



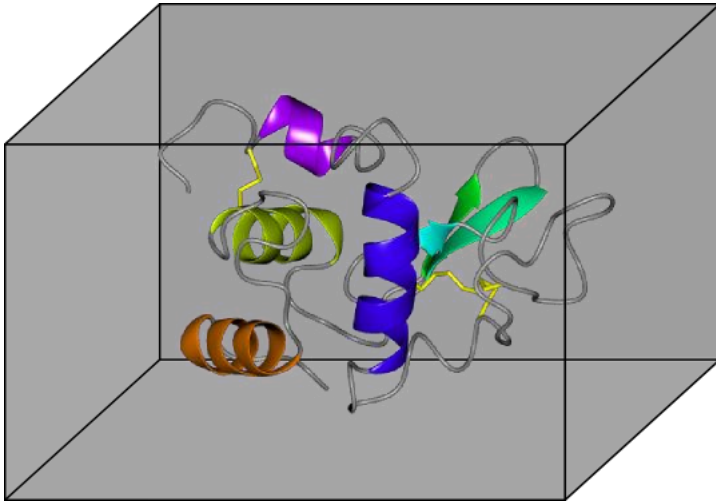
Molecular Replacement

**Ronan Keegan
(original by Gabor Bunkoczi)**

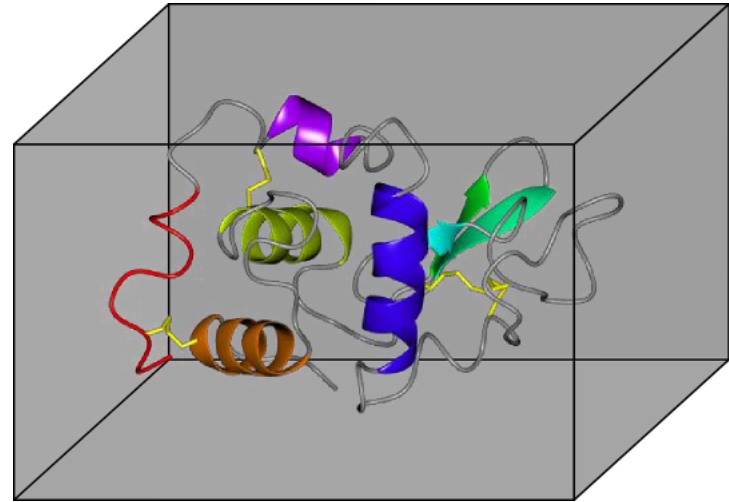
***OIST/CCP4 Workshop
6 December 2011***



Idea



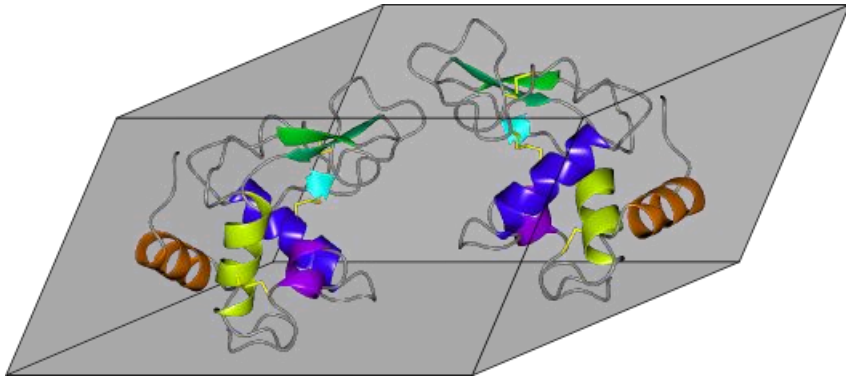
h	k	l	F	ϕ
0	0	1	12.6	123
0	0	2	2.1	12
0	0	3	69.9	287



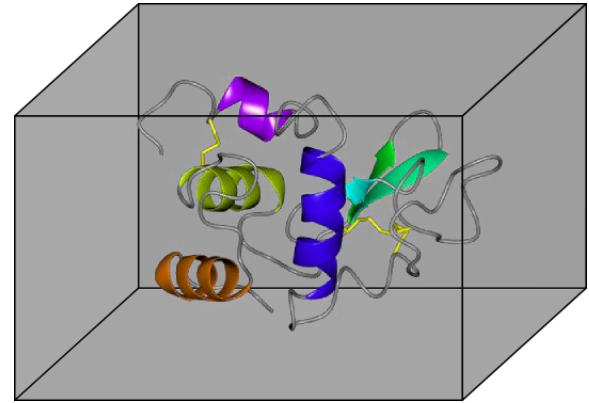
h	k	l	F	ϕ
0	0	1	10.4	113
0	0	2	3.5	18
0	0	3	57.2	265



Problem



Previous crystal form



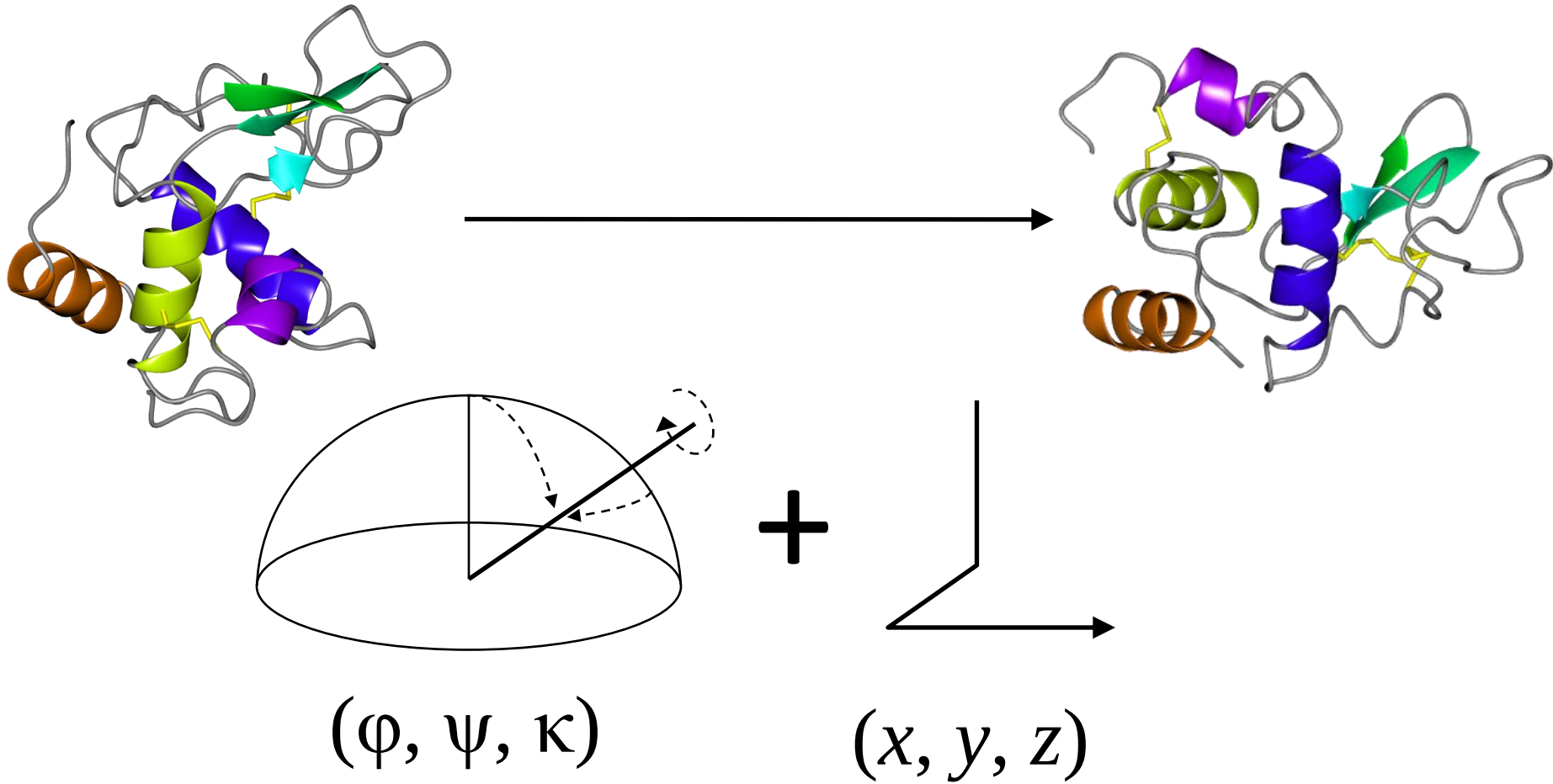
Current crystal form



$$\mathcal{R}(\phi, \psi, \kappa, x, y, z)$$



Separability

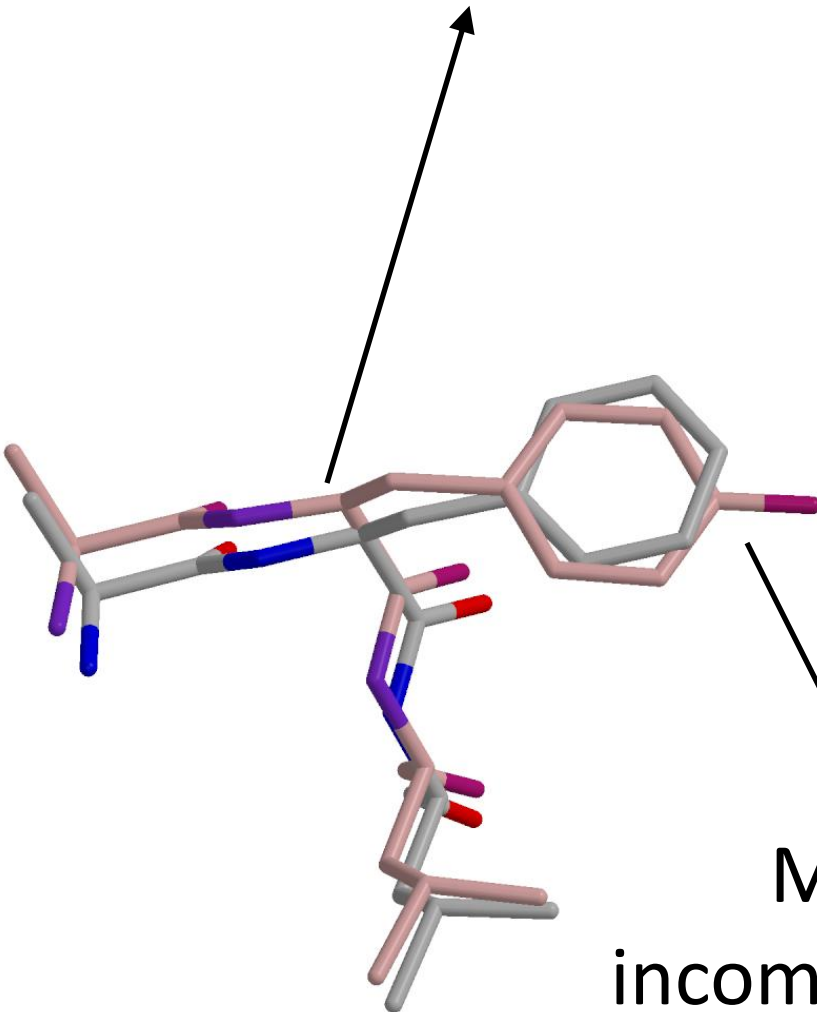
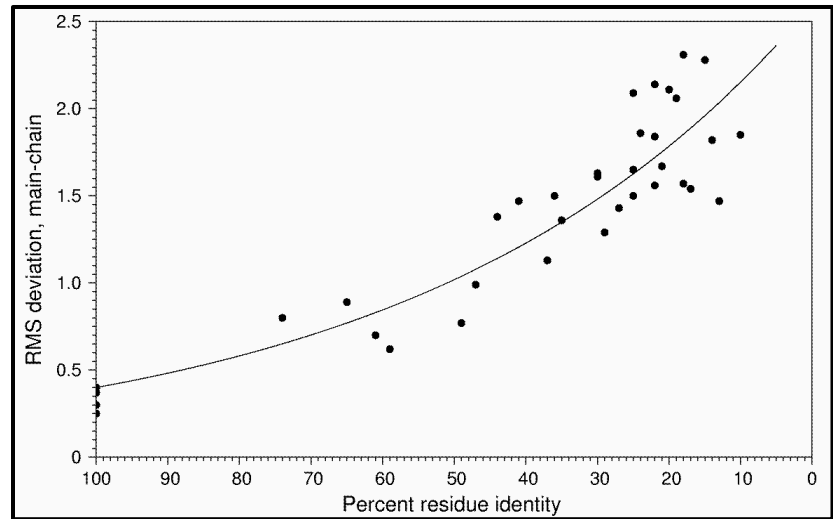




