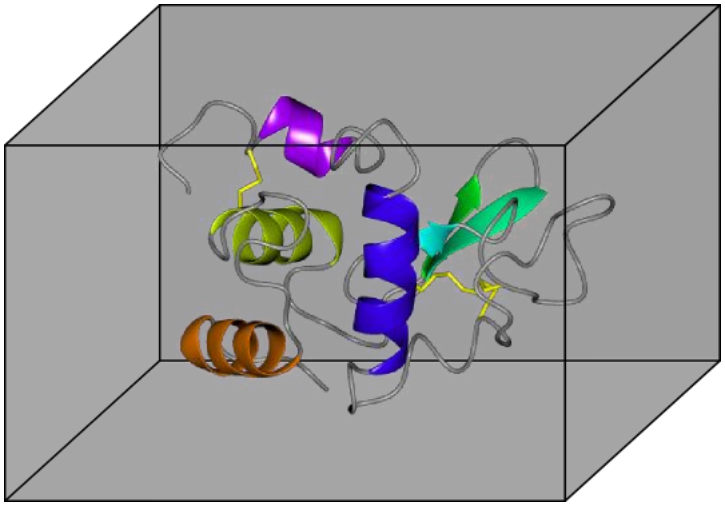
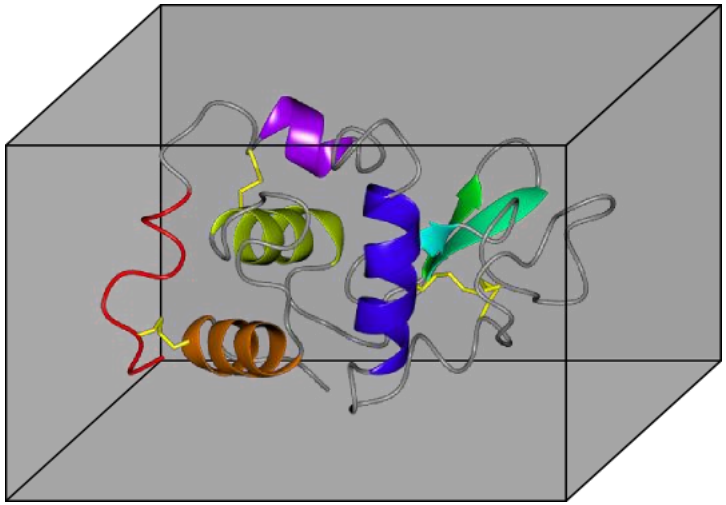
Molecular Replacement

Ronan Keegan
(original by Gabor Bunkoczi)

OIST/CCP4 Workshop
6 December 2011

Idea

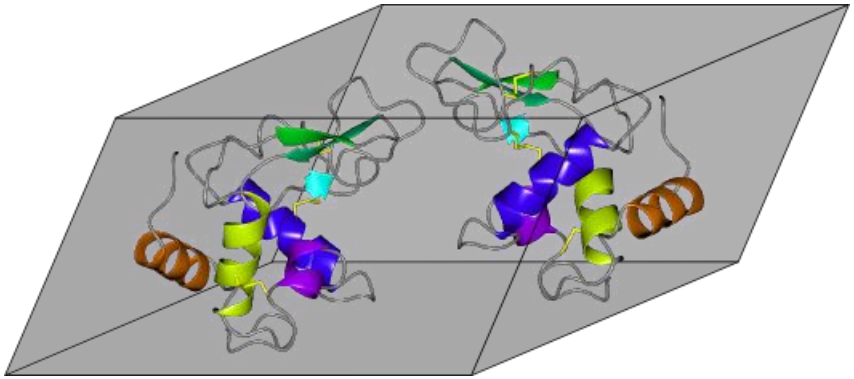


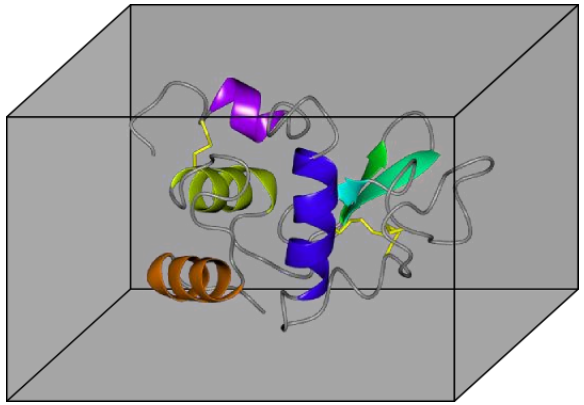


h	k	l	F	ϕ
0	0	1	12.6	123
0	0	2	2.1	12
0	0	3	69.9	287

h	k	l	F	ϕ
0	0	1	10.4	113
0	0	2	3.5	18
0	0	3	57.2	265

Problem



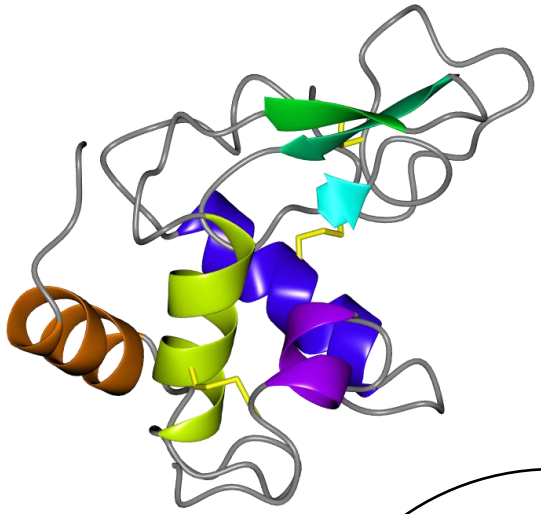


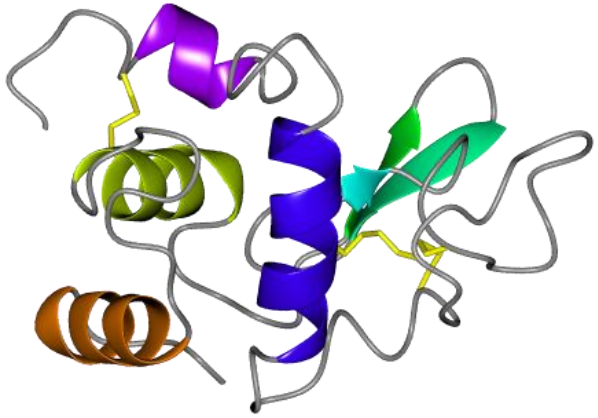
Previous crystal form

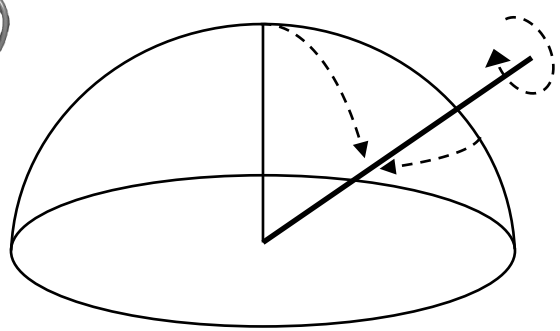
Current crystal form

$$\Re\left(\phi,\psi,\kappa,x,y,z\right)$$

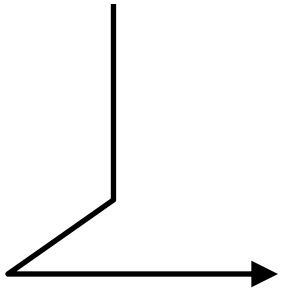
Separability







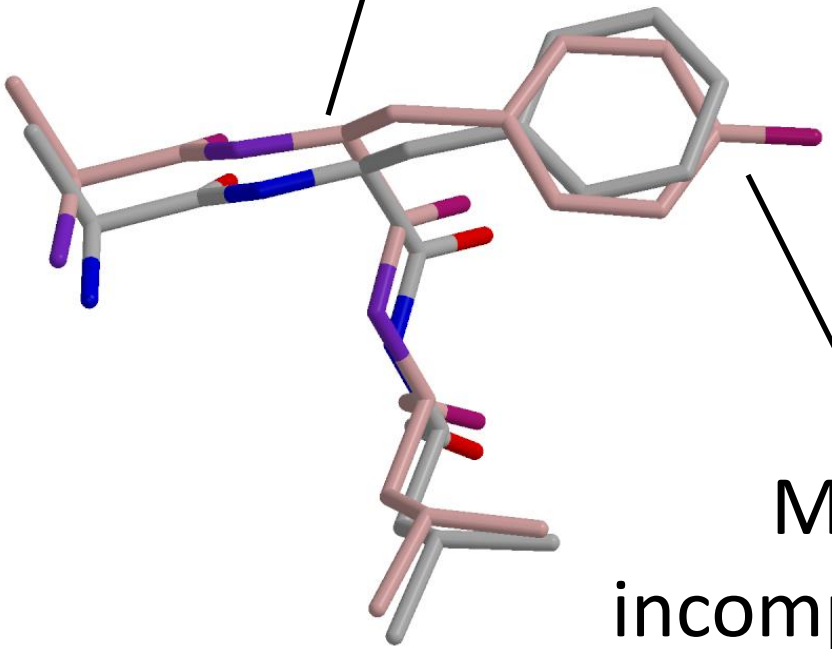
+



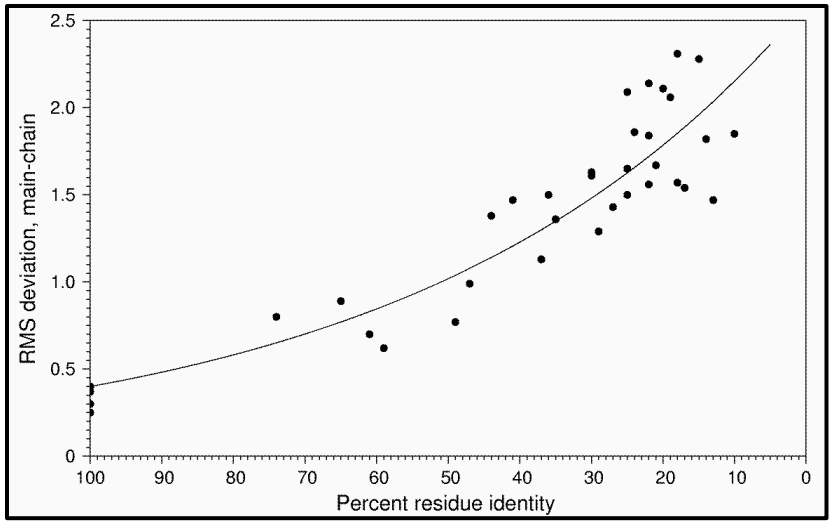
(φ, ψ, κ) (x, y, z)

Model error

Position errors



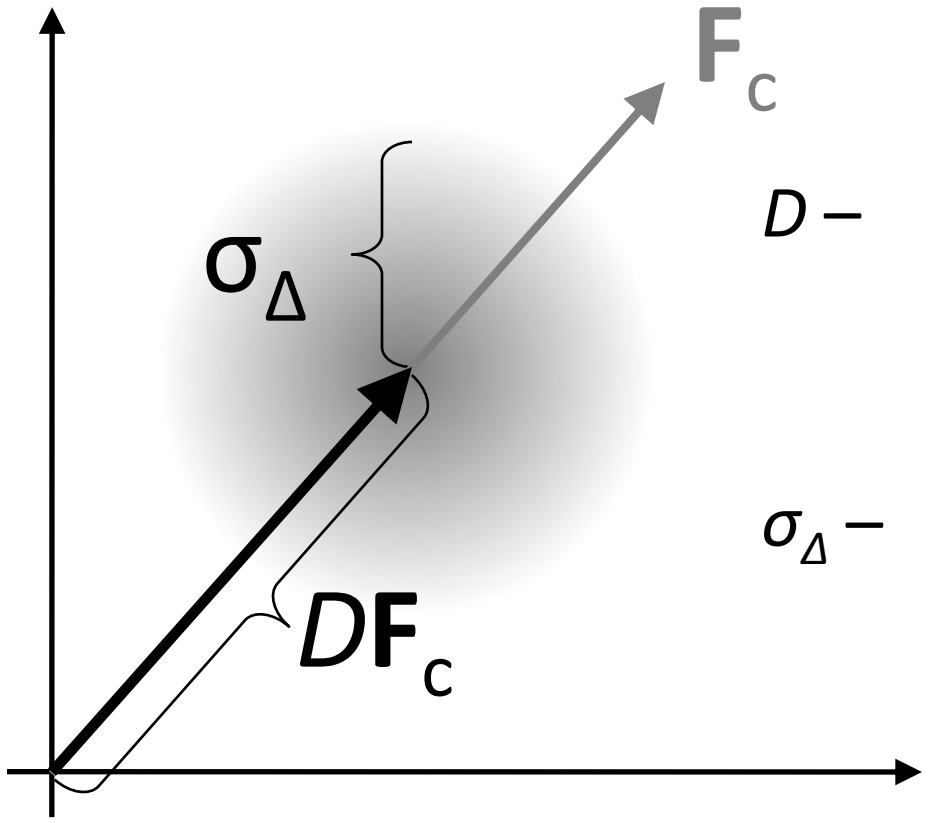
Calculate from sequence identity (Chothia & Lesk, 1986)



