1.pdb

refine_map_coeffs.mtz

targ.1.pdb

targ.1.pdb

targ.1.pdb

targ.1.pdb

target_rosetta_0001.pdb

target_rosetta_best.pdb

target_autobuild.pdb

Molecular replacement and model-building
Structure determination of cab55348 (using template supplied by user)  
1.9 Å, 28% sequence identity (AutoMR alone fails with R/Rfree=0.47/0.53) 
MR model: blue, Final model: green
Sample Rosetta models in cycles 1 and 2.
MR model: blue

Final model: pink

Map from refined MR model

Rebuilding in a poor section of the starting model
MR model: blue

Final model: pink

Rosetta models cycle 1: green

Map from refined MR model

Rebuilding in a poor section of the starting model
MR model: blue    Final model: pink

Best-scoring Rosetta model cycle 1: yellow

Map from refined MR model

Rebuilding in a poor section of the starting model
MR model: blue  Final model: pink  Rosetta models cycle 1: green

AutoBuild Map from cycle 1

Rebuilding in a poor section of the starting model
MR model: blue  Final model: pink  Rosetta models cycle 2: yellow

AutoBuild Map from cycle 1

Rebuilding in a poor section of the starting model
AutoBuild model cycle 2

Map CC: 0.78  R/Rfree = 0.26/0.31

Rebuilding in a poor section of the starting model
Wizards

• **AutoSol Wizard:** Structure solution (MIR/MAD/SAD) with HYSS/Phaser/Solve/Resolve

• **AutoBuild Wizard:** Iterative density modification, model-building and refinement with Resolve/phenix.refine/Elbow; model rebuilding in place; touch-up of model; simple OMIT; SA-OMIT; Iterative-build OMIT; OMIT around atoms in a PDB file; protein, RNA, DNA model-building

• **LigandFit Wizard:** automated fitting of flexible ligands

• **AutoMR Wizard:** Phaser molecular replacement followed by automatic rebuilding
MODEL-BUILDING TOOLS

- **phenix.find_ncs**: Find and evaluate NCS from density, heavy-atom sites, or model
- **phenix.apply_ncs**: Apply NCS operators to a single chain
- **phenix.build_one_model**: Resolve rapid model-building with real-space refinement
- **phenix.phase_and_build**: Improve map by model-building and refinement, then build full model
- **phenix.find_helices_strands**: Trace chain or build secondary structure from a map
- **phenix.mr_rosetta**: Combine Rosetta structure-modeling with Phenix
• **phenix.refine**: fully automatic/fully flexible refinement, SA-refinement, NCS identification, TLS, torsion-angle refinement, twin refinement

• **phenix.xtriage**: twinning, twin laws, anisotropy, anomalous signal, outliers, space group

• **phenix.builder**: ligand structures and CIF definitions from SMILES, PDB....

• **phenix.ligand_identification**: identify ligand density with class-specific libraries

• **phenix.validation**, **phenix.model_vs_data**, **phenix.real_space_correlation**, **phenix.get_cc_mtz_mtz**: Molprobity and density analysis of structures and density maps

• **phenix.pdbtools**, **phenix.reflection_file_editor**: manipulate PDB and mtz files

• ...and many more: see **phenix.doc** and **www.phenix-online.org**