Model error

Position errors

Calculate from
sequence identity
(Chothia & Lesk, 1986)

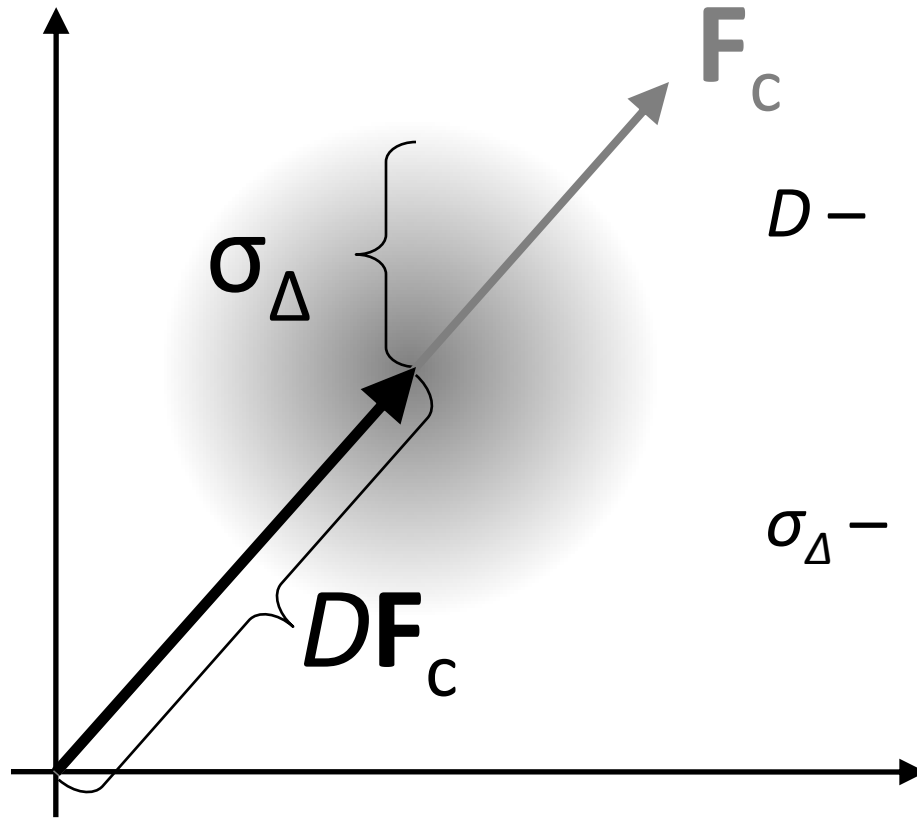


Model
incompleteness

Calculate from unit
cell composition



Likelihood function



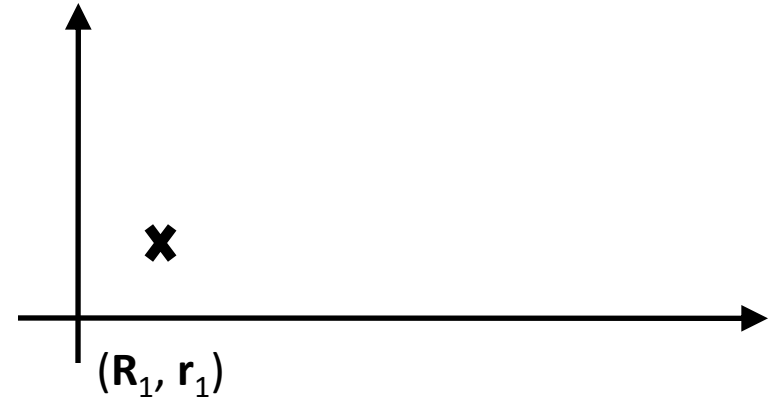
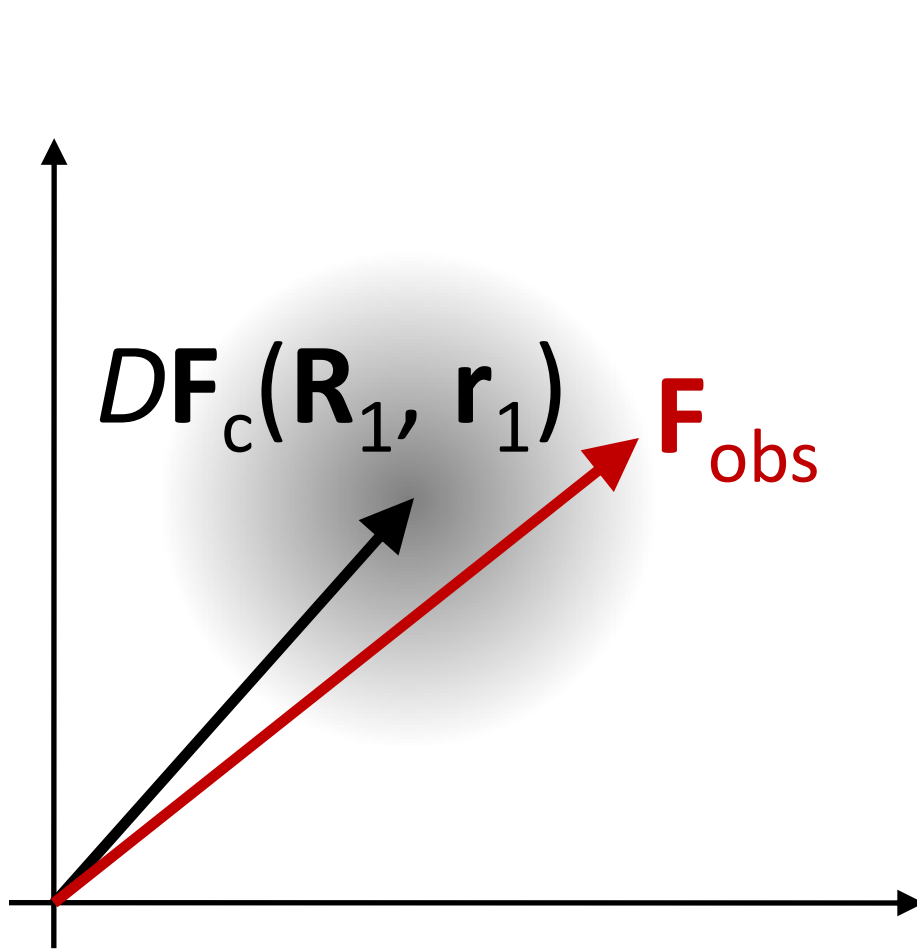
F_c – model structure factor
Dependent on orientation
and position of model

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

σ_Δ – uncertainty
Dependent on model
errors and incompleteness

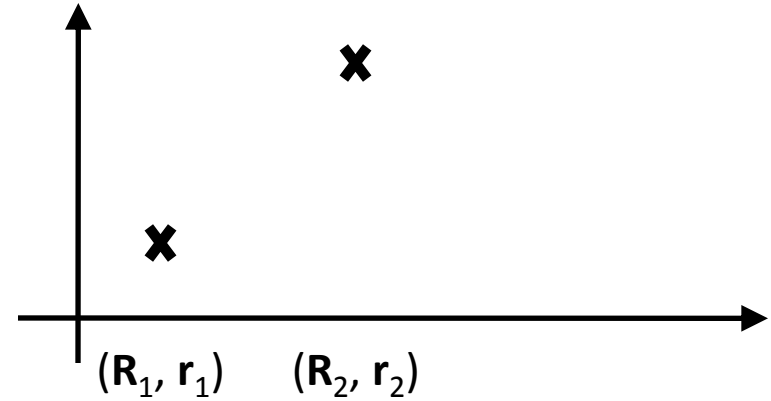
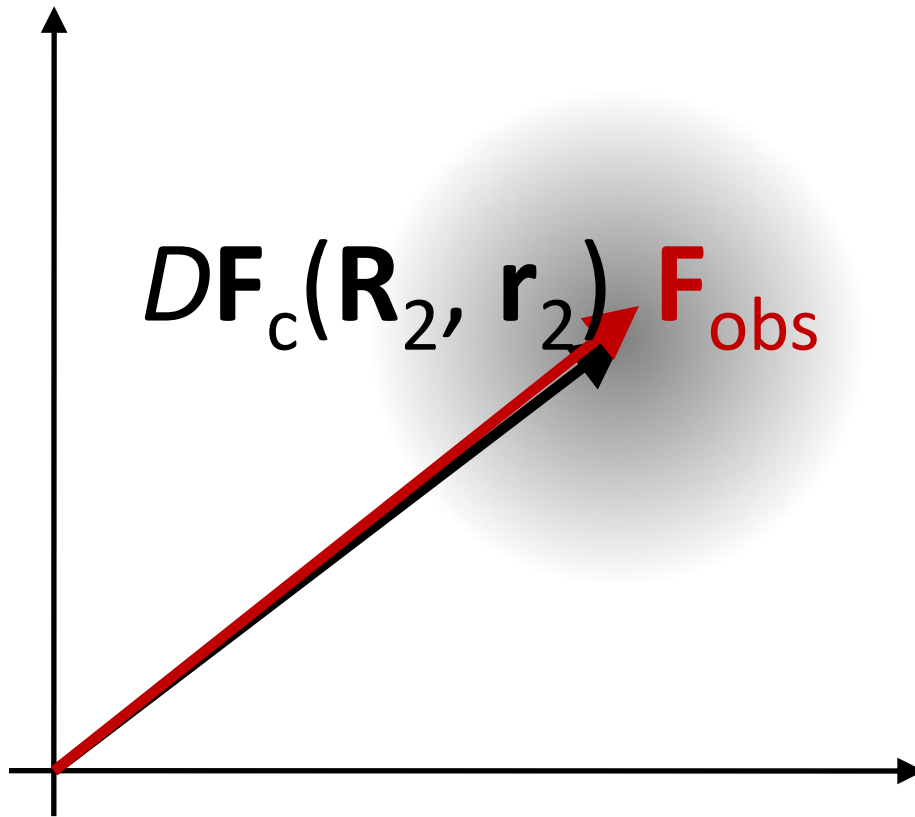