Model incompleteness

Calculate from unit cell composition

Likelihood function

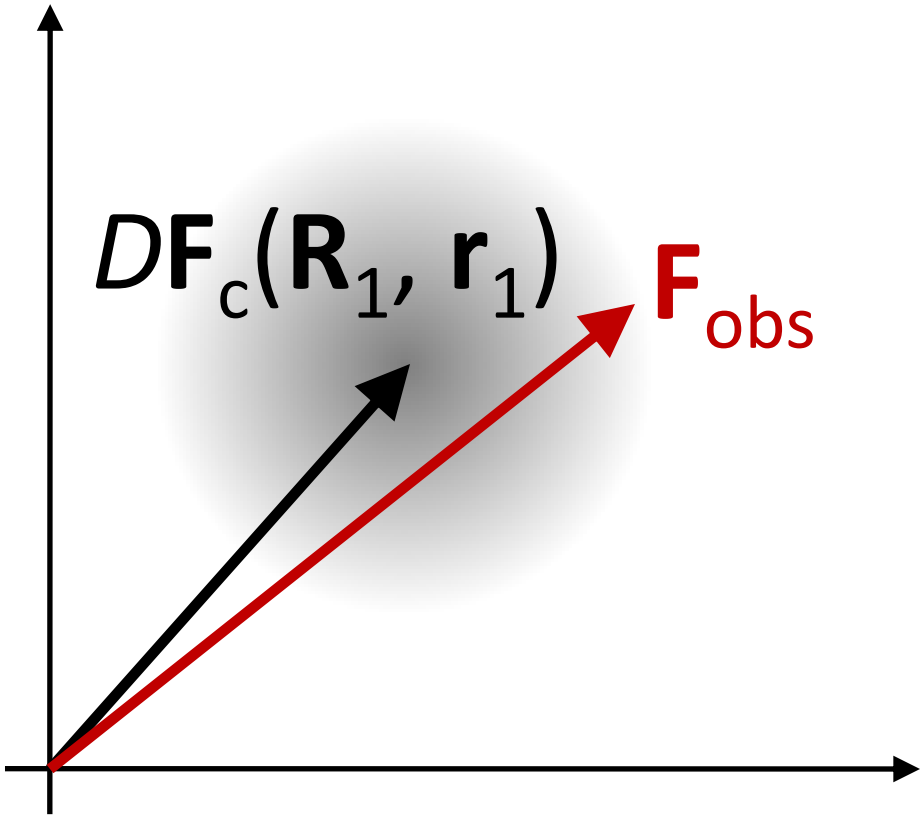


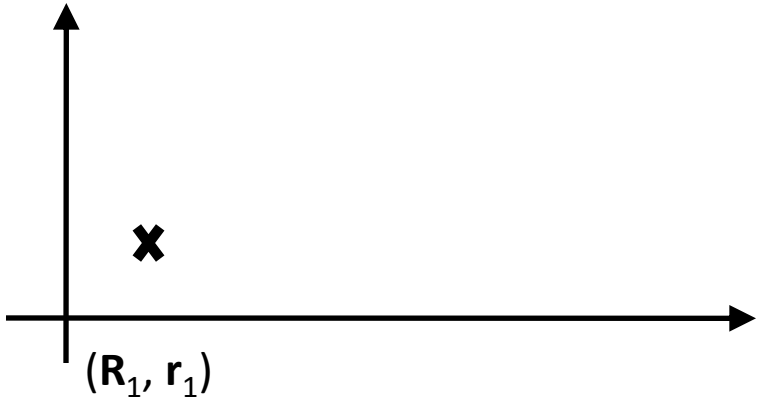
$\mathbf{F}_c$ – model structure factor
Dependent on orientation and position of model

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

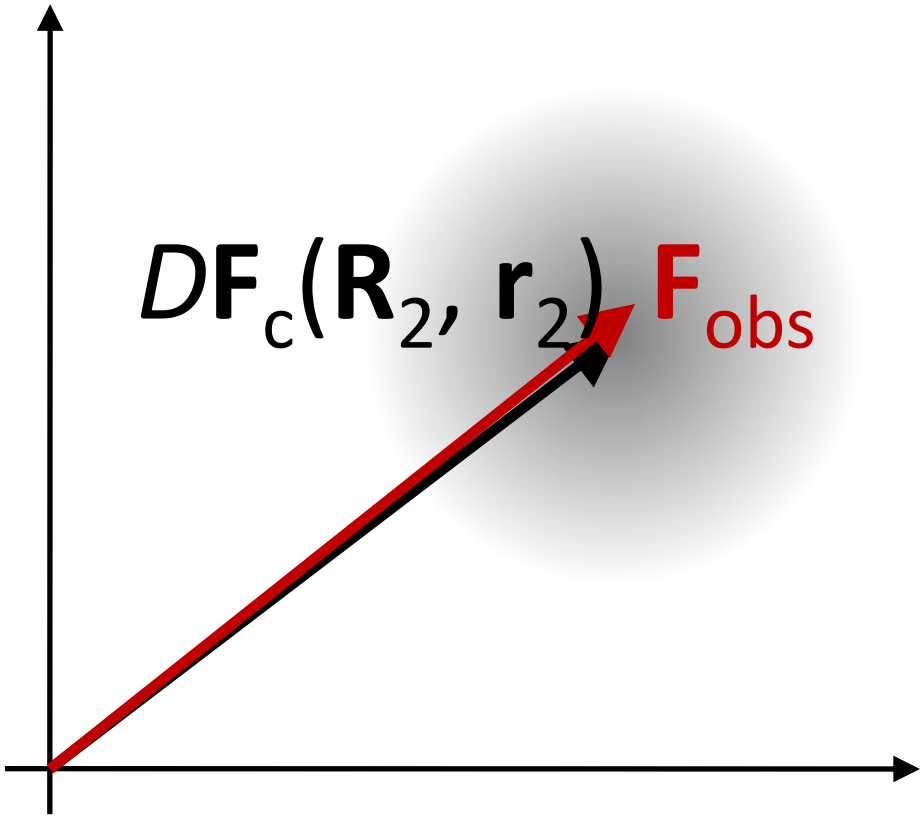
σ_Δ – uncertainty
Dependent on model errors and incompleteness

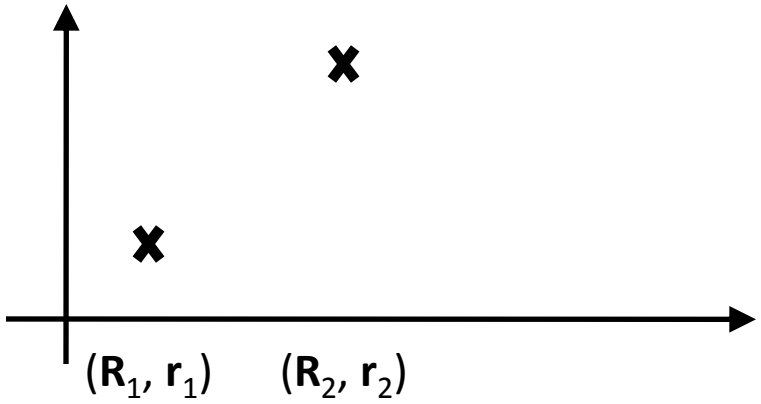
Reflection likelihood



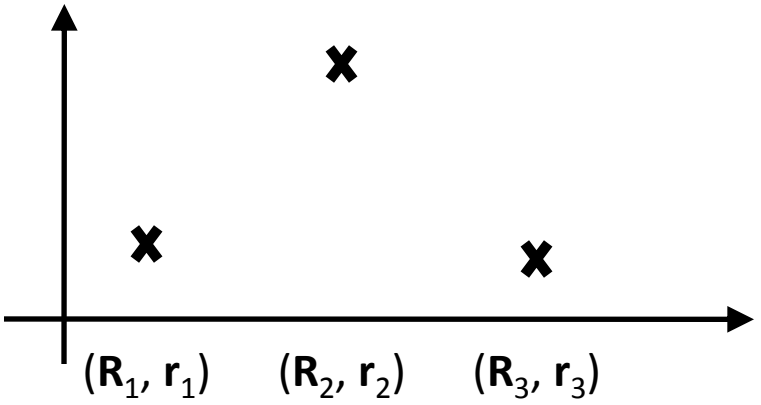


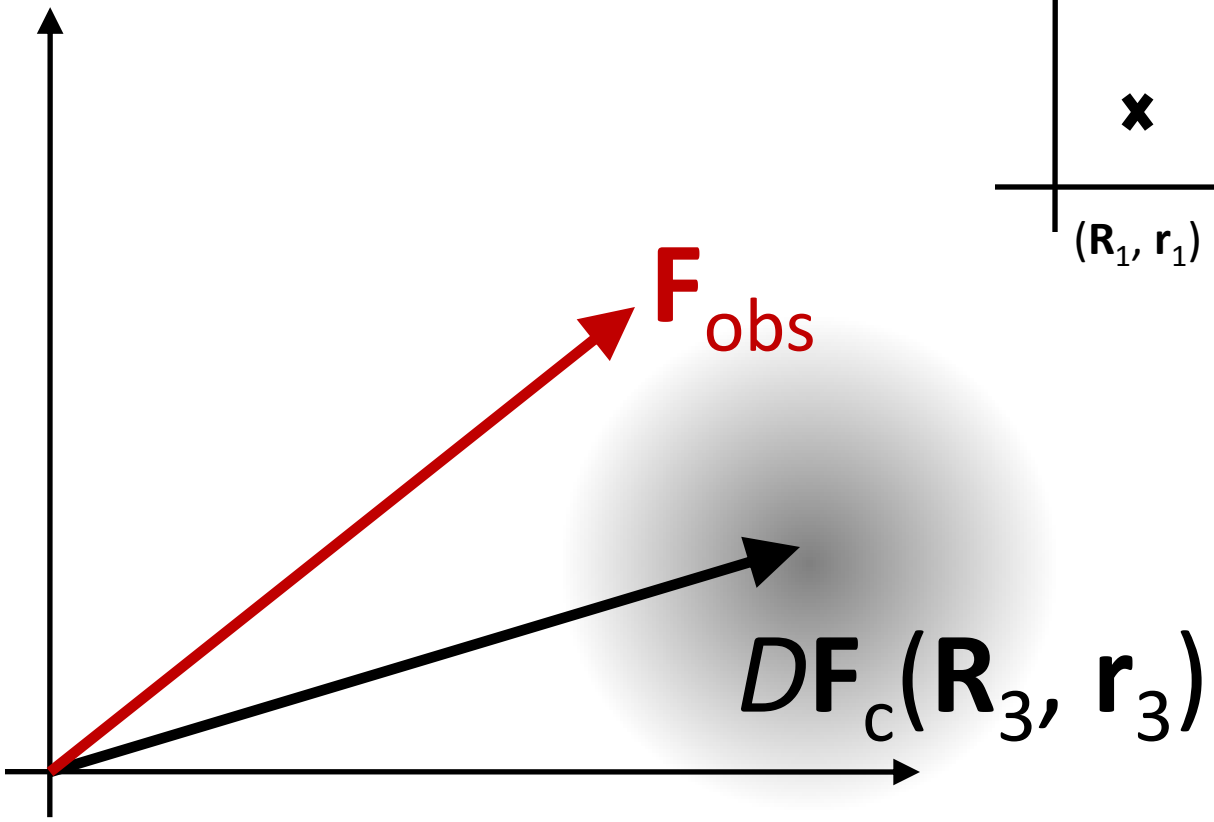
Reflection likelihood





Reflection likelihood



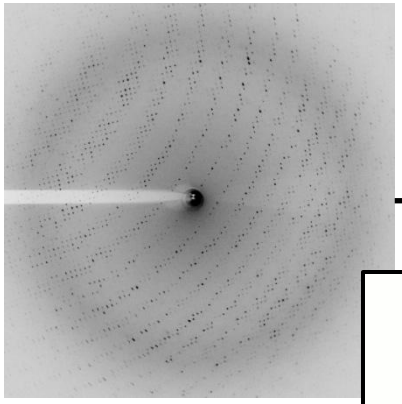


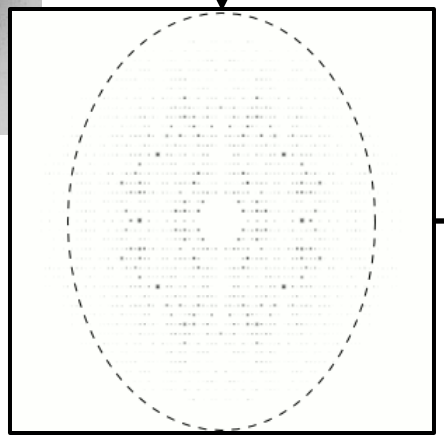
More about Maximum Likelihood

"Liking Likelihood" – A. McCoy (2004),
*Acta Cryst **D**. 60; 2169*

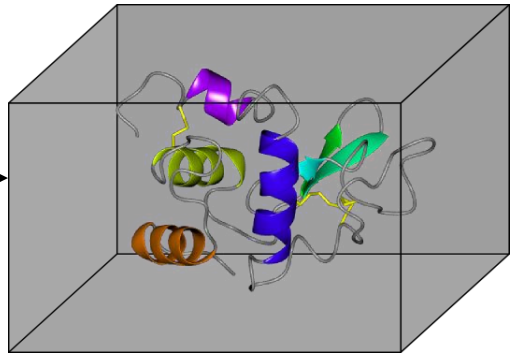
Using ML methods to account for errors and incompleteness in the search model Phaser can increase the chances of successful molecular replacement

