Experimental phasing

Charles Ballard
(original Gábor Bunkóczi)
CCP4 Workshop
7 December 2011
Anomalous diffraction

$F_P$ – protein

$F_A$ – anomalous substructure

$\Delta_{\text{ano}}$
Phasing

Substructure location → Anomalous substructure

Substructure refinement → Electron density map

Phase calculation
Phase calculation in SAD

\[ F_o^+, F_o^- \] – measurements

\[ F_o^+ - F_o^- \] – calculated from substructure \((-2F_A'')\)

\[ \alpha_1, \alpha_2, \alpha_{\text{best}} \] – phase angles

Approximate map given by

\[ mF_{\text{best}} e^{i\alpha_{\text{best}}} \]

With \( F_{\text{best}} = \frac{|F^+| + |F^-|}{2.0} \), \( m = \cos \left( \frac{1}{2} (\alpha_1 - \alpha_2) \right) \)
SAD likelihood function

Probability of experimental observations. Observations correlated, multivariate distribution.

\[
P(F_O^+, F_O^{-}; H^+, H^{-*}) \rightarrow P(F_O^+, F_O^{-}; H^+, H^{-}) \rightarrow P(F_O^+, \alpha^+, F_O^-, \alpha^-; H^+, H^{-})
\]

Factor joint probability into two parts

\[
P(F_O^+, F_O^{-}; H^+, H^{-*}) = P(F_O^+; F_O^-, H^+, H^{-*})P(F_O^-, H^{-*})
\]

Marginalise phases \(\alpha^+\) and \(\alpha^-\)
SAD function – normal scattering

\[ P(F_o^-, \alpha^-; H^-) = \]

\[ \frac{F_o^-}{\pi \Sigma^-} \exp \left( - \frac{\left| F_o^- e^{i\alpha^-} - DH^{-*}\right|^2}{\Sigma^-} \right) \]

\(\Sigma^-\) accounts for errors between \(F_o^-\) and \(H^{-*}\).

Primarily normal scattering.
SAD function - anomalous

\[ P(F^+_o; F^-_o, H^+, H^-^*) = \]

\[ \frac{2F^+_o}{\Sigma^+} \exp \left( -\frac{(F^+_o - F^+_c)^2}{\Sigma^+} \right) e^{i\theta} \left( \frac{2F^+_o F^+_c}{\Sigma^+} \right) \]

\( \Sigma^+ \) accounts for errors between \( F^+_o \) and \( F^+_c \), where

\[ F^+_c = |H^+ + D_\phi e^{i\alpha_\phi} (F^- e^{i\alpha_-} - H^-^*)| \]
SAD function

Total likelihood is integral product of the distributions under the black circle

**Normal scattering**

**Anomalous scattering**

Phase probability
Phaser phasing scenarios

**Known anomalous substructure**
- missing atoms
- model errors (xyz, occ, B)
- weak normal component

**Known partial structure**
- missing (anomalous) atoms
- (partial) model errors
- weak anomalous component
SAD phasing

Parameters:
- atom type
- position
- occupancy
- B-factor
- $f''$

Substructure atoms $H^+, H^-$ estimates SAD-function
MR-SAD phasing

Input:
- partial structure
- possible atom types

Heavy atom parameters

$H^+, H^-$ estimates

SAD-function
SAD function

Scenario:
anomalous substructure

Scenario:
partial MR model
Log-likelihood gradient maps

Peak height dependent on $f’/f”$ ratio
- distinguish atom type
- (fooled if normal scattering is present at site)

Very sensitive to detail
- e.g. anisotropic temperature motion

Iterative
- improvements in substructure description increases sensitivity
Log-likelihood gradient maps

<table>
<thead>
<tr>
<th>Element</th>
<th>( \Delta_{\text{ano}} )</th>
<th>LLG1</th>
<th>LLG2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Br</td>
<td>26.5</td>
<td>28.5</td>
<td>-</td>
</tr>
<tr>
<td>Co</td>
<td>5.0</td>
<td>5.0</td>
<td>18.5</td>
</tr>
<tr>
<td>Na1</td>
<td>-</td>
<td>-</td>
<td>7.6</td>
</tr>
<tr>
<td>Na2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Na3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Element</th>
<th>( f' )</th>
<th>( f'' )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>P</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td>K</td>
<td>0.25</td>
<td>0.41</td>
</tr>
<tr>
<td>Ca</td>
<td>0.28</td>
<td>0.50</td>
</tr>
</tbody>
</table>
F1-ATPase testcase

3500 residue protein complex
81 S atoms, 3.50 Å resolution
Wavelength: 1.77 Å ($f_s''=0.72$ e$^{-}$)
Anomalous signal extends to 21.5 Å, or to 12.7 Å very optimistically.