Reflection likelihood



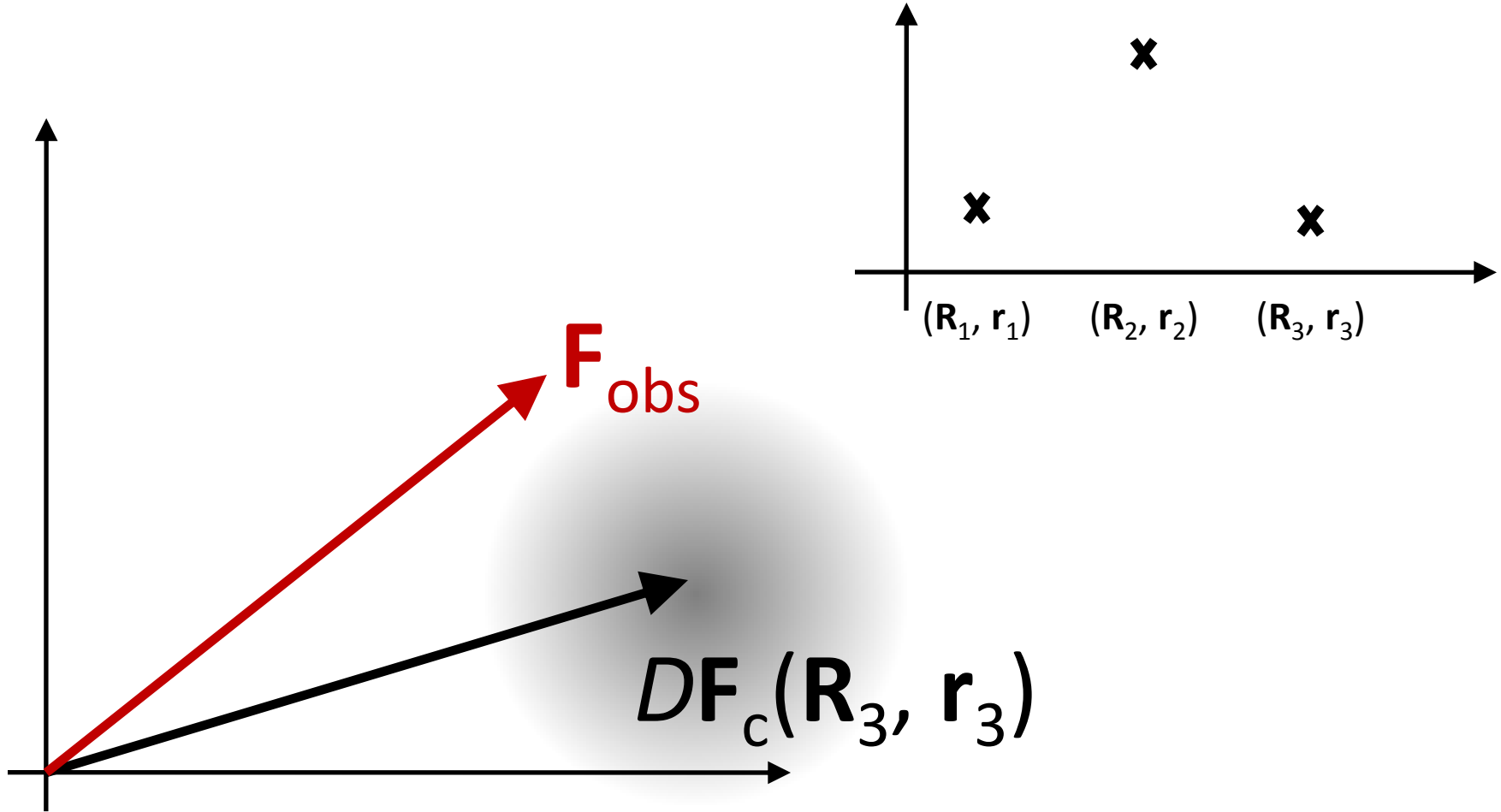


Reflection likelihood





Reflection likelihood





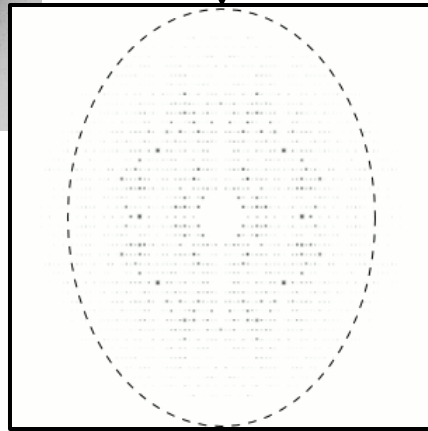
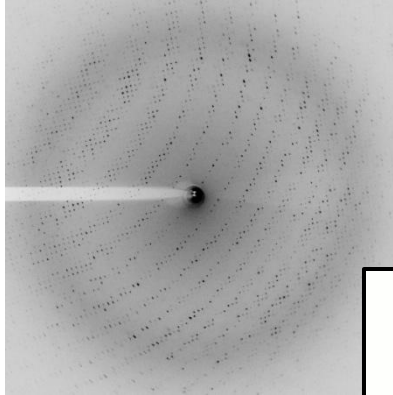
More about Maximum Likelihood

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

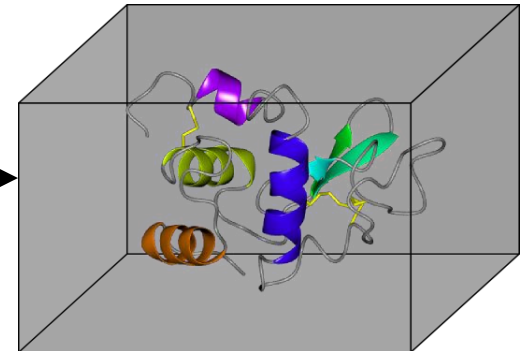
Using ML methods to account for errors
and incompleteness in the search model
Phaser can increase the chances of
successful molecular replacement



Phaser workflow

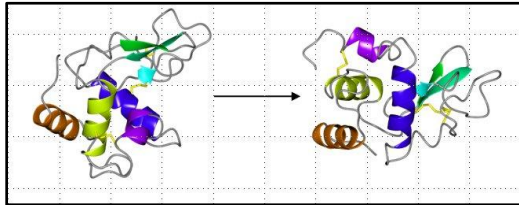


Molecular
Replacement

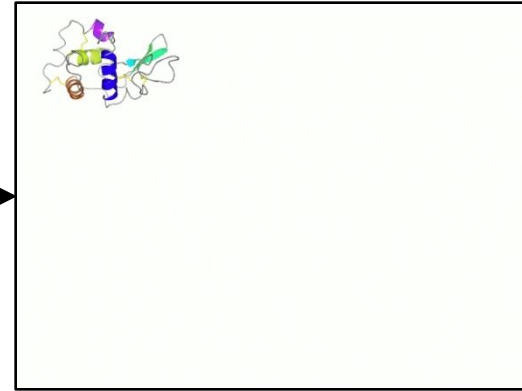




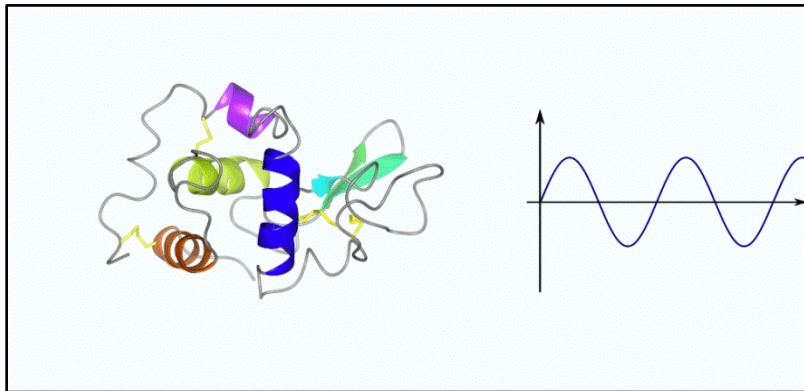
Molecular replacement workflow



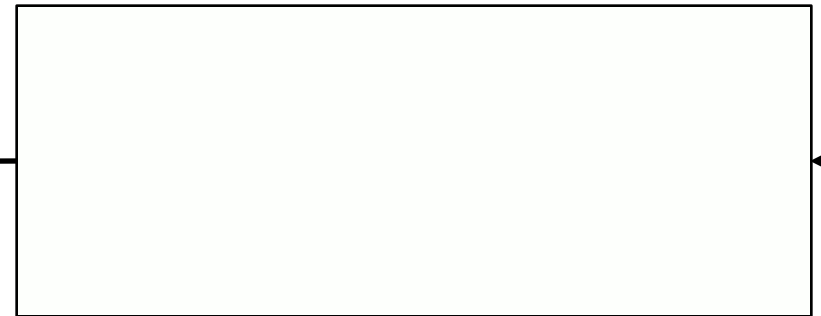
Rotation search



Translation search



Refinement

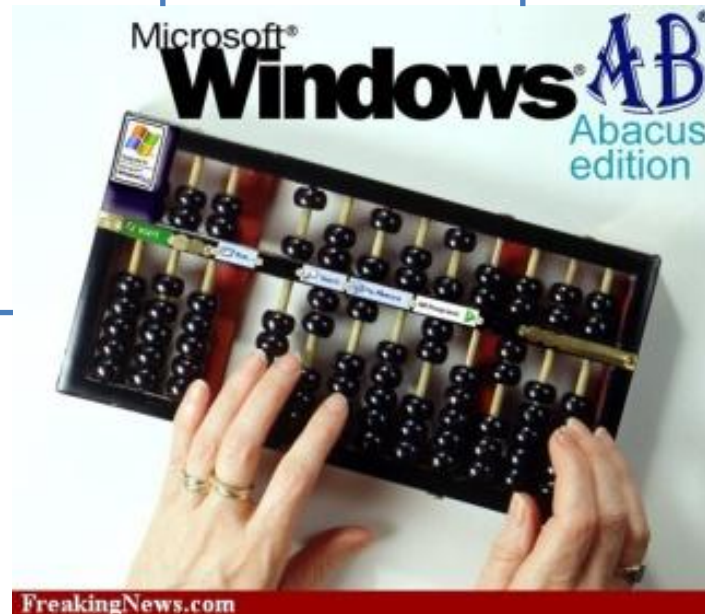
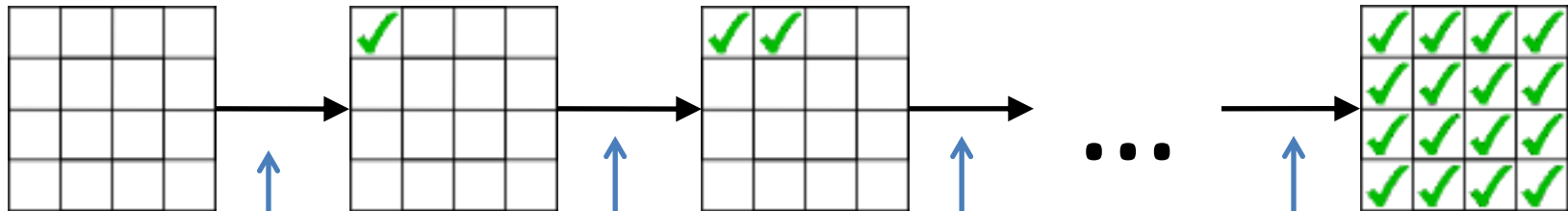