Phaser workflow

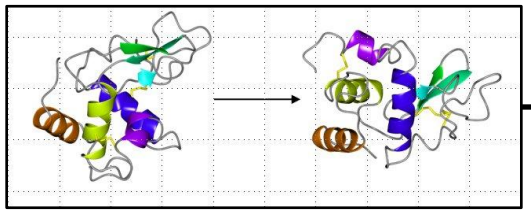




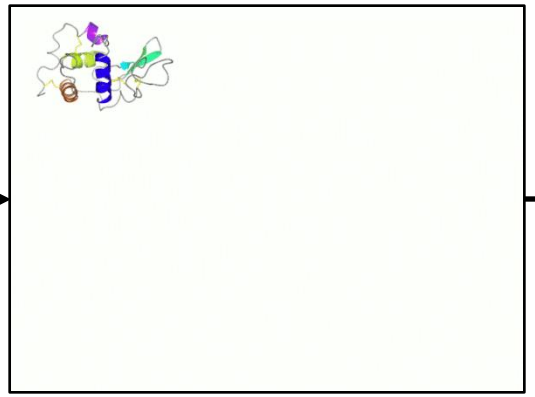
Molecular Replacement



Molecular replacement workflow

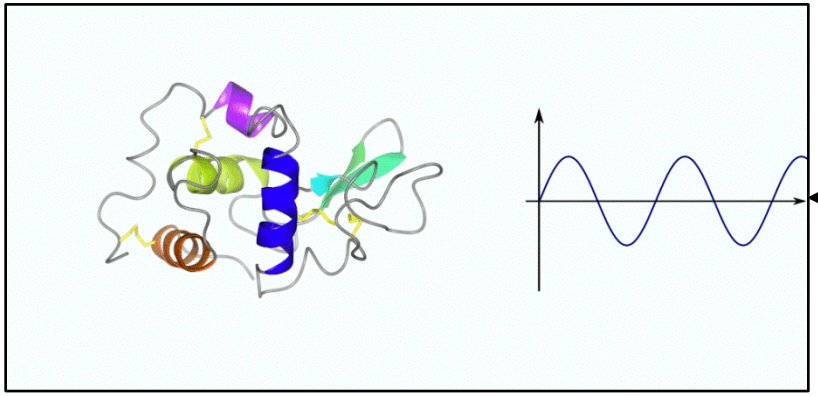


Rotation search



Translation search

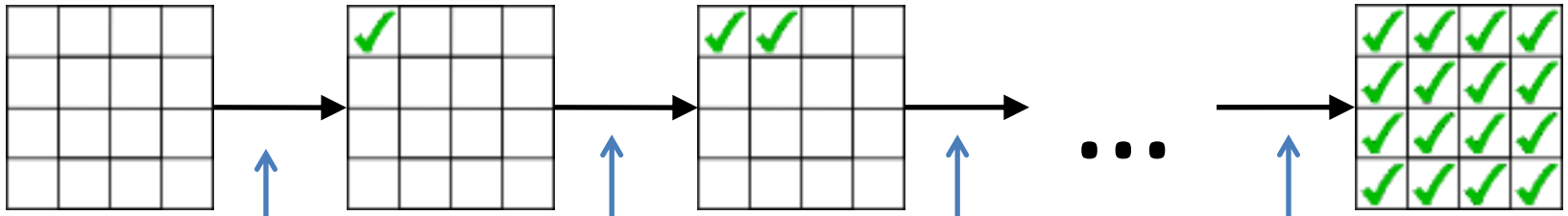
Packing test



Refinement

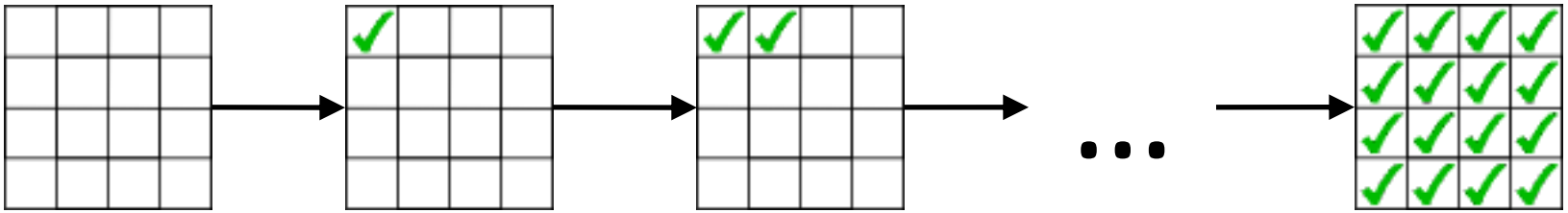
Brute search

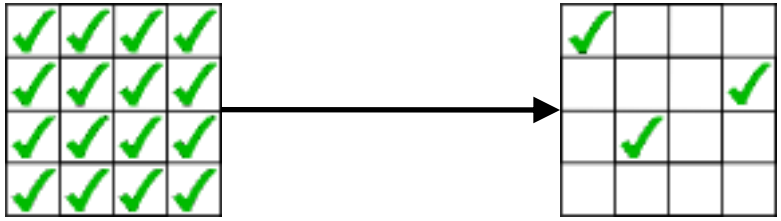
Calculate likelihood for each gridpoint



Brute search

Calculate likelihood for each gridpoint



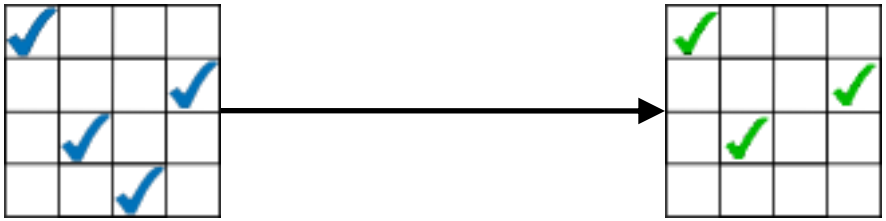


Find peaks

Fast search

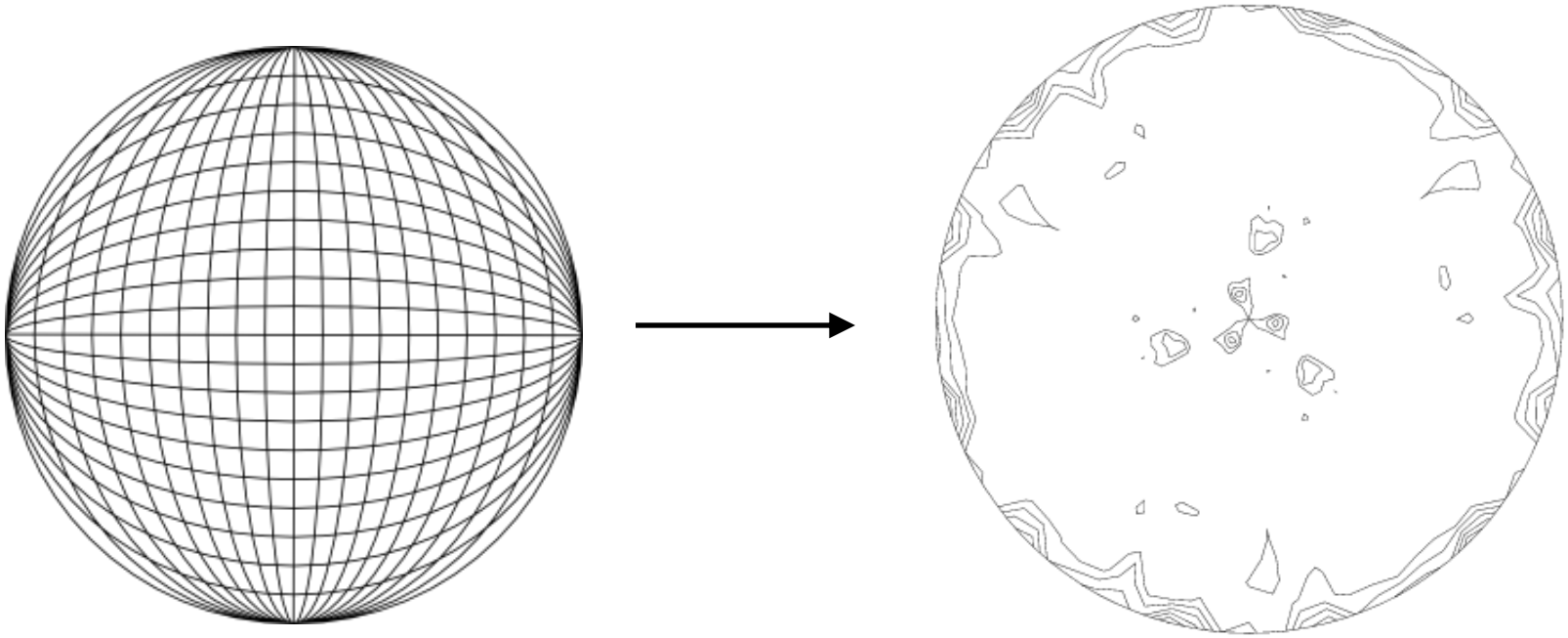
Calculate approximation by FFT

Peak search



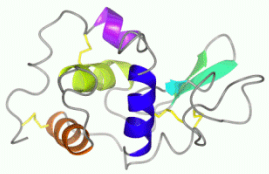
“Rescore” with likelihood

Rotation search

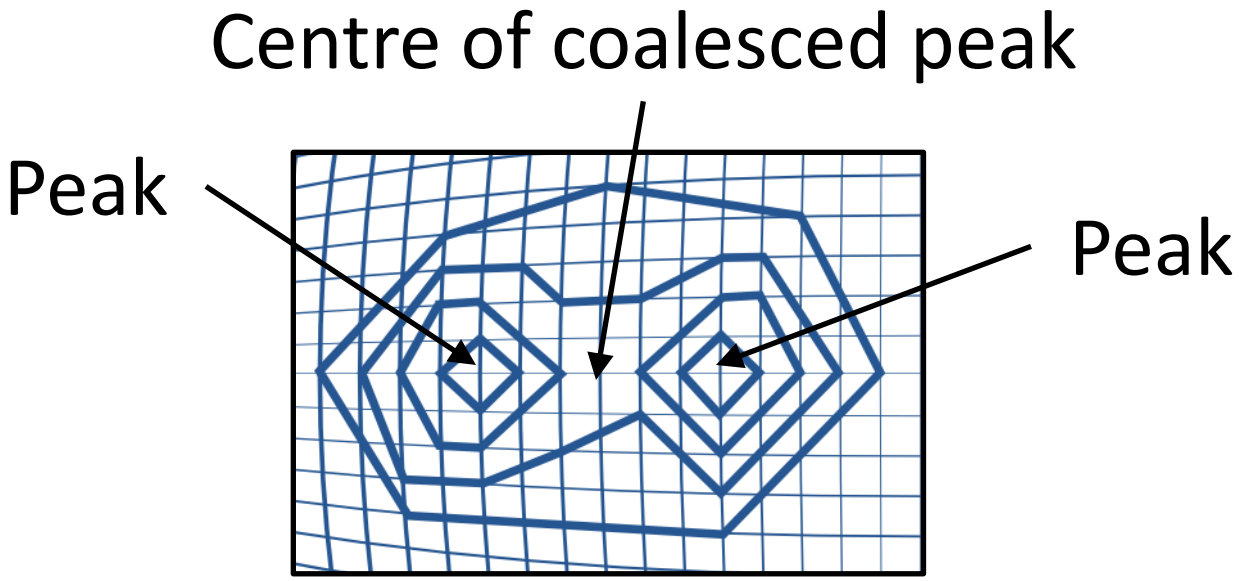


⇒ Typically small signal

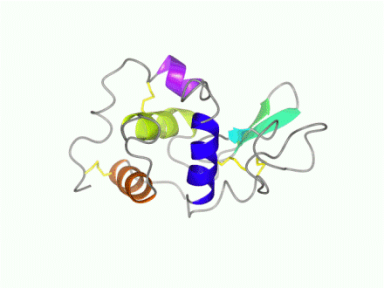
⇒ Dependent only on the point group



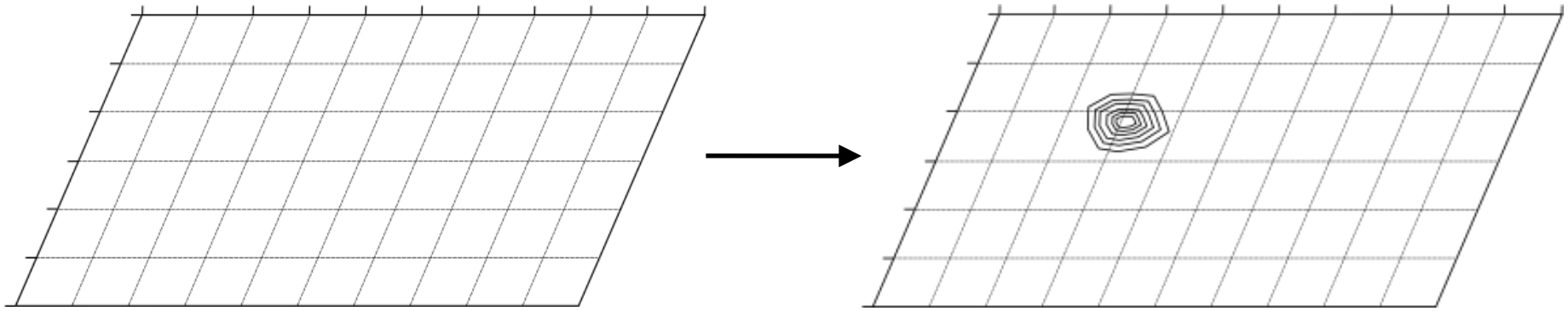
Rotation peaks



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



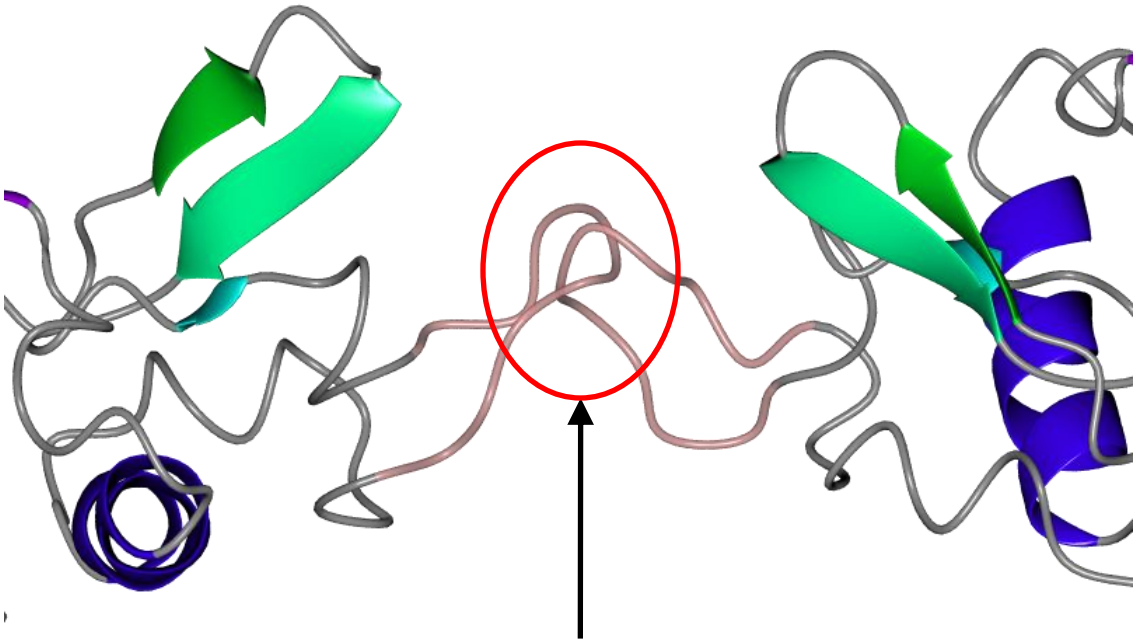
Translation search



⇒ Solutions are recognizable (high Z-score)