1bmf -
Abrahams J.P., Leslie A.G.W., Lutter R., Walker J.E.
Nature (1994) **370**:621

2ck3 - high resolution (1.95 Å ) equivalent
Bowler M.D., Montgomery M.G. Leslie, A.G Walker J.E.
### F1 ATPase testcase

<table>
<thead>
<tr>
<th>pdb 2CK3</th>
<th>Present sites</th>
<th>Found sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Met</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td>73</td>
<td>8</td>
</tr>
<tr>
<td>F1-ATPase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 ATPase 3α chains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeast ATPase</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
F1 ATPase testcase

ADP in active site
(not included in model)

Located sites
Phaser

Macrocycle
- Outlier rejection
- Substructure refinement

Substructure editing
- Find new atoms in LLG map
- Reject atoms with low occupancy

Phase calculation

Automated mode

convergence?
Data flow in substructure solution

Measured data → Anomalous differences

Electron density map → Anomalous substructure → Solution

Phasing
Rescore trials with phaser

Insulin full search against 1.7 Å data

- Clearly separated solution and non-solution groups
- Impossible to identify solutions
CC for phaser-refined substructures

Insulin full search against 1.7 Å data

Increased CC for solutions
Completion: improved substructure

Lower CC for non-solutions
Refinement: reduced overfitting

Overall: better discrimination
Fibronectin testcase

Full search with CC rescoring

Solution buried in noise

Rudiño-Piñera E, Ravelli RB, Sheldrick GM, Nanao MH, Korostelev VV, Werner JM, Schwarz-Linek U, Potts JR, Garman EF

*J Mol Biol.* **368**, 833-44
Fibronectin testcase

Full search with phaser rescoring

Rudiño-Piñera E, Ravelli RB, Sheldrick GM, Nanao MH, Korostelev VV, Werner JM, Schwarz-Linek U, Potts JR, Garman EF

*J Mol Biol.* **368**, 833-44
Substructure seeding

Patterson seeding

Dual-space recycling

Refinement and scoring

Substructure is available at this point

Compare:
- atoms compatible with the difference Patterson
- atoms compatible with the SAD LLG map
## Comparison

### Transaminase, 60 Se sites

<table>
<thead>
<tr>
<th>Method</th>
<th>Patterson seeding</th>
<th>LLG-map seeding</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC scoring</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>Phaser scoring</td>
<td>20</td>
<td>25</td>
</tr>
</tbody>
</table>

### KSM42, 160 Se sites

<table>
<thead>
<tr>
<th>Method</th>
<th>Patterson seeding</th>
<th>LLG-map seeding</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC scoring</td>
<td>23</td>
<td>28</td>
</tr>
<tr>
<td>Phaser scoring</td>
<td>28</td>
<td>37</td>
</tr>
</tbody>
</table>
When to use

- Phaser scoring
  - default run
  - default give no solution

- LLG-map seeding
  - default run
  - large substructures
  - with phaser scoring
**Phaser**

**Mode**
- SAD (start from heavy atoms)
- MRSAD (start with partial MR model)

**Reflection data**
- $F^+/F^-$
- possible space groups

**Scatterer information**
- atom type and wavelength
- measured $f'/f''$ values

**Absolute scale**
- macromolecule weight
Results

Phase information
- FWT, PHWT
- HLA, HLB, HLC, HLD
- PHIB
- FLLG, PHILLG

PDB file
anomalous substructure

Best map or initial map for density modification

Phase probability for density modification or refinement

Centroid phase for density modification/refinement (use HL-coefficients!)

Log-likelihood gradient map (flat after automated mode!)

Logfile
www.phaser.cimr.cam.ac.uk
Acknowledgements

Python-based Hierarchical Environment for Integrated Xtallography

Computational Crystallography Initiative / LBNL
Paul Adams, Ralf Grosse-Kunstleve, Pavel Afonine, Nathaniel Echols, Jeffrey Headd, Nigel Moriarty, Nicholas Sauter, Peter Zwart

Los Alamos National Lab
Tom Terwilliger, Li-Wei Hung

University of Cambridge
Randy Read, Airlie McCoy, Robert Oeffner, Gabor Bunkoczi

Duke University
Jane and Dave Richardson, Vincent Chen