Packing test



Brute search

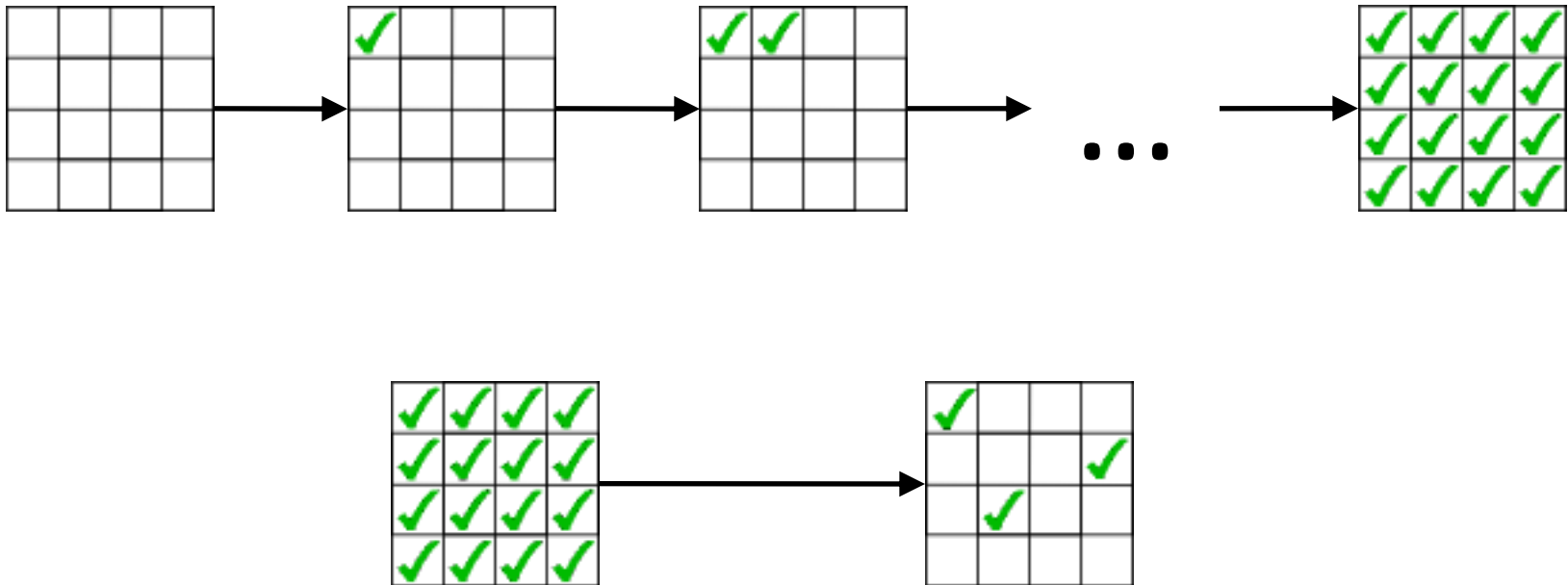
Calculate likelihood for each gridpoint





Brute search

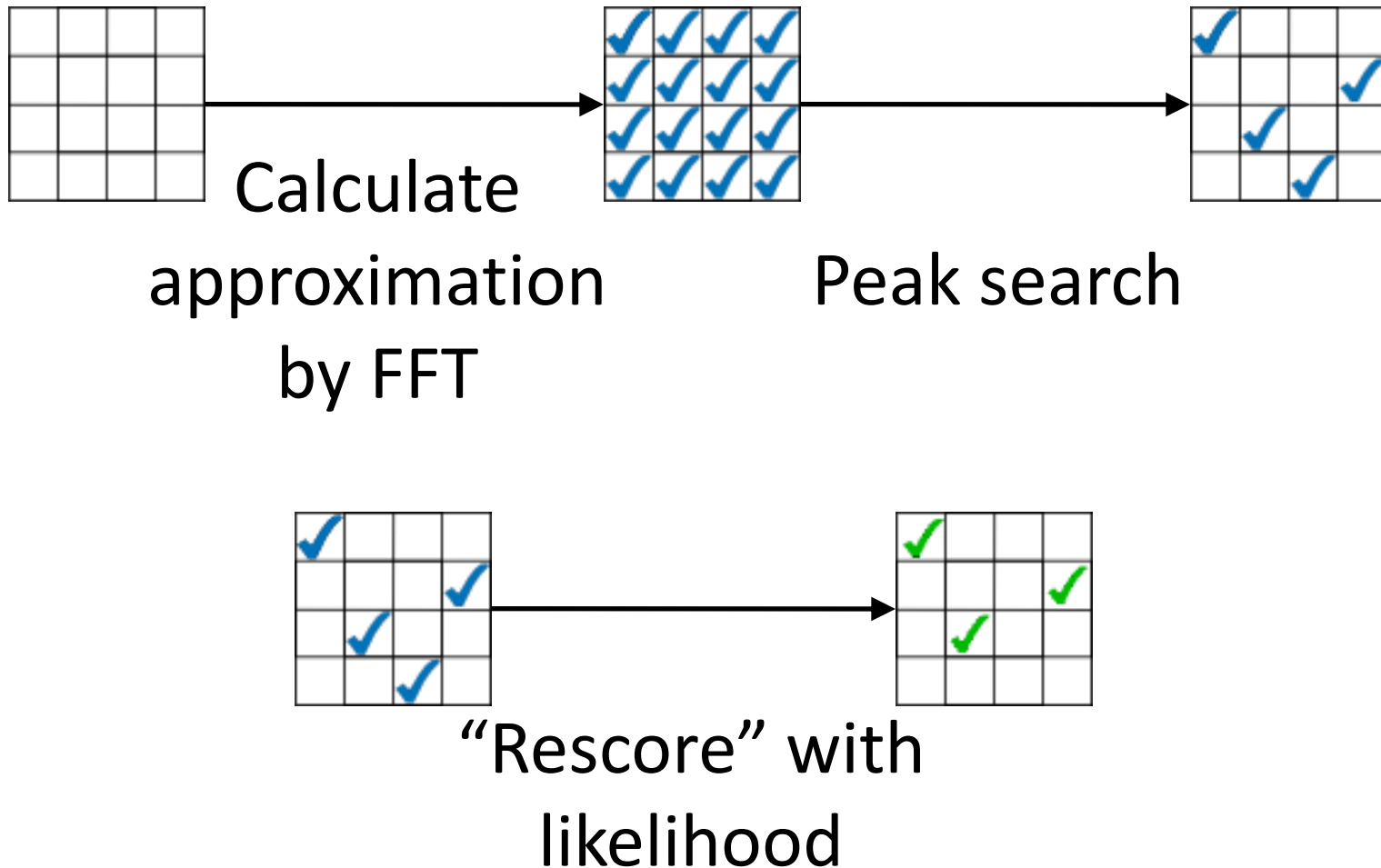
Calculate likelihood for each gridpoint



Find peaks

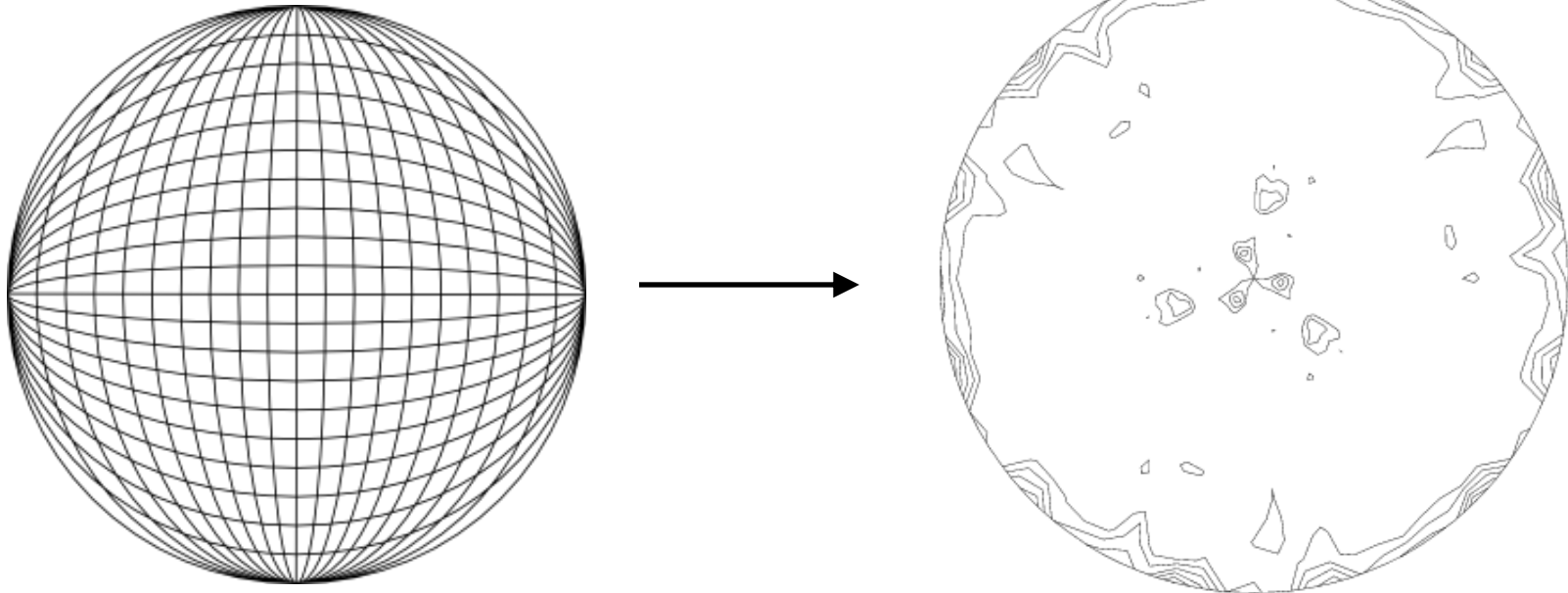


Fast search



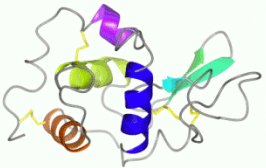


Rotation search



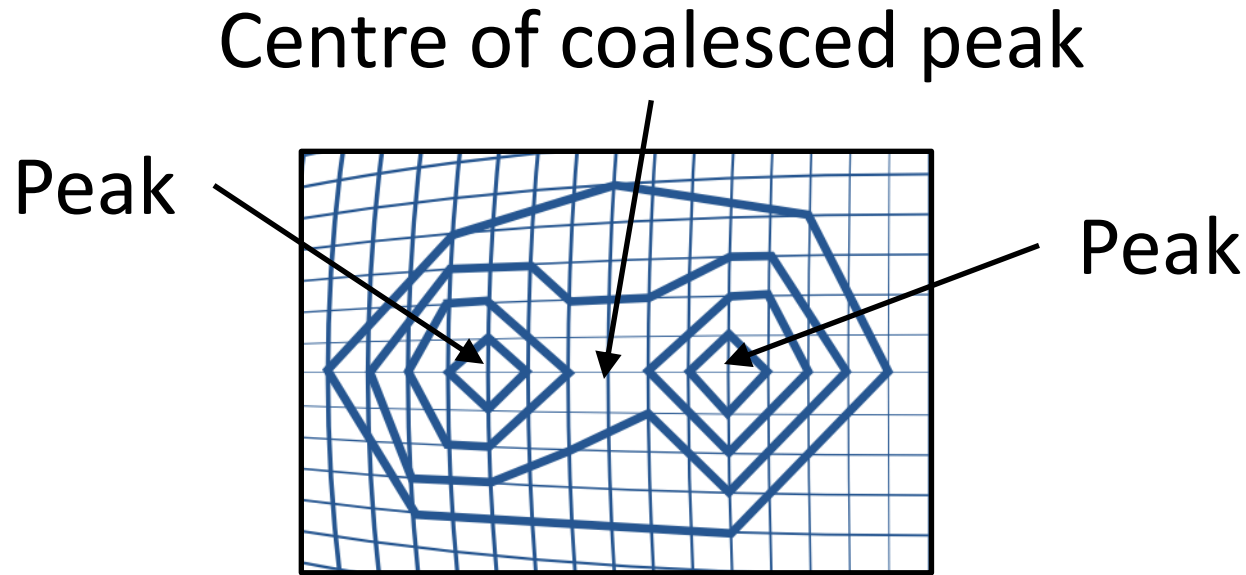
⇒ Typically small signal

⇒ Dependent only on the point group

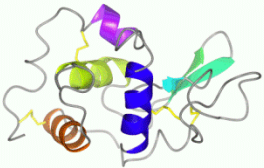




Rotation peaks

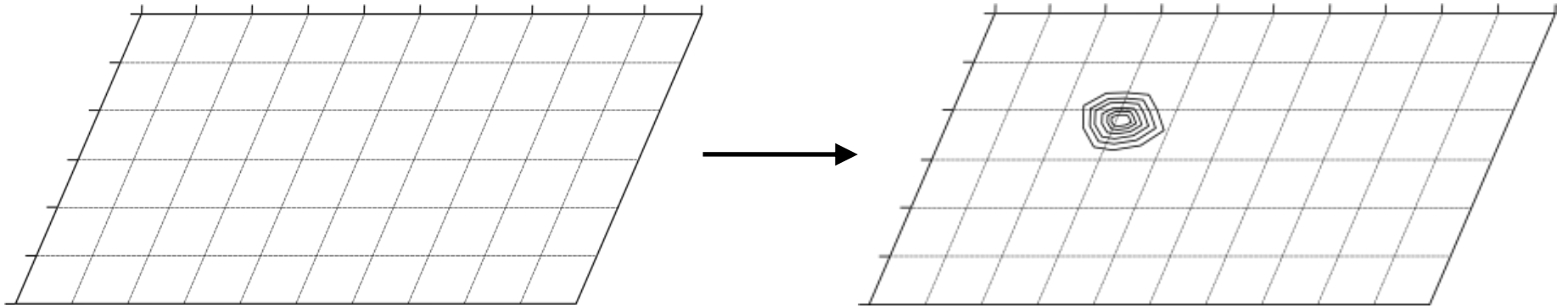


Translation function is very sensitive to orientation and may not find a solution in such a case!