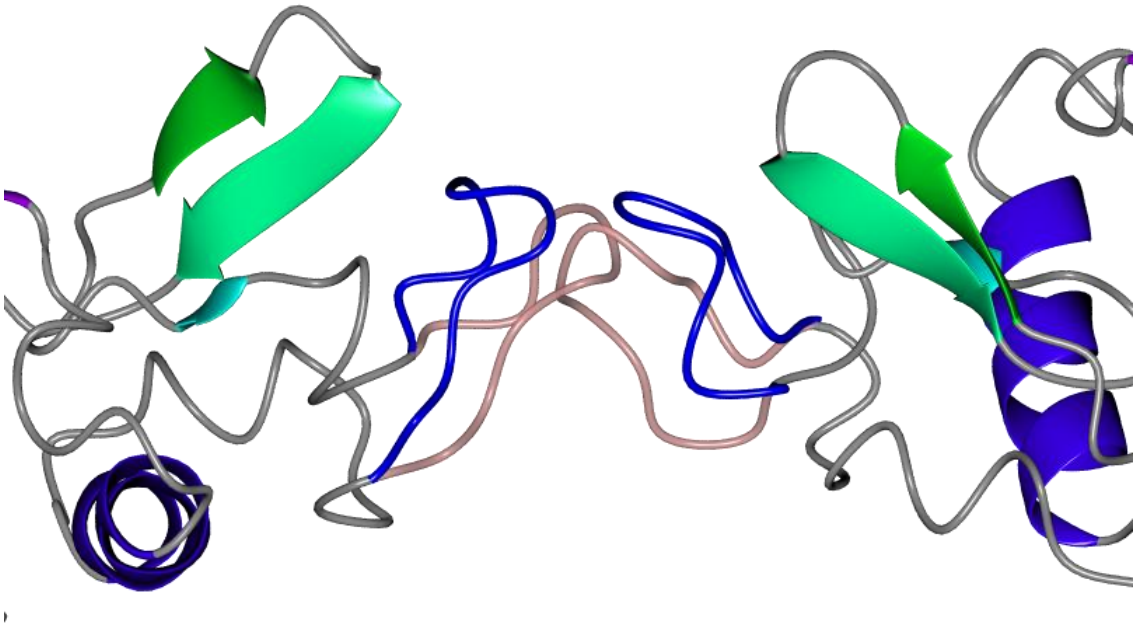
⇒ Very sensitive to correct orientation

Packing



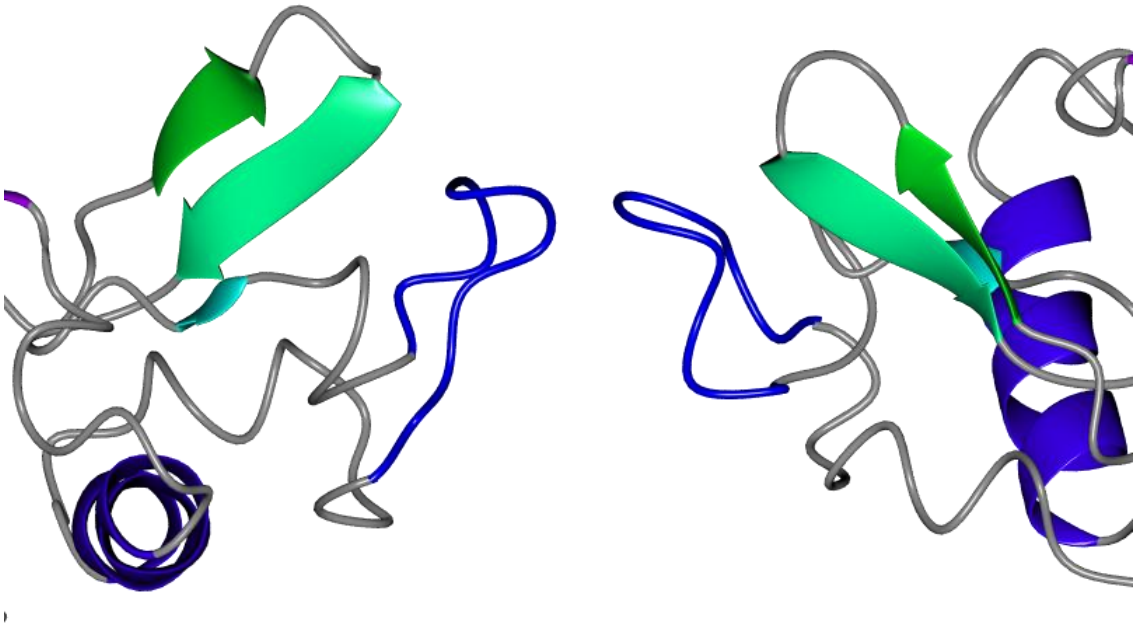
Clashes between models decrease credibility in solution

Packing



Surface loops may change conformation

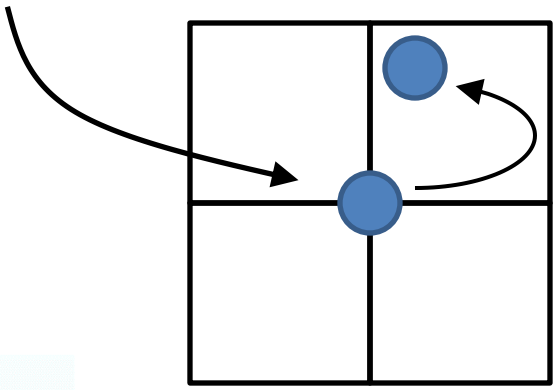
Packing



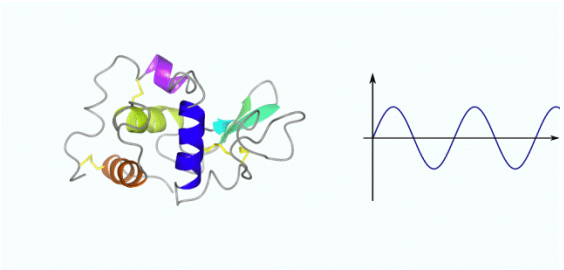
A small number of clashes
is acceptable

Refinement

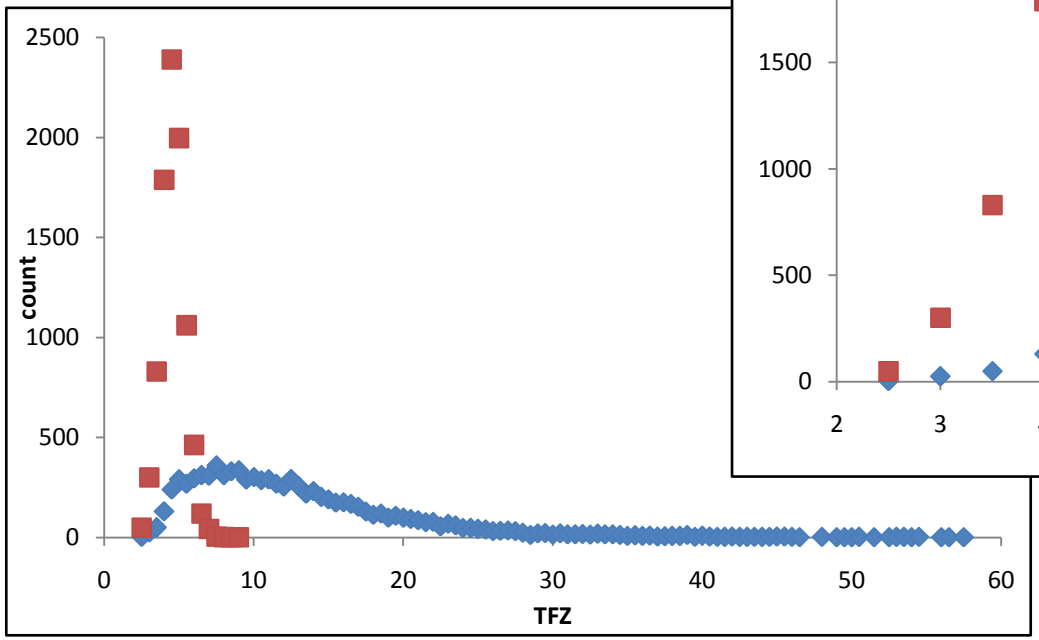
The rotation and translation functions were performed on a (not very fine) grid

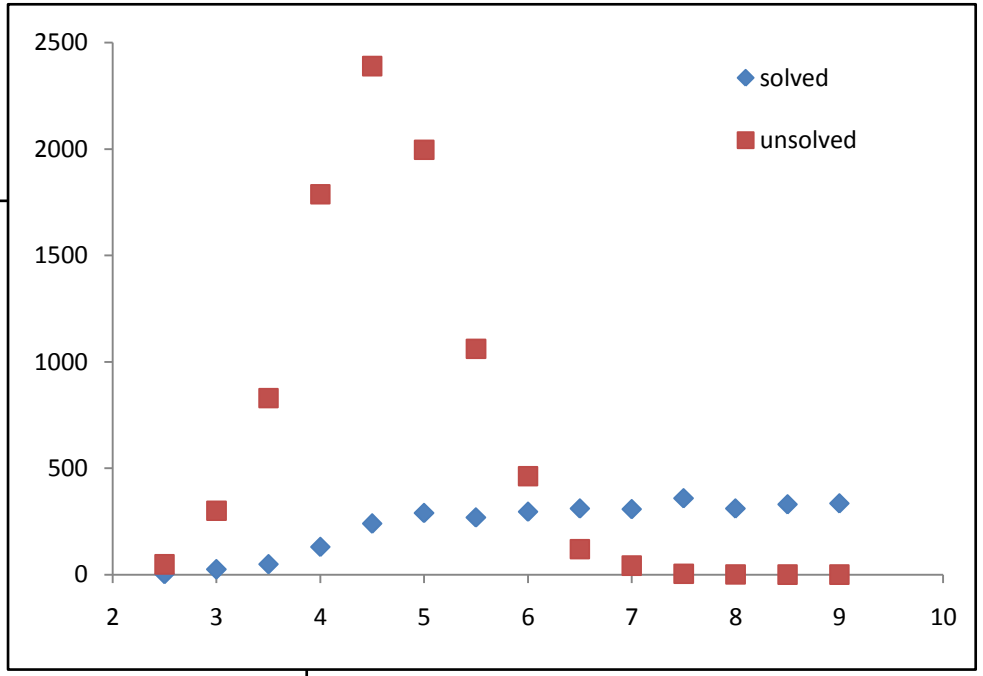


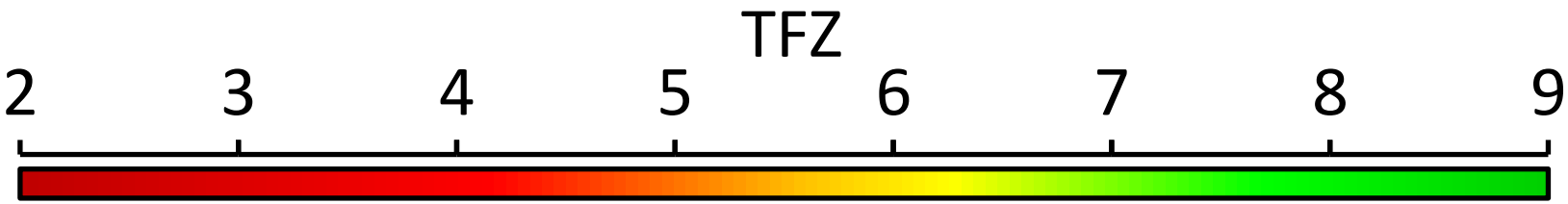
The solution can be improved if the grid is taken away and the rotational and translational parameters optimized



