Molecular Replacement

Ronan Keegan
(original by Gabor Bunkoczi)

OIST/CCP4 Workshop
6 December 2011
Idea

<table>
<thead>
<tr>
<th>h</th>
<th>k</th>
<th>l</th>
<th>F</th>
<th>φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>12.6</td>
<td>123</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2.1</td>
<td>12</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>3</td>
<td>69.9</td>
<td>287</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>h</th>
<th>k</th>
<th>l</th>
<th>F</th>
<th>φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10.4</td>
<td>113</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3.5</td>
<td>18</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>3</td>
<td>57.2</td>
<td>265</td>
</tr>
</tbody>
</table>
Problem

Previous crystal form

Current crystal form

\( R(\phi, \psi, \kappa, x, y, z) \)
Separability

\[(\phi, \psi, \kappa) + (x, y, z)\]
Model error

Position errors

Calculate from sequence identity (Chothia & Lesk, 1986)

Model incompleteness

Calculate from unit cell composition
Likelihood function

\[ F_c \] – model structure factor
Dependent on orientation and position of model

\[ D \] – Luzzati-factor \((0 \leq D \leq 1)\)
Dependent on model errors

\[ \sigma_{\Delta} \] – uncertainty
Dependent on model errors and incompleteness
Reflection likelihood

\[ DF_c(R_1, r_1) \]

\[ F_{\text{obs}} \]
Reflection likelihood

\[ D \mathbf{F}_c(R_2, r_2) \mathbf{F}_{\text{obs}} \]
Reflection likelihood

\[ F_{\text{obs}}(R_3, r_3) \]
More about Maximum Likelihood


Using ML methods to account for errors and incompleteness in the search model, Phaser can increase the chances of successful molecular replacement.
Phaser workflow
Molecular replacement workflow

Rotation search

Translation search

Refinement

Packing test
Brute search

Calculate likelihood for each gridpoint
Brute search

Calculate likelihood for each gridpoint

Find peaks
Fast search

Calculate approximation by FFT

Peak search

“Rescore” with likelihood
Rotation search

⇒ Typically small signal
⇒ Dependent only on the point group
Rotation peaks

Centre of coalesced peak

Translation function is very sensitive to orientation and may not find a solution in such a case!
Translation search

⇒ Solutions are recognizable (high Z-score)
⇒ Very sensitive to correct orientation
Packing

Clashes between models decrease credibility in solution
Packing

Surface loops may change conformation
Packing

A small number of clashes is acceptable
The rotation and translation functions were performed on a (not very fine) grid.

The solution can be improved if the grid is taken away and the rotational and translational parameters optimized.
Identifying solutions

![Diagram showing the relationship between TFZ and count, with categories for solved and unsolved cases.]
Partial structure

Log Likelihood gain vs. FRF (% of best)
Partial structure
Model search order

65% of structure
35% identical model

35% of structure
50% identical model

Which model is easier to find?
Model search order

65% of structure
35% identical model

Better model if data resolution is low

35% of structure
50% identical model

Better model if data resolution is high

Determined automatically!
Tree search with pruning

1\textsuperscript{st} component

Pruned

Propagated

Perform search

2\textsuperscript{nd} component
Running Phaser via CCP4i

- **mode**
- **MTZ file**
- **target details**
- **search model**
- **specify search**
- **RUN IT!**
Fast Rotation Function

Euler angles (CCP4)

Top LLG and Z-scores for FRF
Fast Translation Function

Top LLG and Z-scores for FRF

FRF solution number

Fractional translation
Packing

Phaser does packing check after FTF

**Clashes** = C\(\alpha\) atoms closer than 2Å

Default number of clashes = 0

Think about increasing to 2 or 5
Solution Files

.sol file produced at end of job
• Contains summary of all solutions
• Each solution contains rotations and usually translations - 3DIM vs 6DIM
• One line per model located
• .sol file can be read back into Phaser in later jobs

<table>
<thead>
<tr>
<th>Z-score</th>
<th>Have I solved it?</th>
</tr>
</thead>
<tbody>
<tr>
<td>less than 5</td>
<td>no</td>
</tr>
<tr>
<td>5 - 6</td>
<td>unlikely</td>
</tr>
<tr>
<td>6 - 7</td>
<td>possibly</td>
</tr>
<tr>
<td>7 - 8</td>
<td>probably</td>
</tr>
<tr>
<td>more than 8</td>
<td>definitely</td>
</tr>
</tbody>
</table>

RFZ = RF Z-score
TFZ = TF Z-score
Models for molecular replacement

- Ab initio models (e.g. Rosetta)
- Structures of homologous proteins
- "mixed" models
- unedited
- homology models
Model improvement

- Structural information in MR models
- Extra information in sequence alignment

Improve model by removing residues that are not present or significantly different in the target
Binary concept

**delete**

```
LGHQH---GS------RYVM
FRKCTNNASKYLCYLF
```

**keep**

CHAINSAW (CCP4)
Sequence similarity

weighting

matrix

LGHQH--GS---RYVM

FRKCTNNASKYLCYL

window  window

Sculptor (CCP4++)
Correcting for insertions

CHAINSAW with structure-based alignment

CHAINSAW

Sculptor

HCYKS--------GIQVR

HCVNSYQSNLDAIKIR

profile-profile sequence-based alignment (FFAS)
Correcting for deletions

CHAINSAW with structure-based alignment

HCVNYSYQSNLDAIKIR

HCYKSK-------GIQVR

profile-profile sequence-based alignment (FFAS)

CHAINSAW

Sculptor
Downweighting

Sequence similarity (residue level)  
Accessible surface area (atom level)
Multi-model strategy

Calculations with CLUSTALW alignments

- Model: CHAINSAW
- Model: Sculptor
- Model: Sculptor + ASA B-factors
- Model: Sculptor + sequence similarity B-factors

Graph showing number of models vs. sequence identity:
- Unsolved
- Multi-model
- CHAINSAW
Interactive model editing
Acknowledgements

Python-based Hierarchical Environment for Integrated Xtallography

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

\[ F_{\text{obs}} \]
Rotation and translation functions

Identical mathematical forms!

Translation function:
Reflection amplitude likelihood

Rotation function:
Reflection amplitude likelihood with $\sigma_\Delta$ is increased to account for unknown translation