Translation search

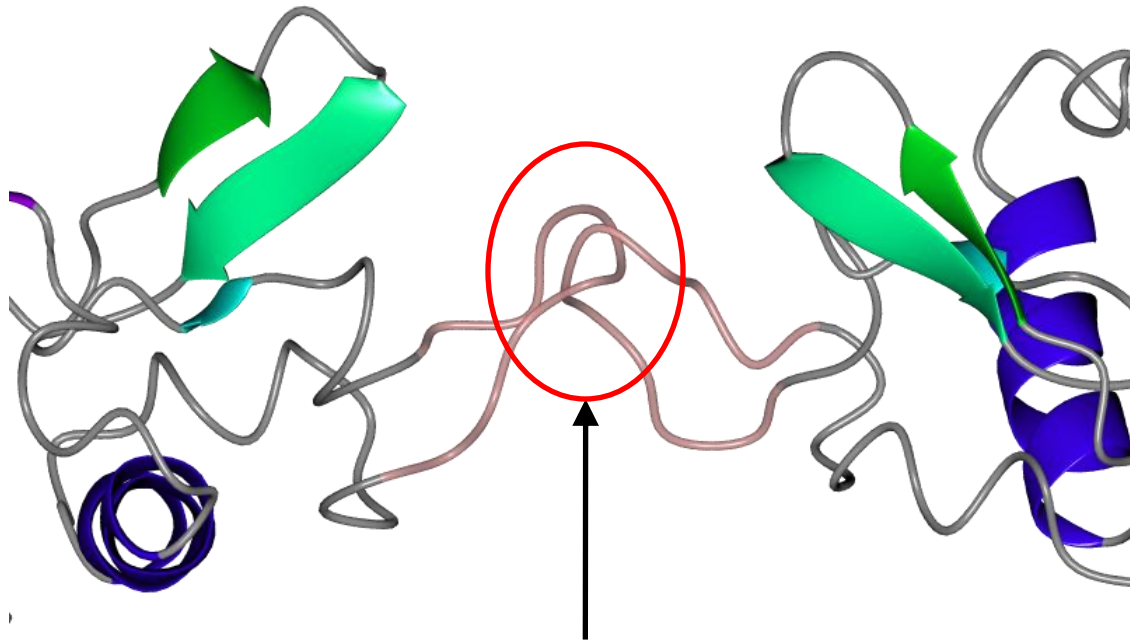


- ⇒ Solutions are recognizable (high Z-score)
- ⇒ Very sensitive to correct orientation

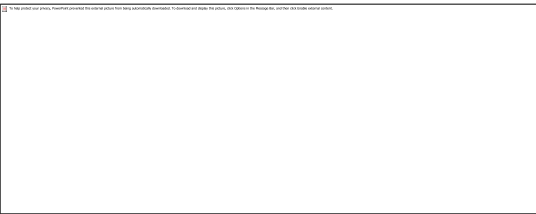




Packing

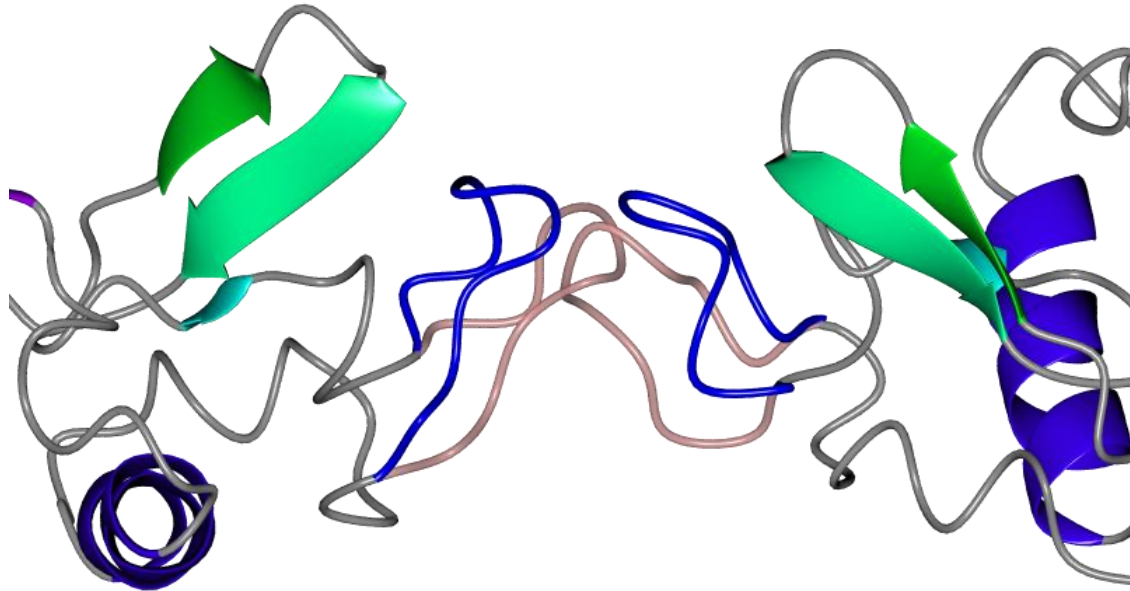


Clashes between models
decrease credibility in
solution





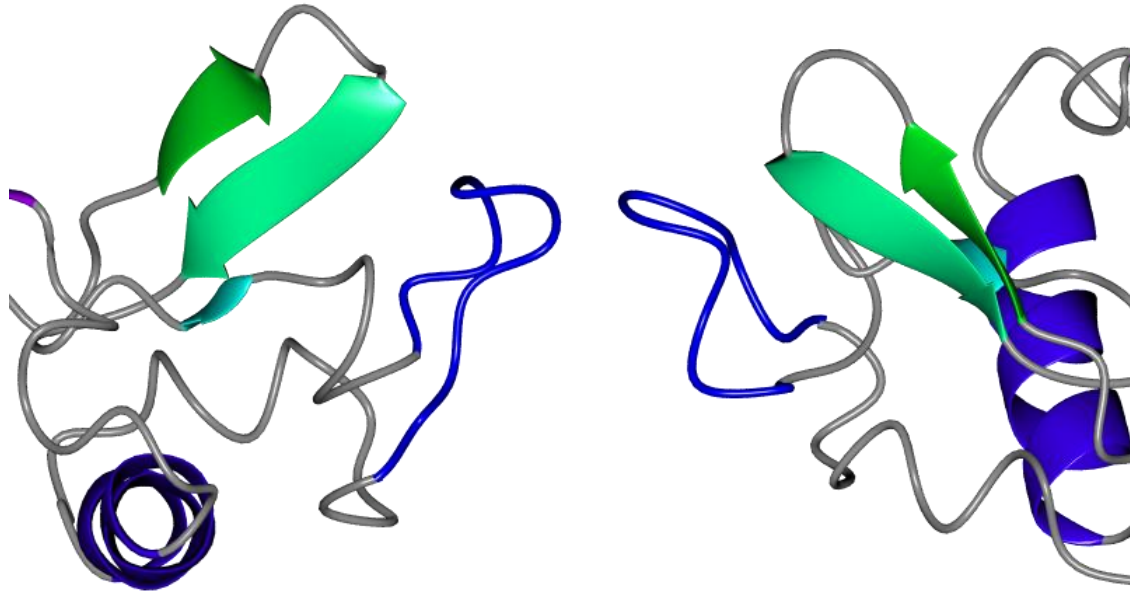
Packing



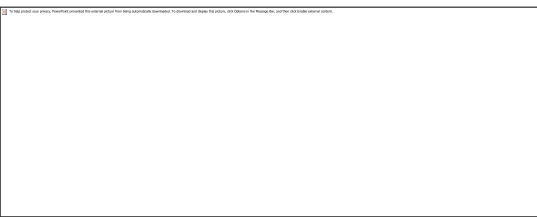
Surface loops may change
conformation



Packing



A small number of clashes
is acceptable

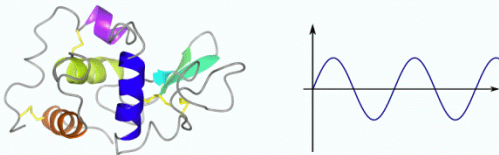
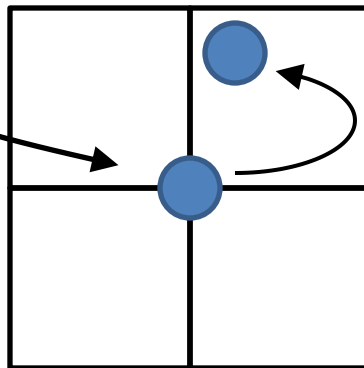




Refinement

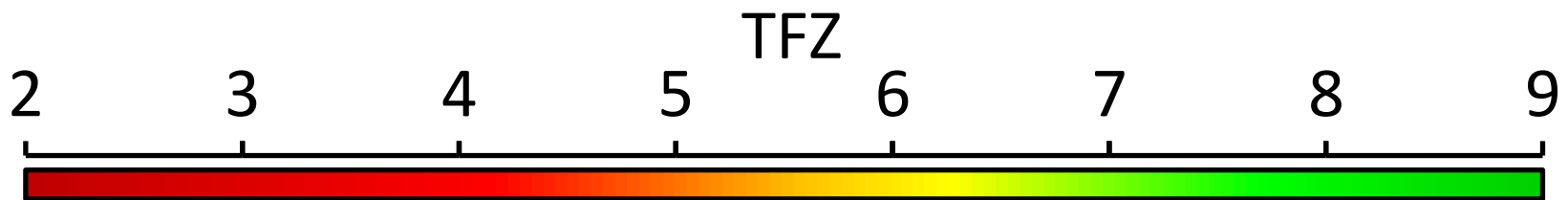
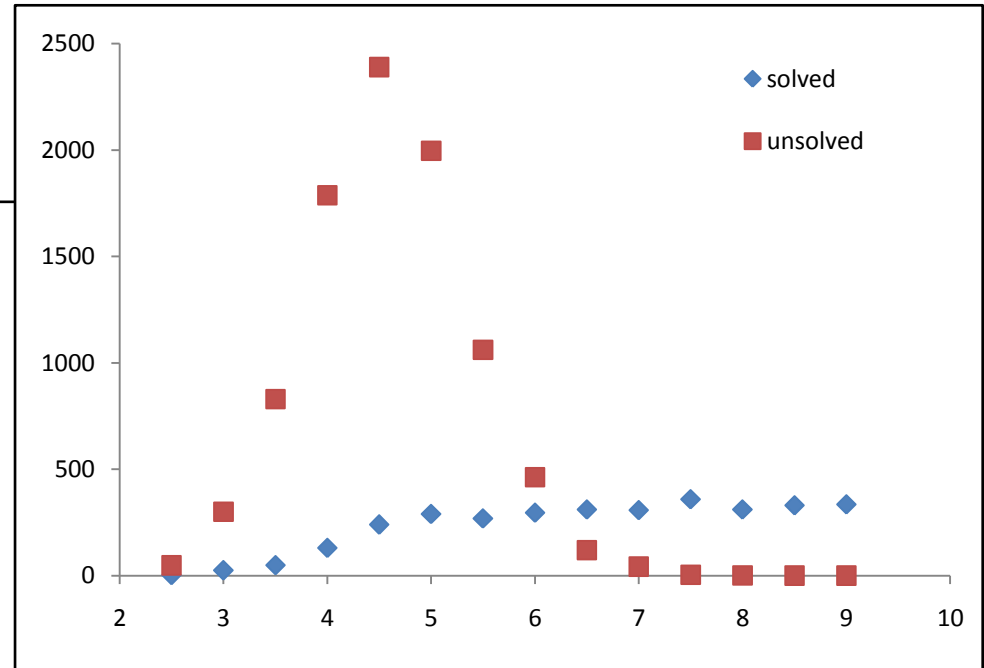
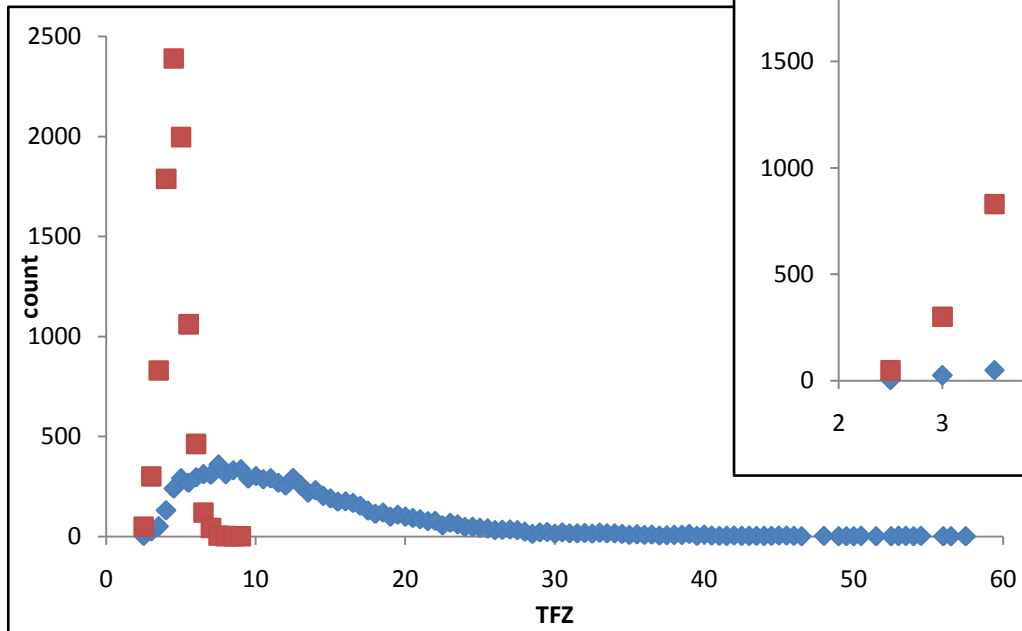
The rotation and translation functions were performed on a (not very fine) grid