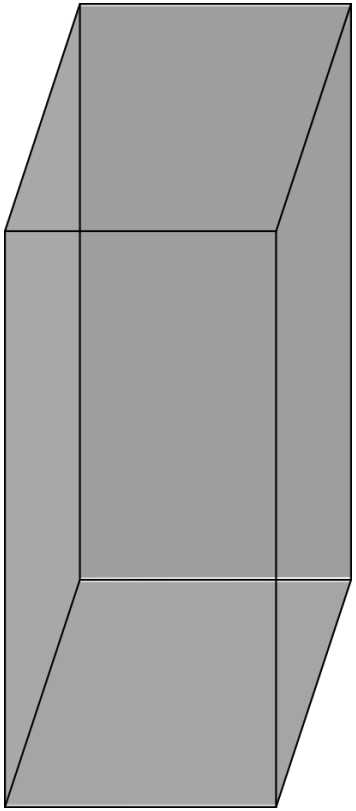
Identifying solutions

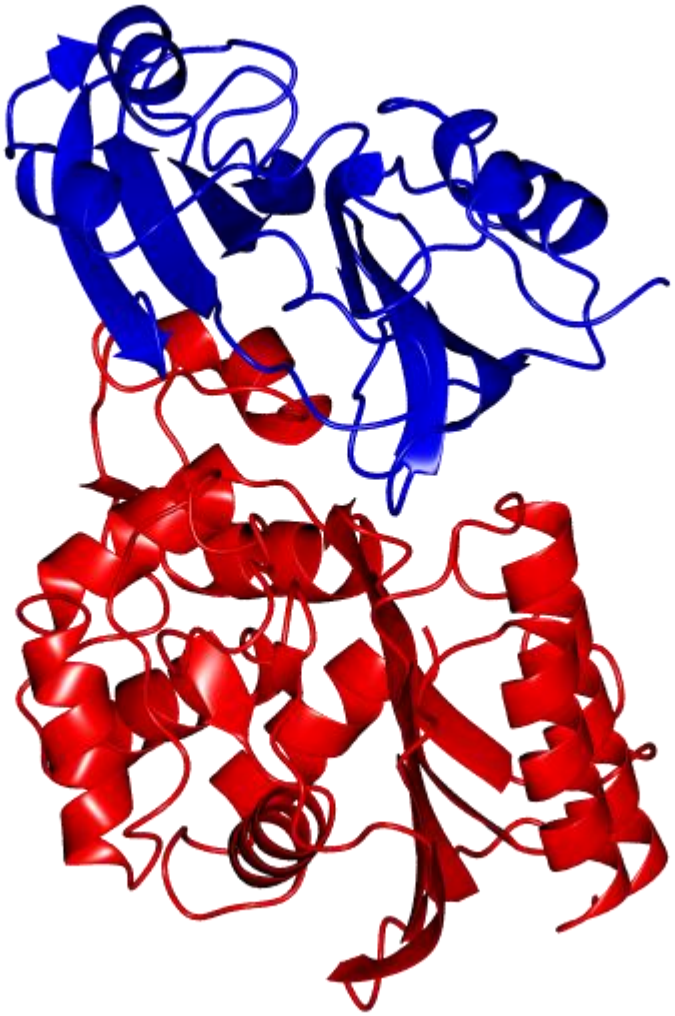






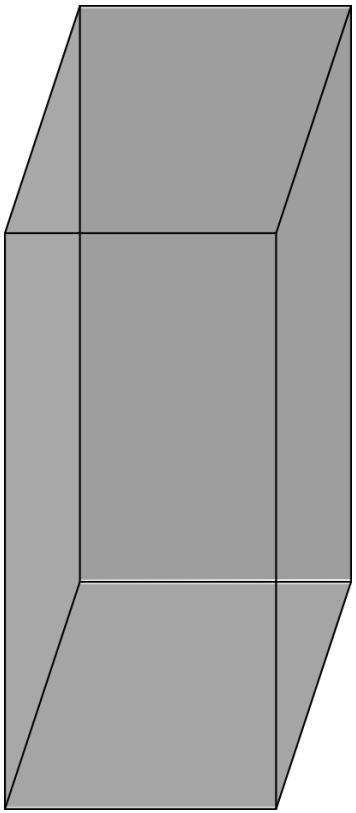
Partial structure

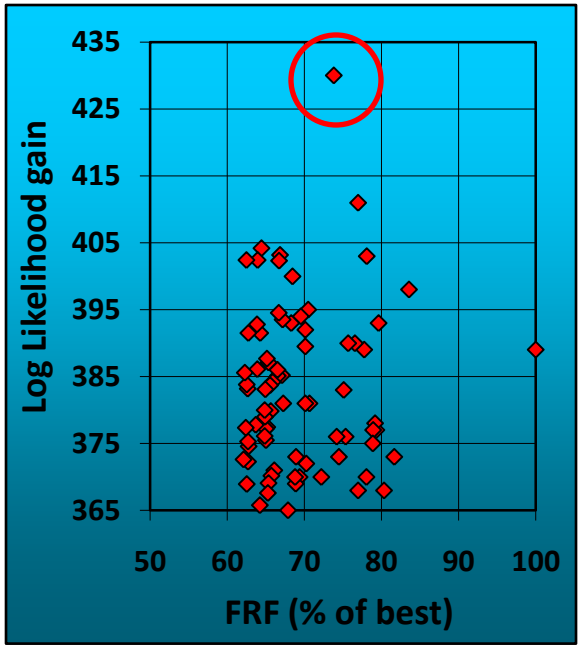




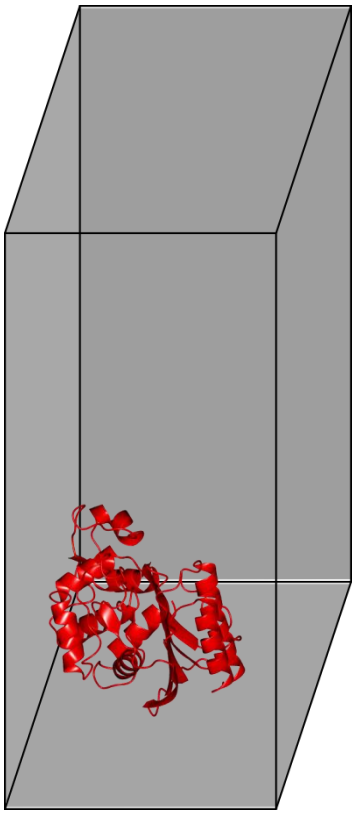
?

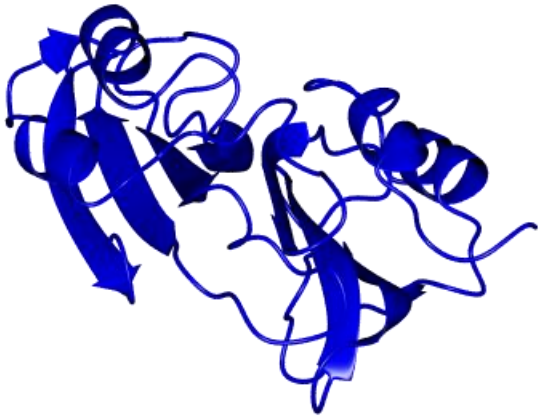
Partial structure

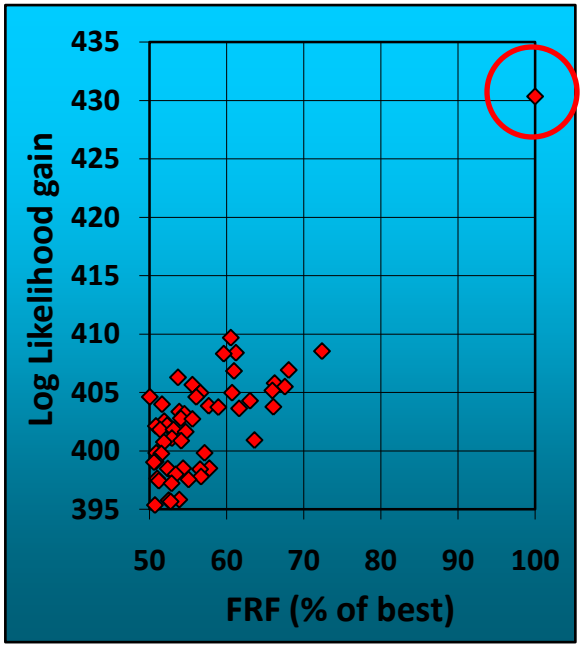




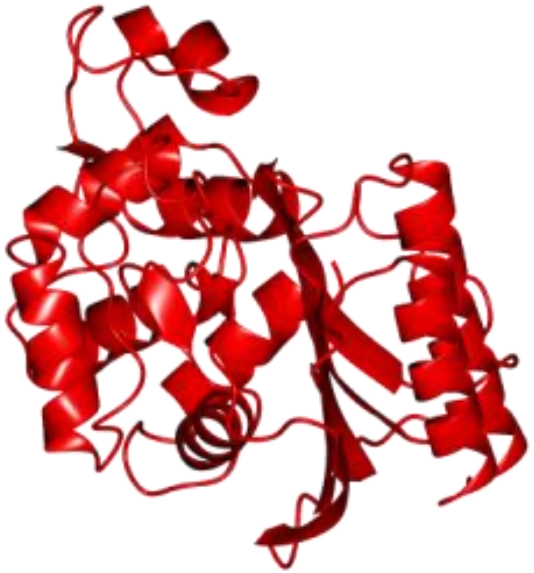
Partial structure

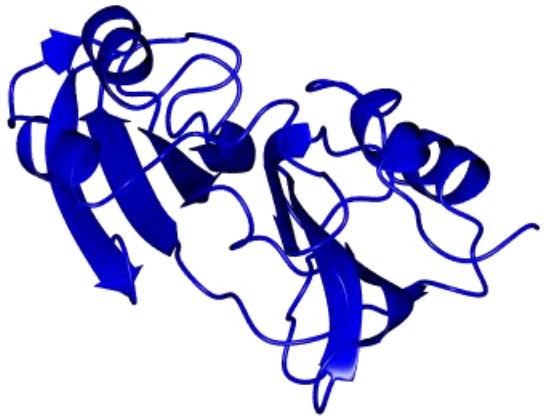






Model search order



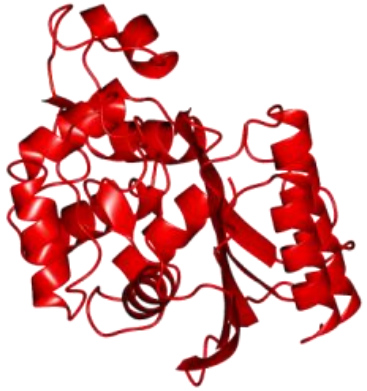


65% of structure
35% identical model

35% of structure
50% identical model

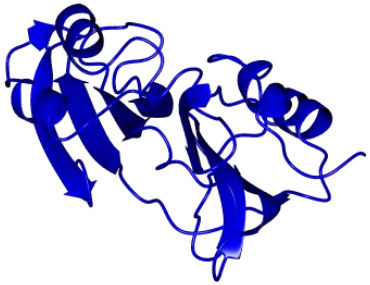
Which model is easier to find?

Model search order



65% of structure
35% identical model

Better model if data resolution is low

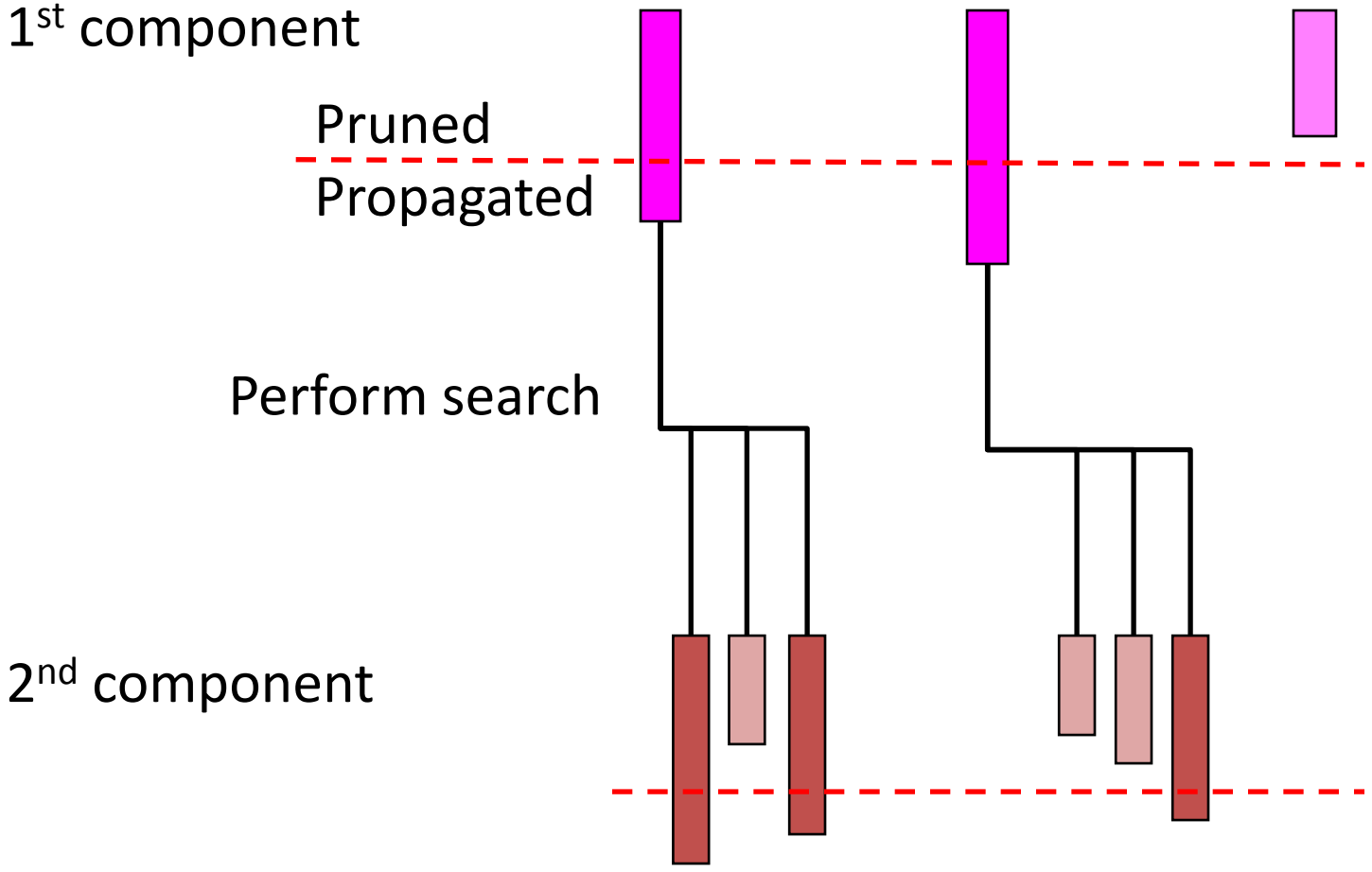


35% of structure
50% identical model

Better model if data resolution is high

Determined automatically!

Tree search with pruning



Running Phaser via CCP4i

mode

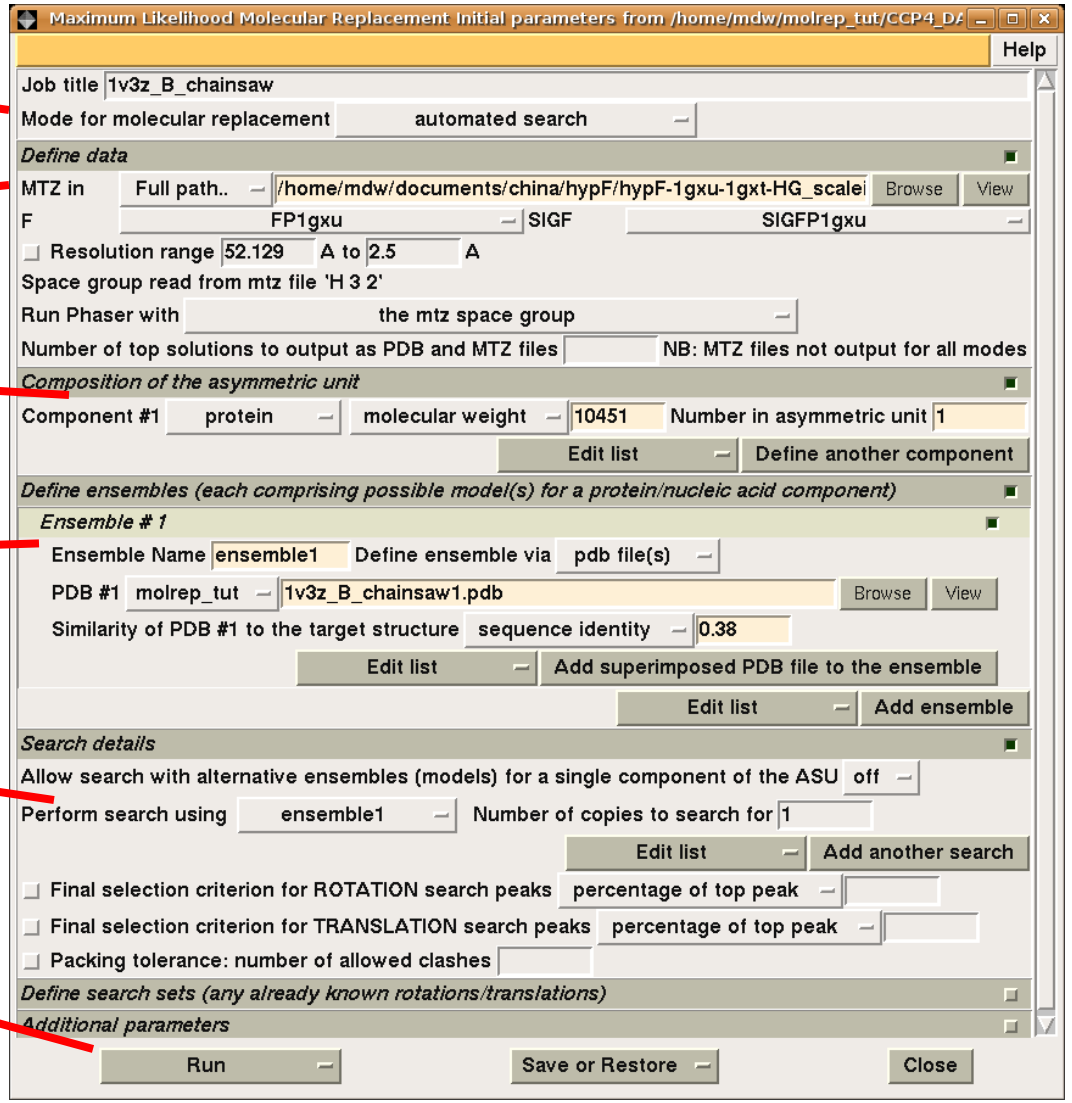
MTZ file

target details

search model

specify search

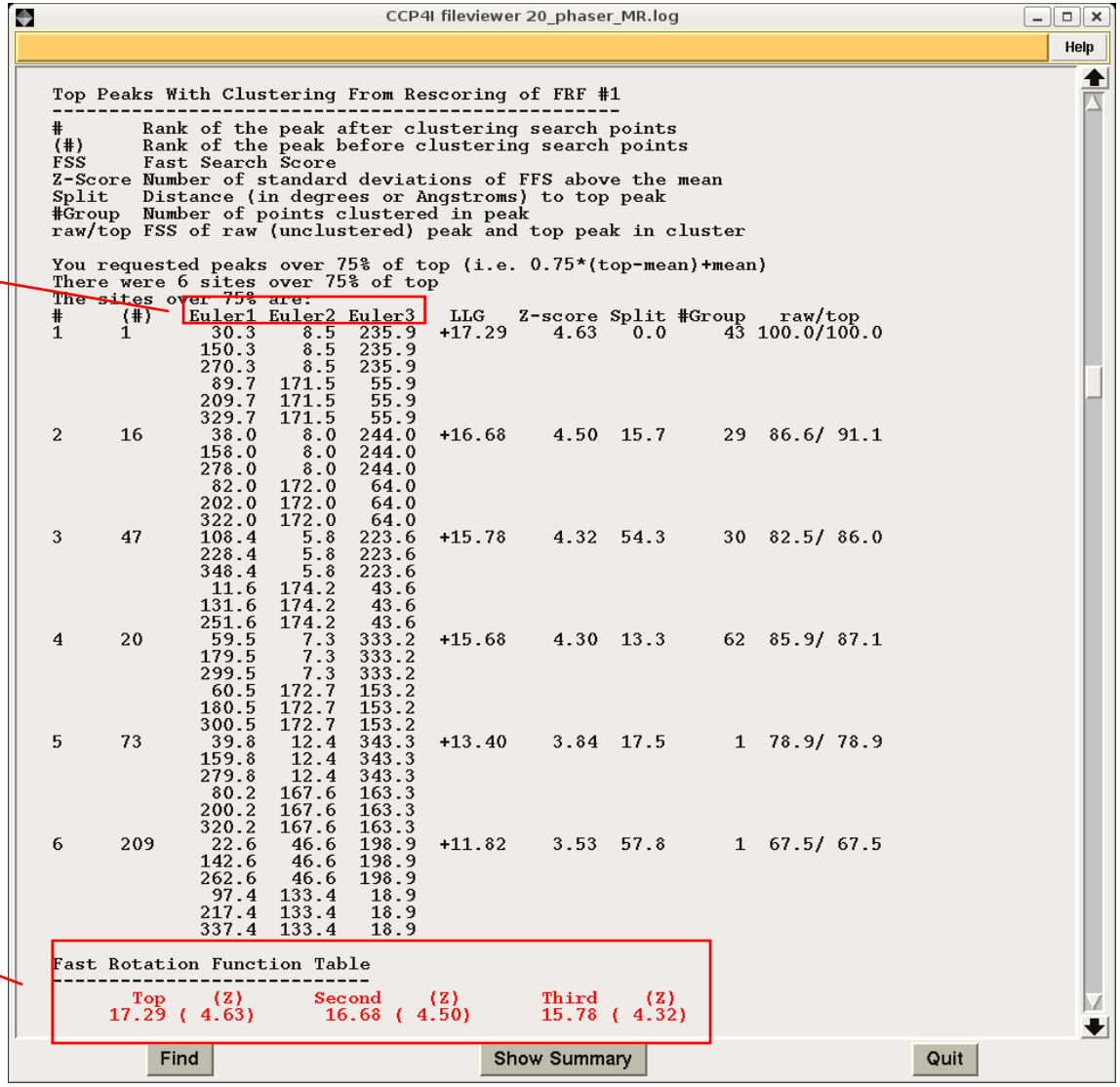
RUN IT!



Fast Rotation Function

Euler angles (CCP4)

Top LLG and Z-scores for FRF



Fast Translation Function

fractional translation

FRF solution number

Top LLG and Z-scores for FRF

CCP4I fileviewer 20_phaser_MR.log

Help

```
Top Peaks With Clustering From Rescoring of SET #1 TRIAL #1
------------------------------------------------------------
#       Rank of the peak after clustering search points
(#)     Rank of the peak before clustering search points
FSS     Fast Search Score
Z-Score Number of standard deviations of FFS above the mean
Split   Distance (in degrees or Angstroms) to top peak
#Group  Number of points clustered in peak
raw/top FSS of raw (unclustered) peak and top peak in cluster

You requested peaks over 75% of top (i.e. 0.75*(top-mean)+mean)
There were 57 sites over 75% of top
The sites over 75% are:
#     (#)    Frac X Frac Y Frac Z   LLG   Z-score Split #Group    raw/top
1     1       0.692  0.691  0.206  +25.09    4.62     0      2  56.60/ 56.60
2     14      0.509  0.989  0.037  +24.74    4.58    29      3  53.10/ 54.87
3     3       0.502  0.981  0.771  +23.07    4.37    12      2  55.89/ 56.07
4     9       0.695  0.695  0.258  +23.06    4.37     8      1  54.10/ 54.10
5     19      0.475  0.844  0.775  +22.62    4.31    16      2  52.43/ 52.43
6     6       0.571  0.450  0.132  +21.80    4.21    17      3  55.18/ 55.18
```

Find | Show Summary | Quit

CCP4I fileviewer 20_phaser_MR.log

Help

```
-----------------
TABLES OF RESULTS
-----------------

 Fast Translation Function Table: Space Group R 3 2
 --------------------------------------------------
#SET #TRIAL    Top     (Z)     Second     (Z)     Third     (Z)    Ensemble
   1      1  25.09 ( 4.62)   24.74 ( 4.58)  23.07 ( 4.37)  ensemble1
   1      2  45.14 ( 7.13)       -     -        -     -    ensemble1
   1      3  23.91 ( 5.00)   21.21 ( 4.63)  20.14 ( 4.49)  ensemble1
   1      4  23.41 ( 4.53)   23.25 ( 4.51)  19.77 ( 4.08)  ensemble1
   1      5  21.86 ( 4.75)   21.46 ( 4.70)  20.37 ( 4.56)  ensemble1
   1      6  21.01 ( 4.78)   19.57 ( 4.61)  19.52 ( 4.60)  ensemble1
 ---- ------
```

Find | Show Summary | Quit

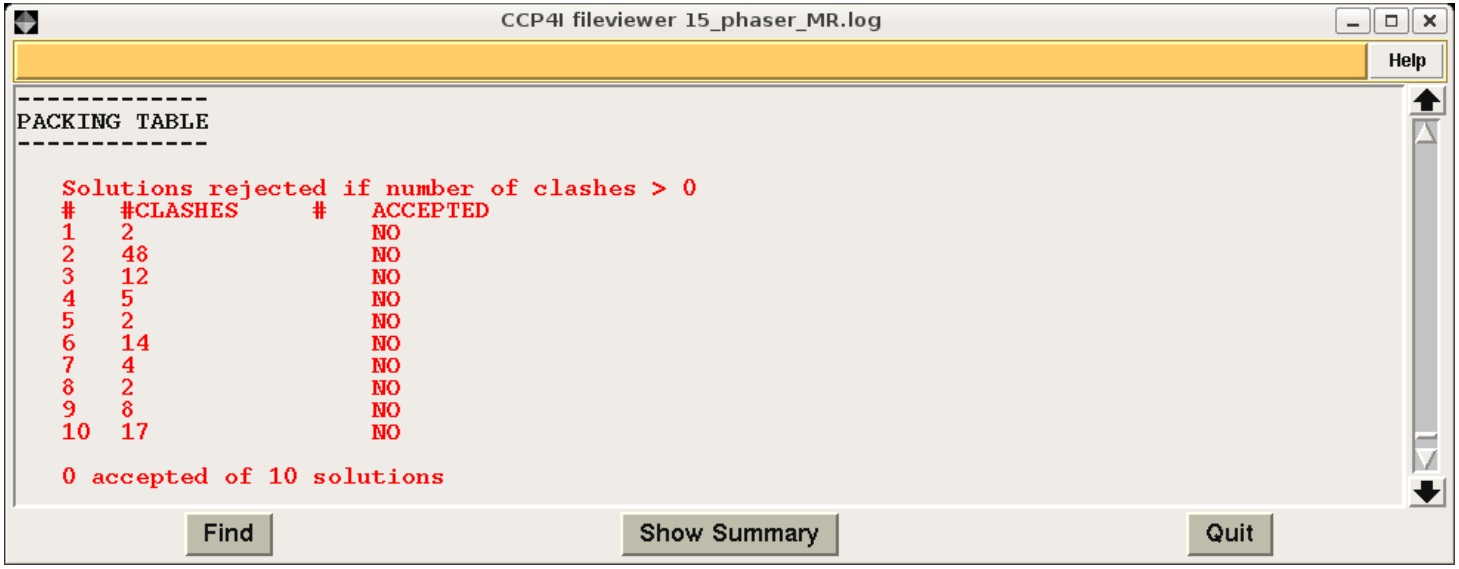
Packing

Phaser does packing check after FTF

Clashes = Cα 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

Z-score	**Have I solved it?**
less than 5	no
5 - 6	unlikely
6 - 7	possibly
7 - 8	probably
more than 8	definitely

RFZ = RF Z-score

TFZ = TF Z-score

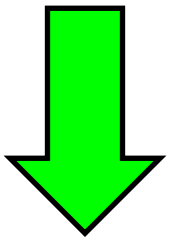
Models for molecular replacement

- ✓ *Ab initio* models (e.g. Rosetta)
- ✓ Structures of homologous proteins
  - ✗ unedited
  - ✓ homology models
  - ✓ “mixed” models