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



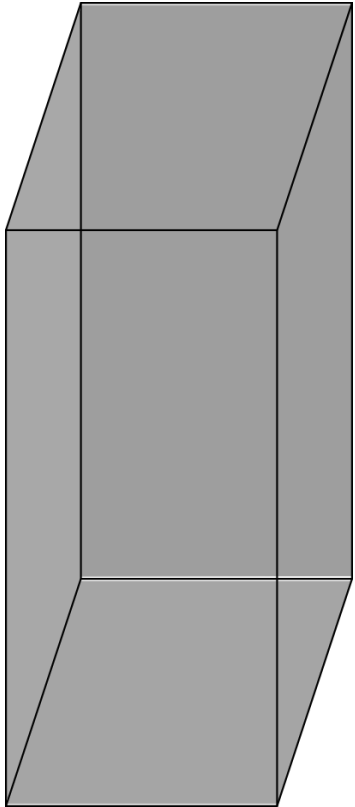


Identifying solutions





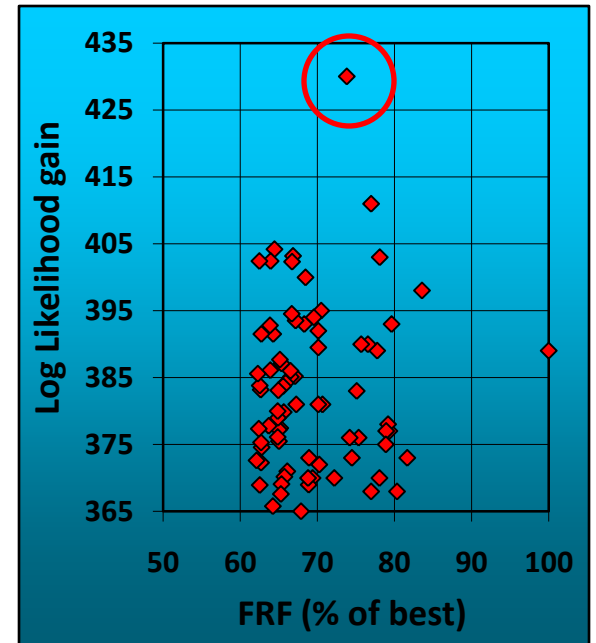
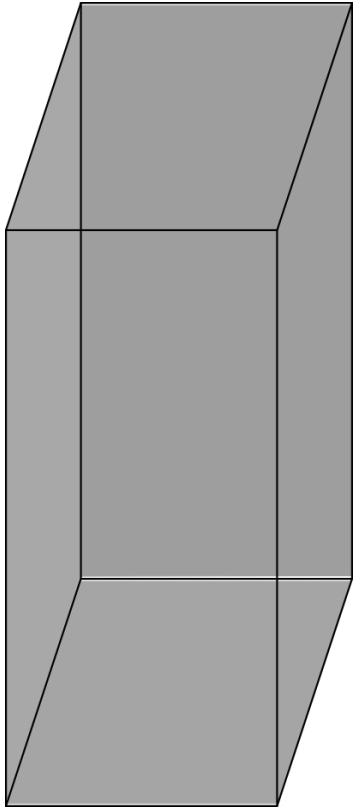
Partial structure



?

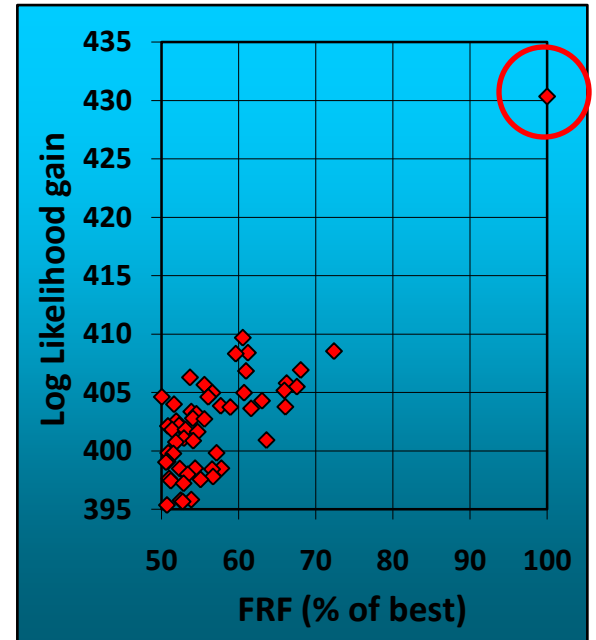
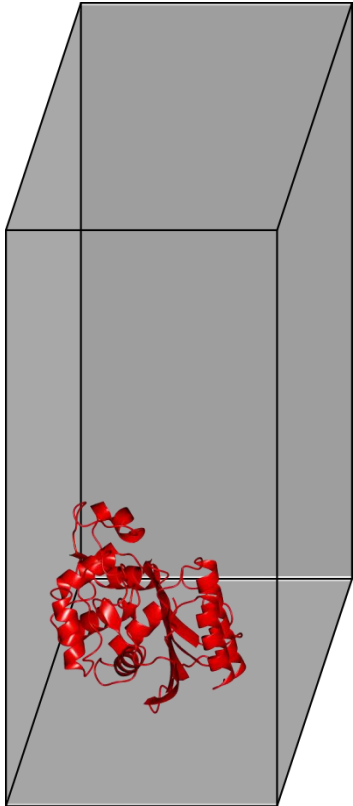


Partial structure





Partial structure

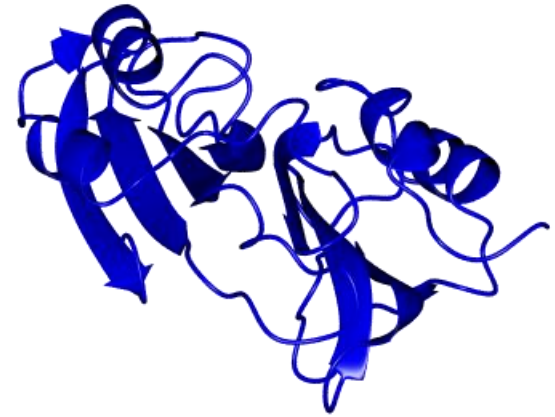




Model search order



65% of structure
35% identical model



35% of structure
50% identical model

Which model is easier to find?

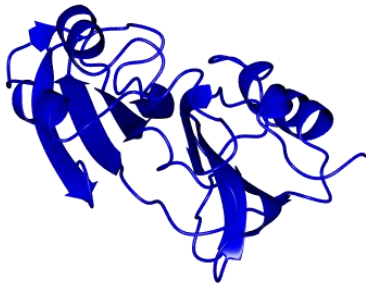


Model search order



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!



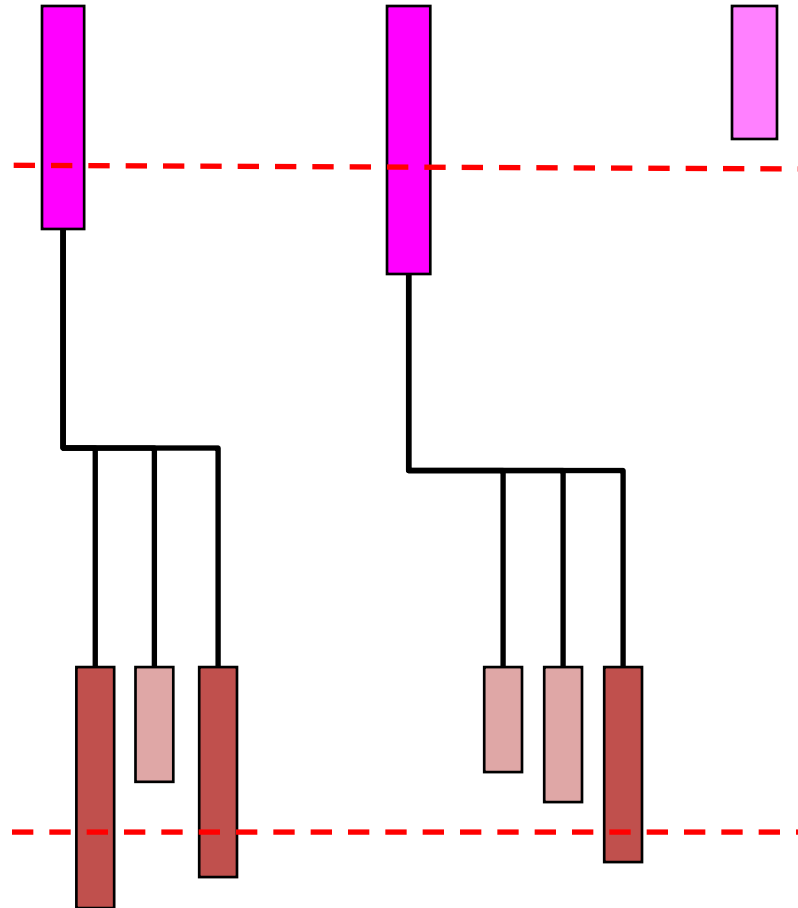
Tree search with pruning

1st component

Pruned
Propagated

Perform search

2nd component





Running Phaser via CCP4i

mode

MTZ file

target details

search model

specify search

RUN IT!