Model improvement

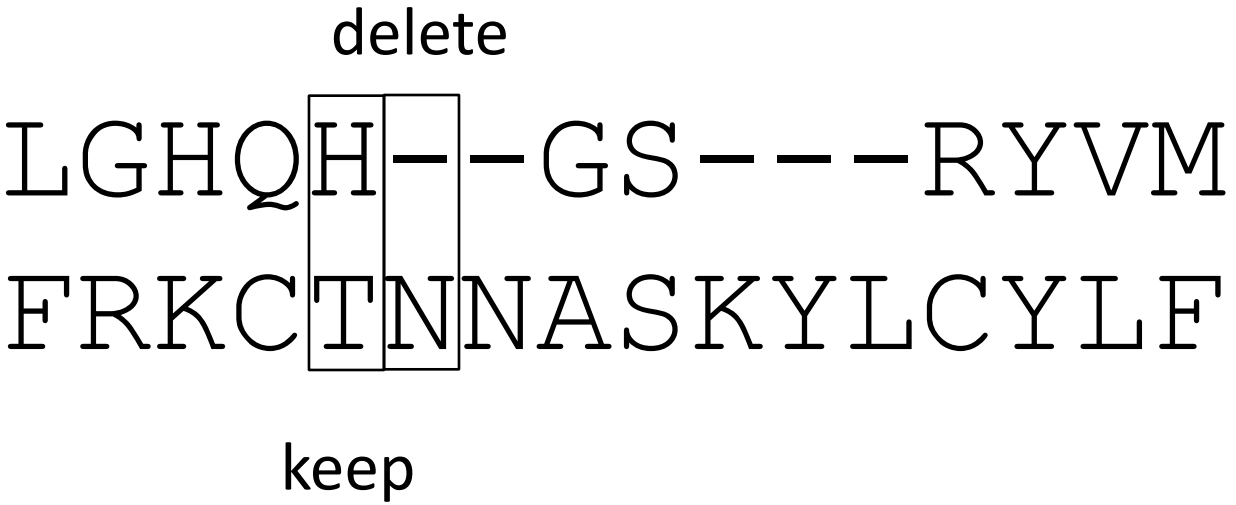
✓ Structural information in MR models

✓ Extra information in sequence alignment



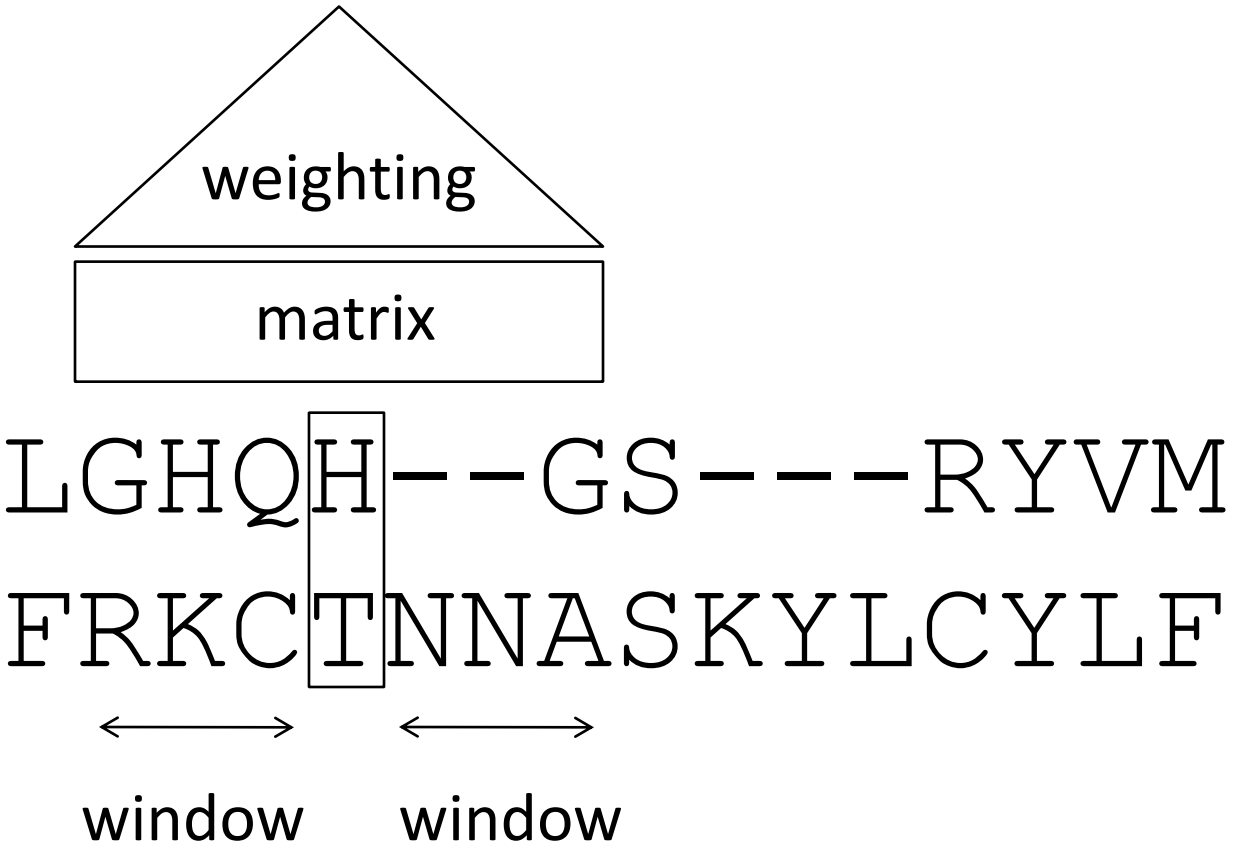
Improve model by removing residues that are not present or significantly different in the target

Binary concept



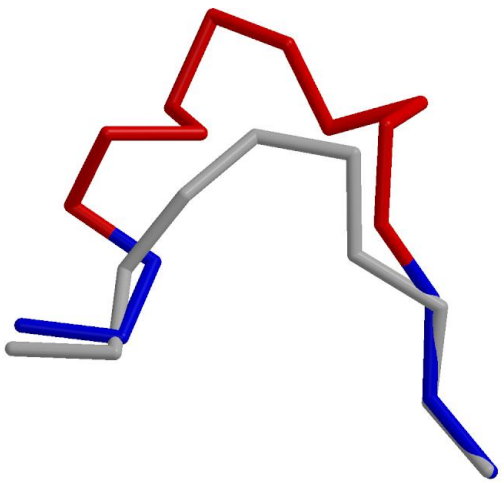
CHAINSAW (CCP4)

Sequence similarity

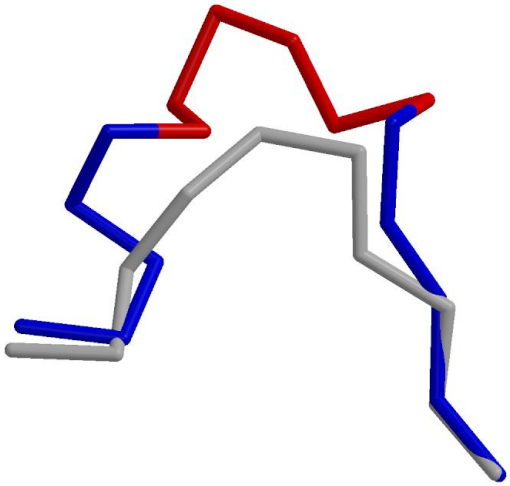


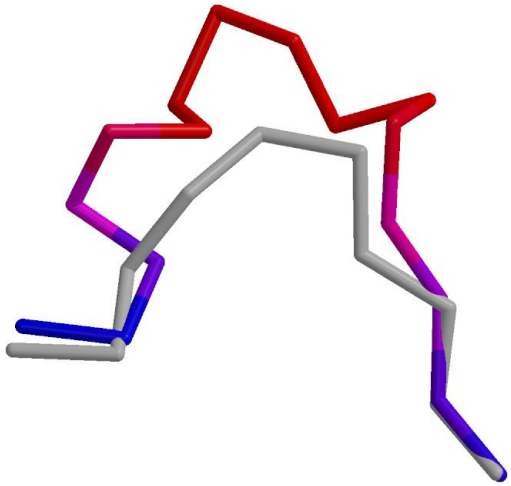
Sculptor (CCP4++)

Correcting for insertions



CHAINSAW with
structure-based alignment





CHAINSAW

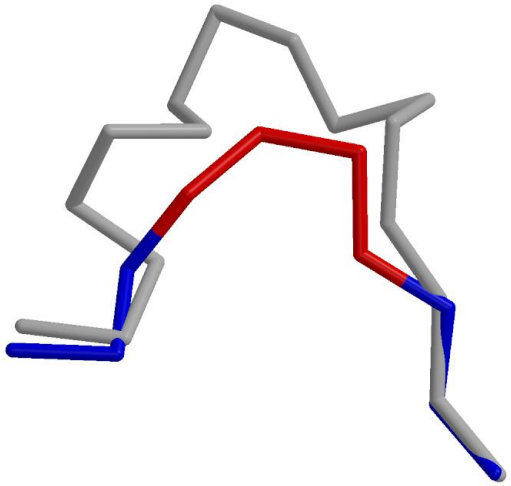
Sculptor

HCYKS------GIQVR

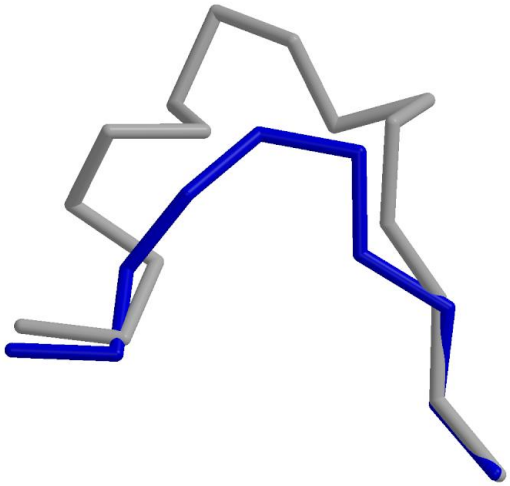
HCVNSYQSNLDAIKIR

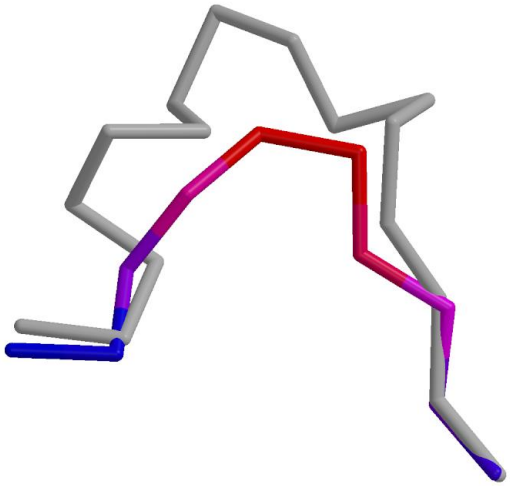
profile-profile sequence-based alignment
(FFAS)

Correcting for deletions



CHAINSAW with
structure-based alignment





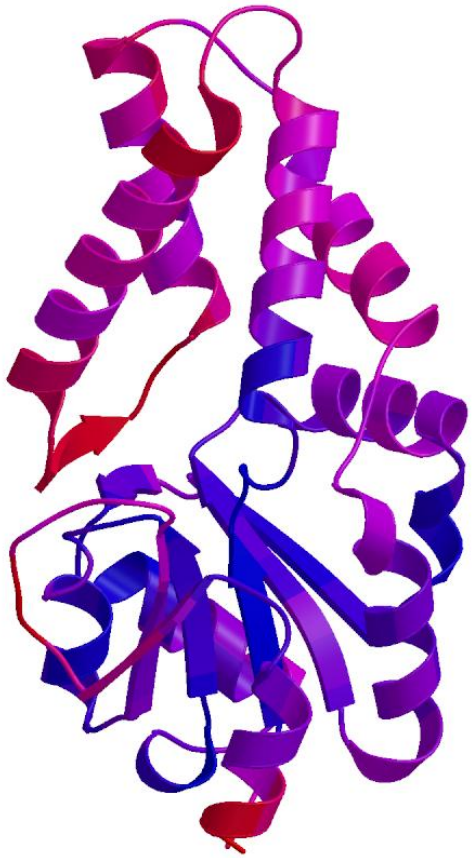
CHAINSAW

Sculptor

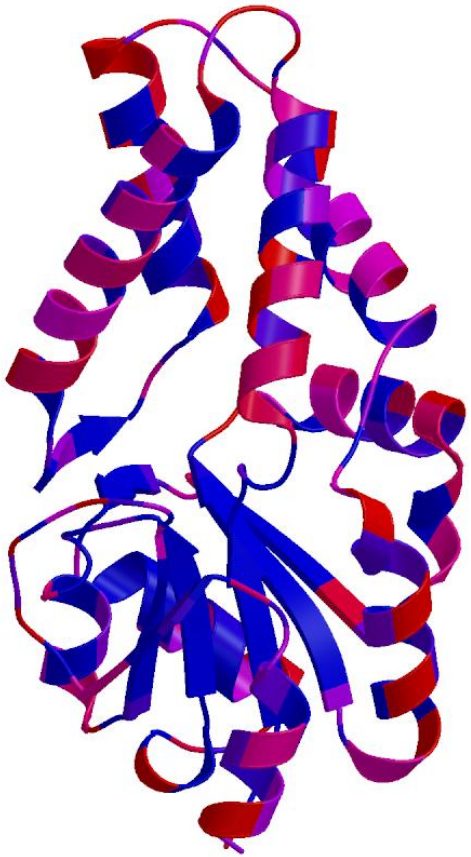
```
HCVNSYQSNLDAIKIR
HCYKS------GIQVR
```

profile-profile sequence-based alignment
(FFAS)

Downweighting



Sequence similarity
(residue level)



Accessible surface area
(atom level)

Multi-model strategy

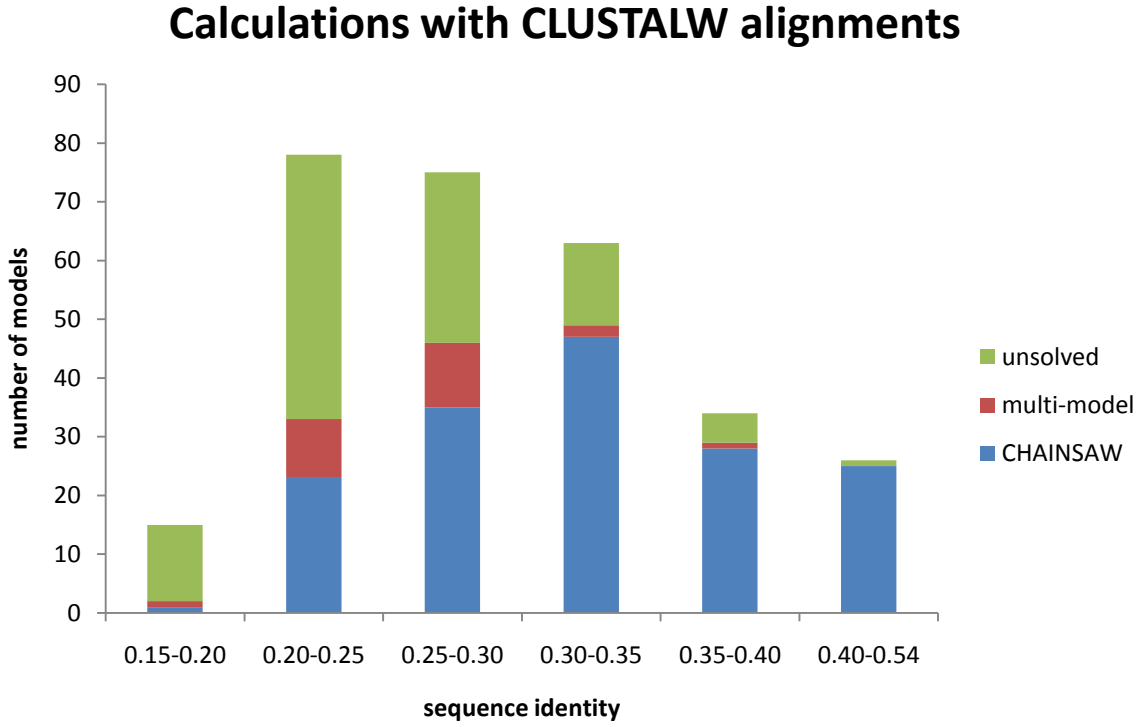
Model
CHAINSAW

Model
Sculptor

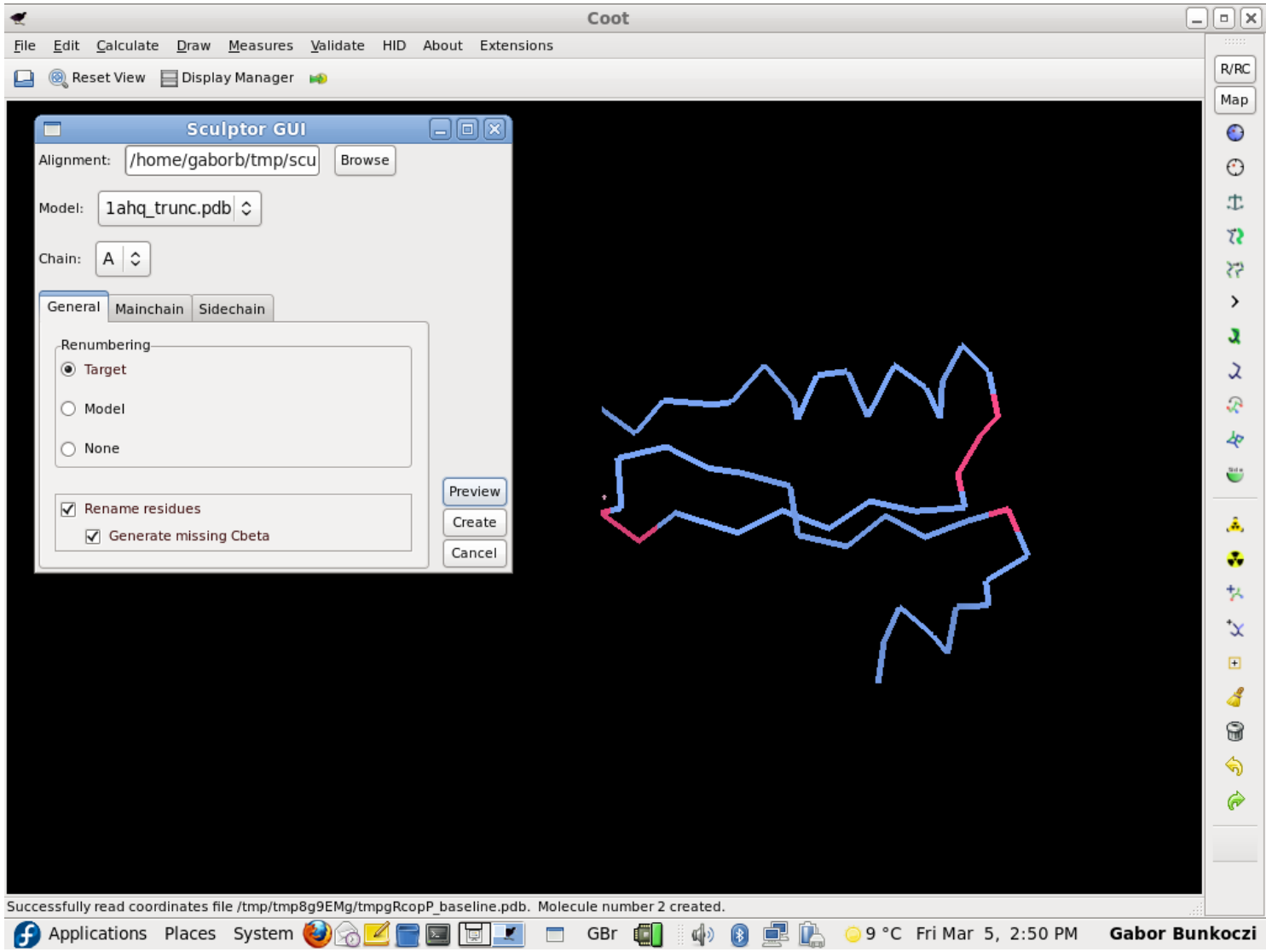
Model
Sculptor
+ASA B-factors

Model
Sculptor
+sequence similarity B-factors

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



Interactive model editing



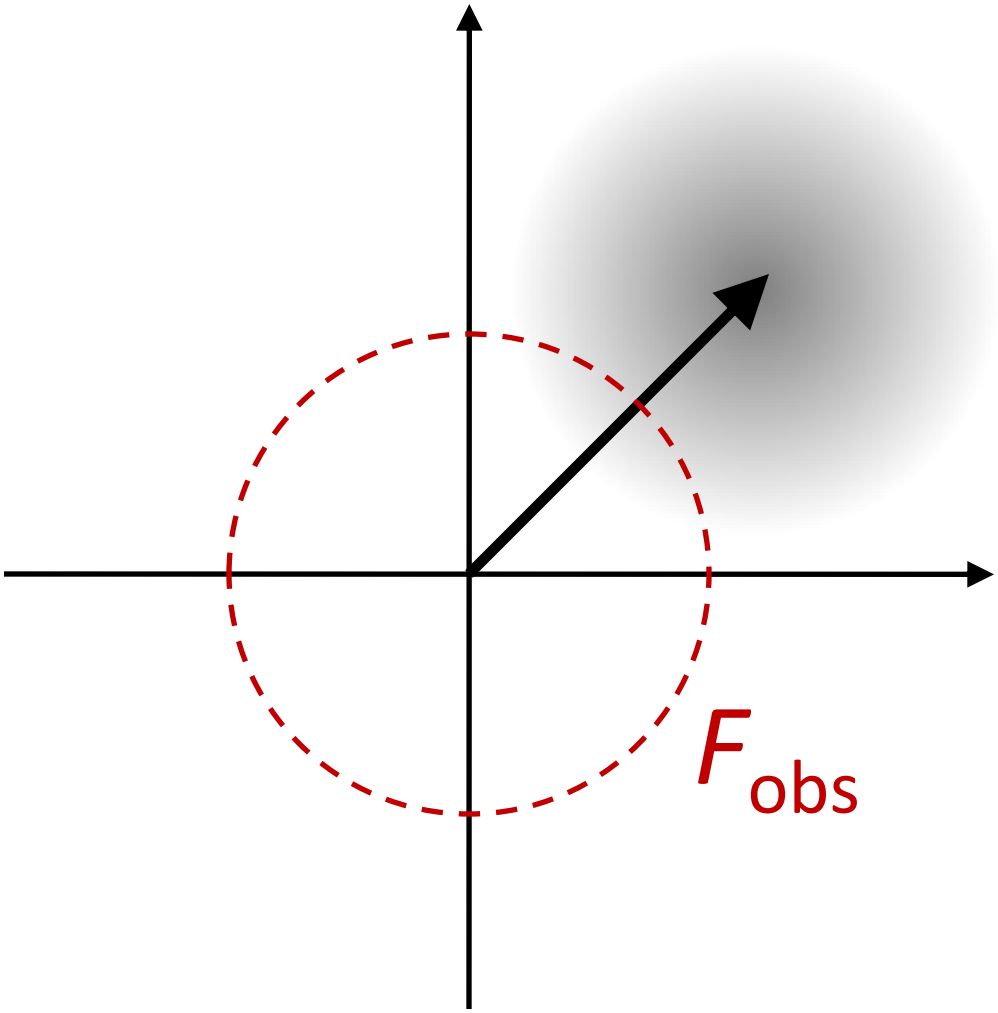
Acknowledgements

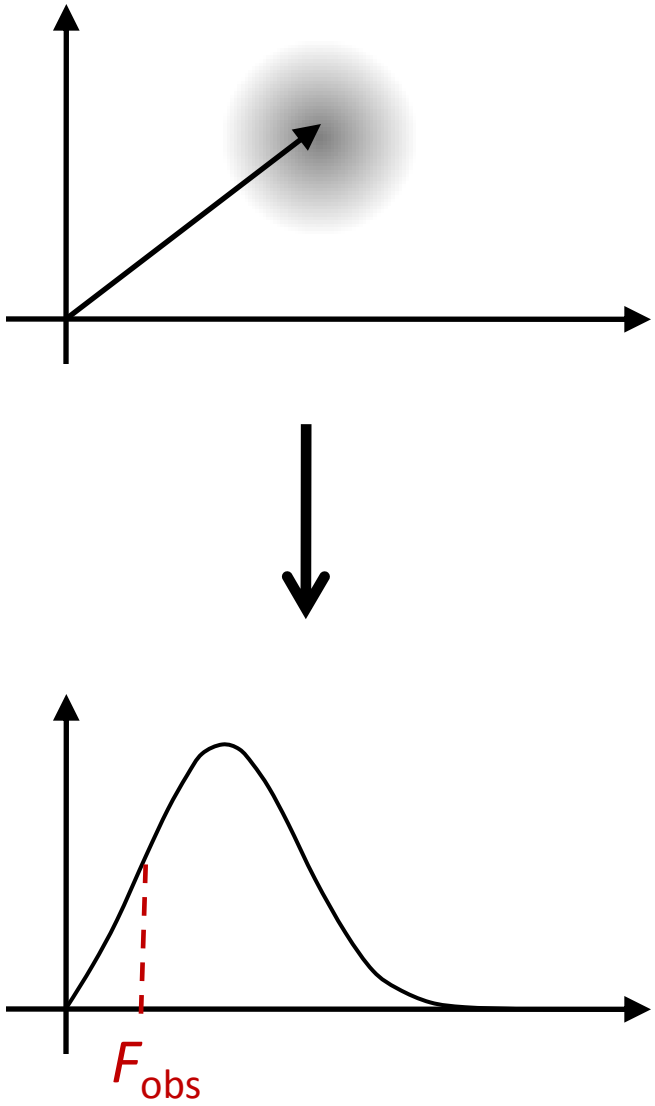
Python-based Hierarchical Environment for Integrated Xtallography

UNIVERSITY OF CAMBRIDGE

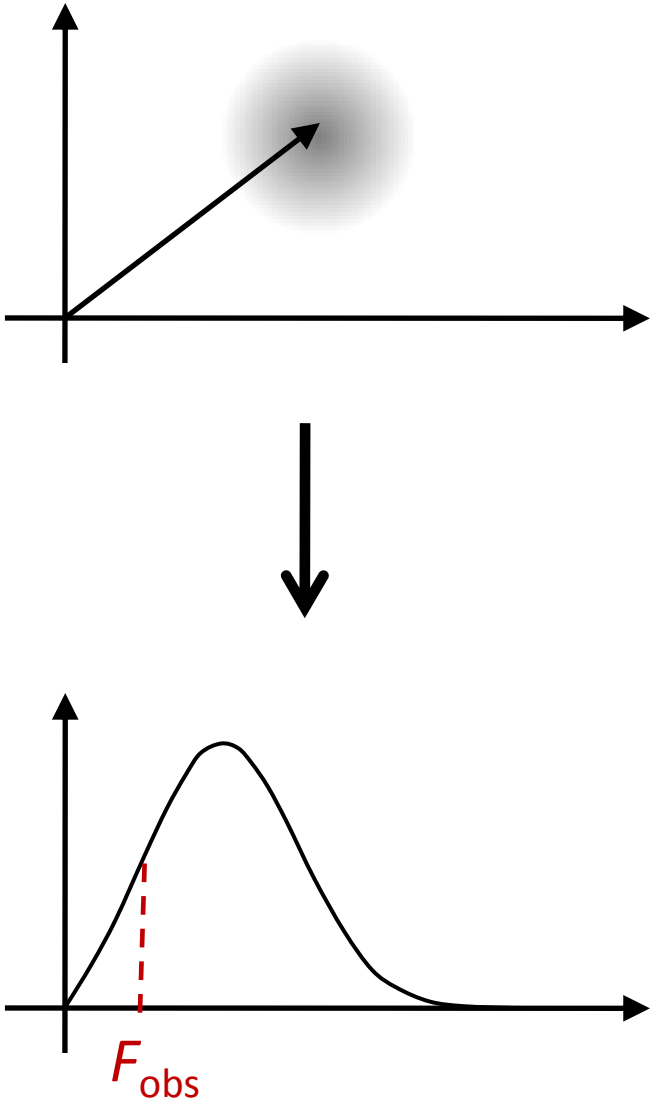
University of Cambridge
Gabor Bunkoczi, Randy Read, Airlie McCoy, Robert Oeffner

Reflection amplitude likelihood





Rotation and translation functions



Identical mathematical forms!

Translation function:
Reflection amplitude likelihood

Rotation function:
Reflection amplitude likelihood with σ_Δ is increased to account for unknown translation