Maximum Likelihood Molecular Replacement Initial parameters from /home/mdw/molrep_tut/CCP4_DF

Job title 1v3z_B_chainsaw

Mode for molecular replacement automated search

Define data

MTZ in Full path.. /home/mdw/documents/china/hypF/hypF-1gxu-1gxt-HQ_scale Browse View

F FP1gxu SIGF SIGFP1gxu

☐ Resolution range 52.129 A to 2.5 A

Space group read from mtz file 'H 3 2'

Run Phaser with the mtz space group

Number of top solutions to output as PDB and MTZ files NB: MTZ files not output for all modes

Composition of the asymmetric unit

Component #1 protein molecular weight 10451 Number in asymmetric unit 1

Edit list Define another component

Define ensembles (each comprising possible model(s) for a protein/nucleic acid component)

Ensemble # 1

Ensemble Name ensemble1 Define ensemble via pdb file(s)

PDB #1 molrep_tut 1v3z_B_chainsaw1.pdb Browse View

Similarity of PDB #1 to the target structure sequence identity 0.38

Edit list Add superimposed PDB file to the ensemble

Edit list Add ensemble

Search details

Allow search with alternative ensembles (models) for a single component of the ASU off

Perform search using ensemble1 Number of copies to search for 1

Edit list Add another search

☐ Final selection criterion for ROTATION search peaks percentage of top peak

☐ Final selection criterion for TRANSLATION search peaks percentage of top peak

☐ Packing tolerance: number of allowed clashes

Define search sets (any already known rotations/translations)

Additional parameters

Run Save or Restore Close



Fast Rotation Function

Euler angles
(CCP4)

Top LLG and Z-
scores for FRF

CCP4I fileviewer 20_phaser_MR.log

Top Peaks With Clustering From Rescoring of FRF #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. $0.75 * (\text{top} - \text{mean}) + \text{mean}$)
There were 6 sites over 75% of top
The sites over 75% are:

#	(#)	Euler1	Euler2	Euler3	LLG	Z-score	Split	#Group	raw/top
1	1	30.3	8.5	235.9	+17.29	4.63	0.0	43	100.0/100.0
		150.3	8.5	235.9					
		270.3	8.5	235.9					
		89.7	171.5	55.9					
		209.7	171.5	55.9					
		329.7	171.5	55.9					
2	16	38.0	8.0	244.0	+16.68	4.50	15.7	29	86.6/ 91.1
		158.0	8.0	244.0					
		278.0	8.0	244.0					
		82.0	172.0	64.0					
		202.0	172.0	64.0					
		322.0	172.0	64.0					
3	47	108.4	5.8	223.6	+15.78	4.32	54.3	30	82.5/ 86.0
		228.4	5.8	223.6					
		348.4	5.8	223.6					
		11.6	174.2	43.6					
		131.6	174.2	43.6					
		251.6	174.2	43.6					
4	20	59.5	7.3	333.2	+15.68	4.30	13.3	62	85.9/ 87.1
		179.5	7.3	333.2					
		299.5	7.3	333.2					
		60.5	172.7	153.2					
		180.5	172.7	153.2					
		300.5	172.7	153.2					
5	73	39.8	12.4	343.3	+13.40	3.84	17.5	1	78.9/ 78.9
		159.8	12.4	343.3					
		279.8	12.4	343.3					
		80.2	167.6	163.3					
		200.2	167.6	163.3					
		320.2	167.6	163.3					
6	209	22.6	46.6	198.9	+11.82	3.53	57.8	1	67.5/ 67.5
		142.6	46.6	198.9					
		262.6	46.6	198.9					
		97.4	133.4	18.9					
		217.4	133.4	18.9					
		337.4	133.4	18.9					

Fast Rotation Function Table

Top	(Z)	Second	(Z)	Third	(Z)
17.29	(4.63)	16.68	(4.50)	15.78	(4.32)

Find Show Summary Quit



Fast Translation Function

fractional
translation

CCP4I fileviewer 20_phaser_MR.log

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. $0.75 * (\text{top-mean}) + \text{mean}$)
 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

FRF solution
number

CCP4I fileviewer 20_phaser_MR.log

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

Top LLG and
Z-scores for
FRF



Packing

Phaser does packing check after FTF

Clashes = C α atoms closer than 2Å

Default number of clashes = 0

Think about increasing to 2 or 5

```
CCP4I fileviewer 15_phaser_MR.log
Help
-----
PACKING TABLE
-----
Solutions rejected if number of clashes > 0
#  #CLASHES  #  ACCEPTED
1   2         #  NO
2  48         #  NO
3  12         #  NO
4   5         #  NO
5   2         #  NO
6  14         #  NO
7   4         #  NO
8   2         #  NO
9   8         #  NO
10 17         #  NO

0 accepted of 10 solutions

Find Show Summary Quit
```



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



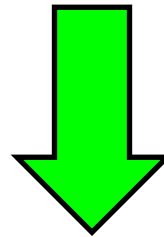
Model improvement



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

delete

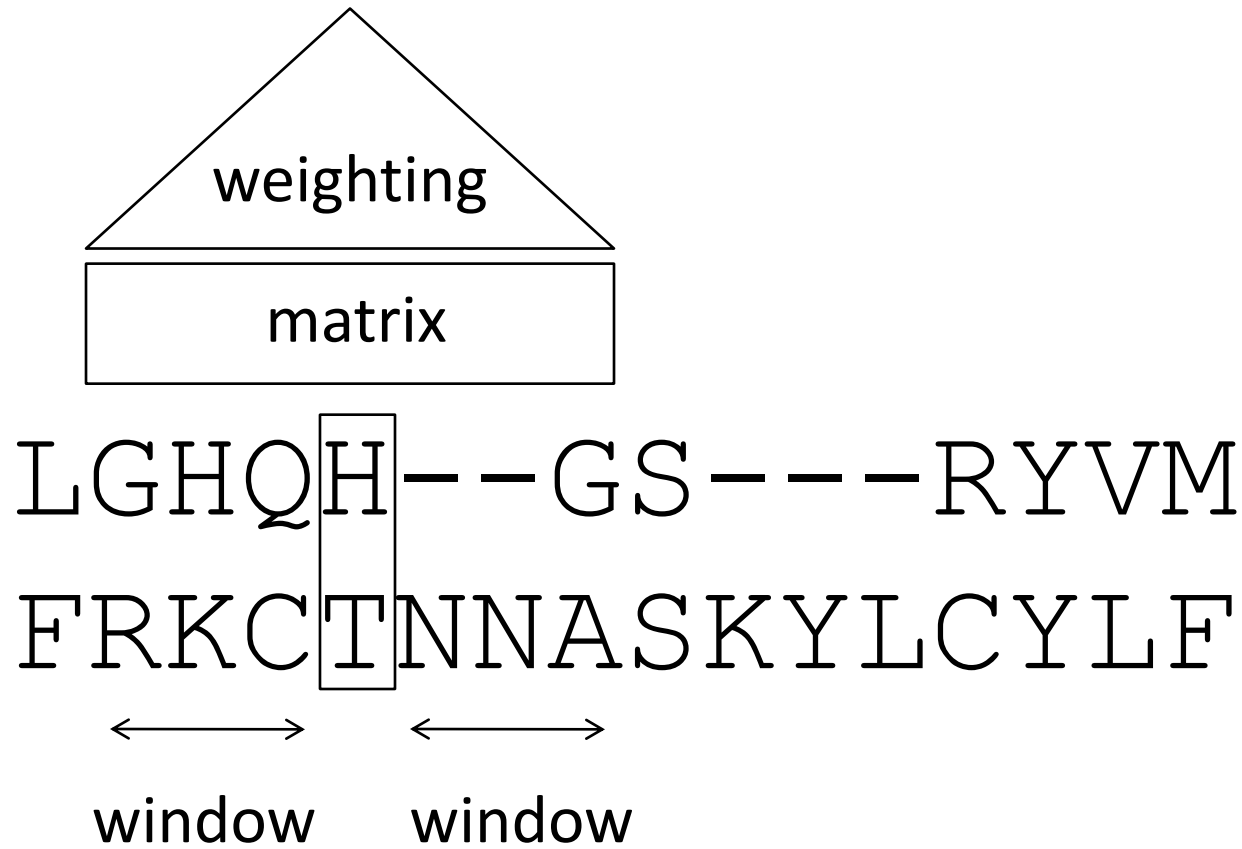
L	G	H	Q	H	-	-	G	S	-	-	-	R	Y	V	M
F	R	K	C	T	N	N	A	S	K	Y	L	C	Y	L	F

keep

CHAINSAW (CCP4)



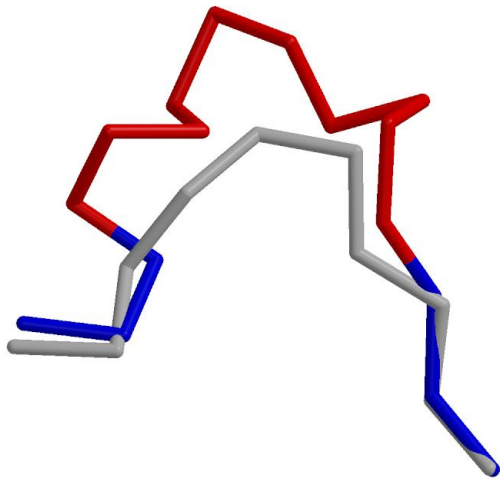
Sequence similarity



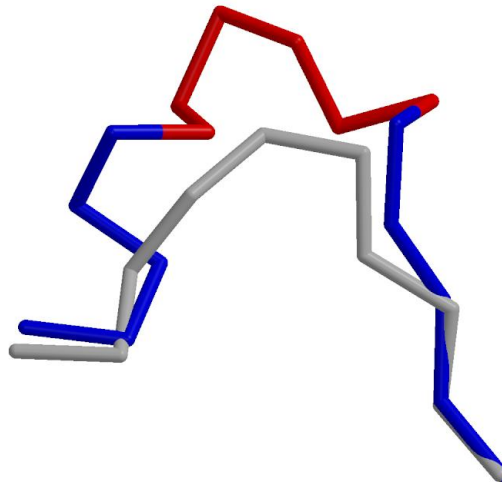
Sculptor (CCP4++)



Correcting for insertions



CHAINSAW with
structure-based alignment



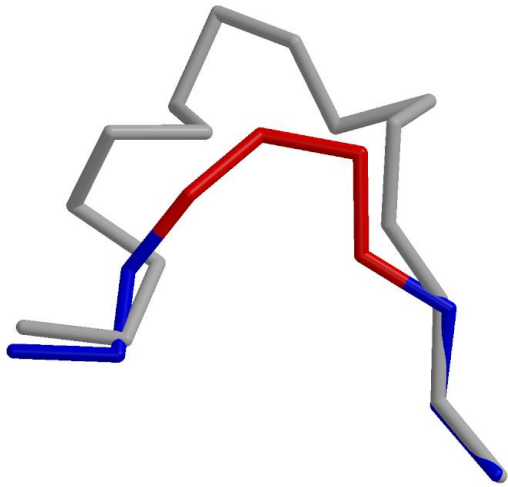
HCYKS-----GIQVR

HCVNSYQSNLDAIKIR

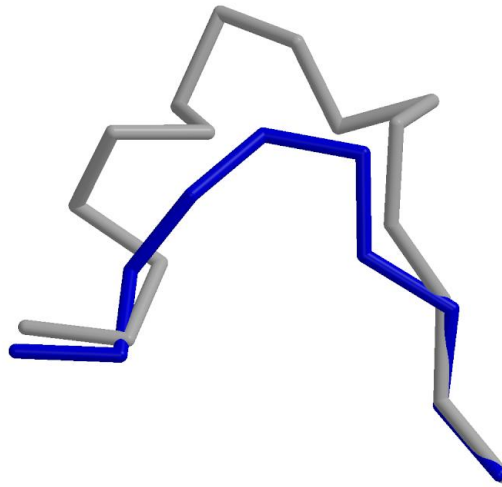
profile-profile sequence-based alignment
(FFAS)



Correcting for deletions



CHAINSAW with
structure-based alignment

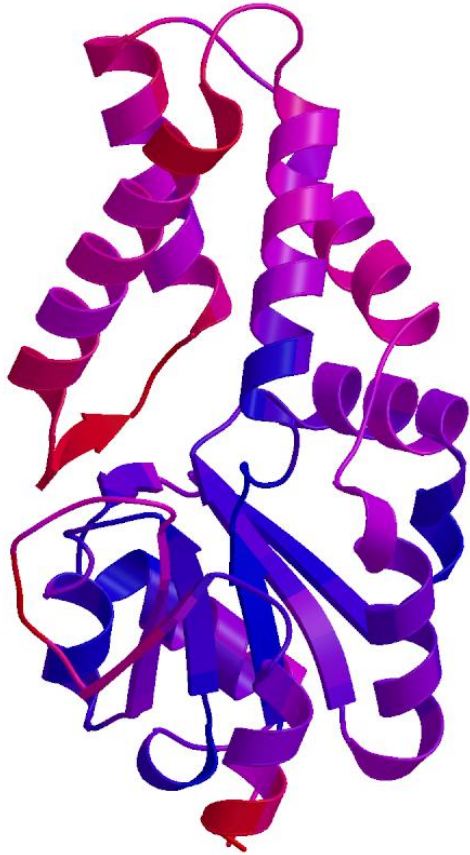


HCVNSYQSNLDAIKIR
HCYKS-----GIQVR

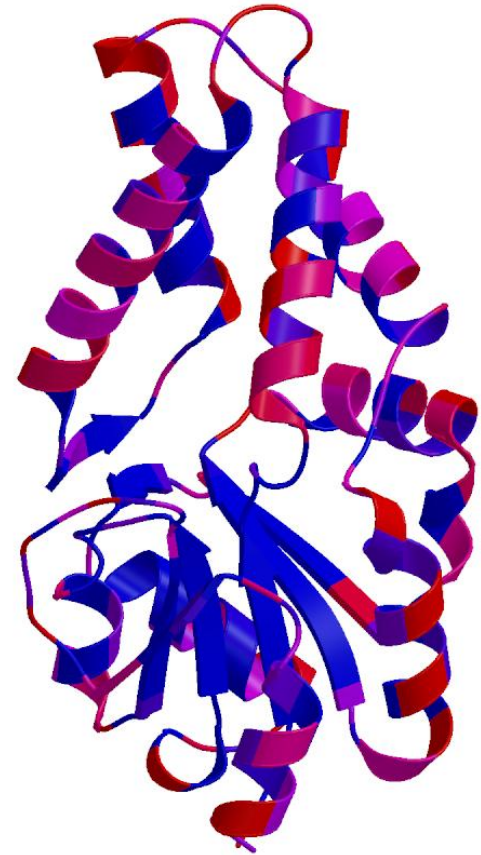
profile-profile sequence-based alignment
(FFAS)



Downweighting



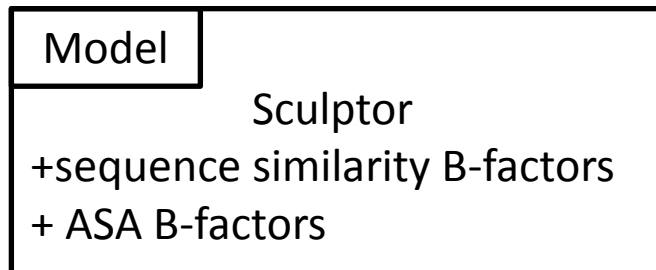
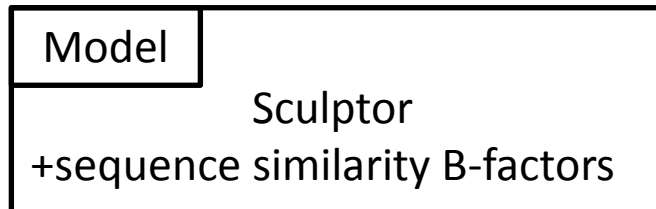
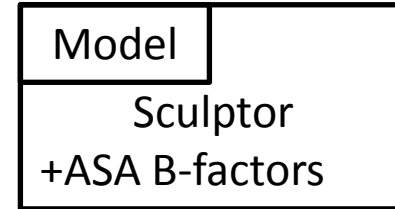
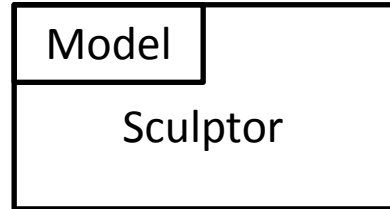
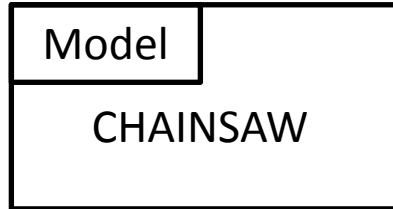
Sequence similarity
(residue level)



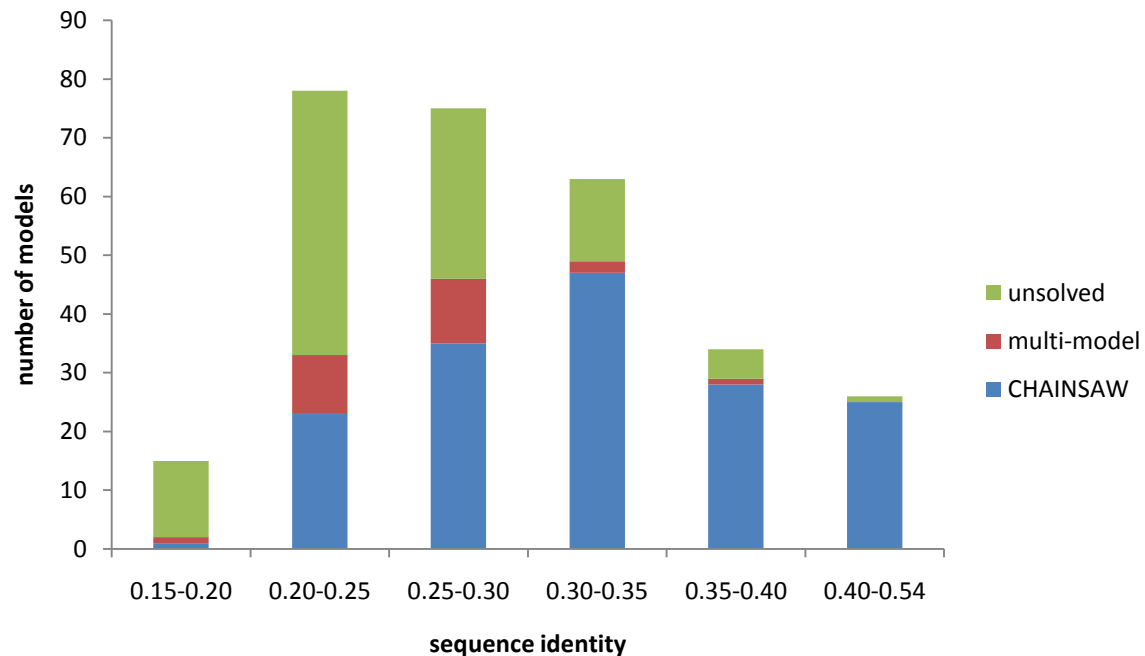
Accessible surface area
(atom level)



Multi-model strategy



Calculations with CLUSTALW alignments





Interactive model editing

Coot

File Edit Calculate Draw Measures Validate HID About Extensions

Reset View Display Manager

Sculptor GUI

Alignment: /home/gaborb/tmp/scu Browse

Model: 1ahq_trunc.pdb

Chain: A

General Mainchain Sidechain

Renumbering

- ☒ Target
- ☐ Model
- ☐ None

☒ Rename residues

☒ Generate missing Cbeta

Preview Create Cancel

Successfully read coordinates file /tmp/tmp8g9EMg/tmpgRcopP_baseline.pdb. Molecule number 2 created.

Applications Places System GBr 9 °C Fri Mar 5, 2:50 PM Gabor Bunkoczi



Acknowledgements

Python-based Hierarchical Environment for Integrated Xtallography



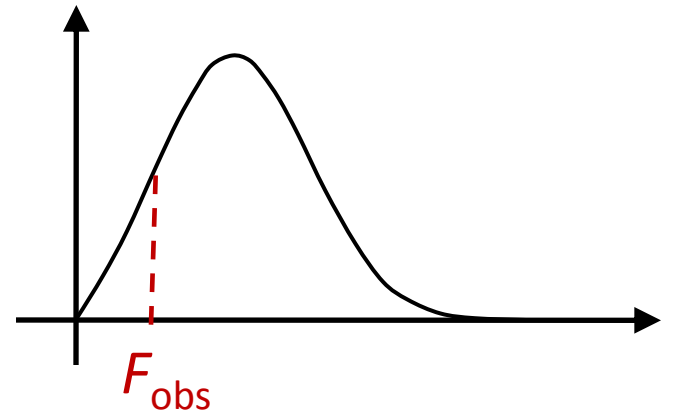
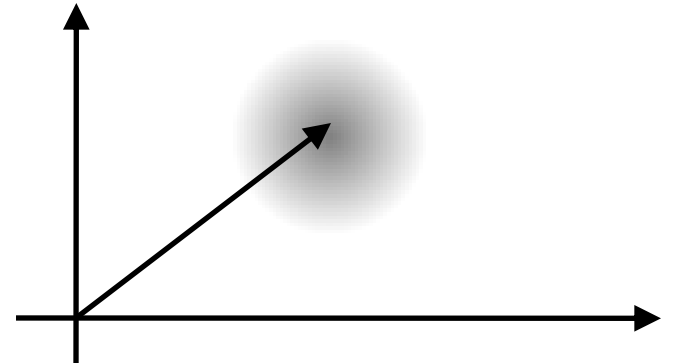
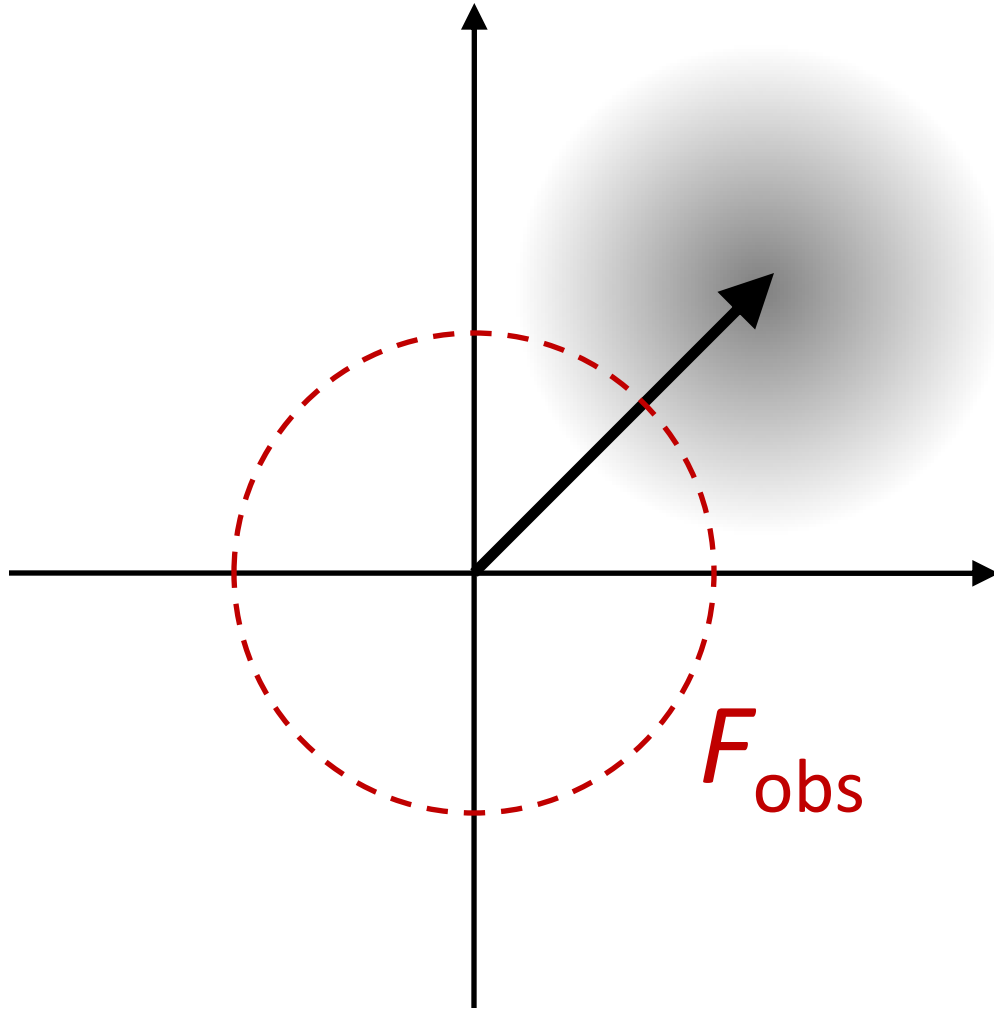
UNIVERSITY OF
CAMBRIDGE

University of Cambridge

Gabor Bunkoczi, Randy Read, Airlie McCoy, Robert Oeffner

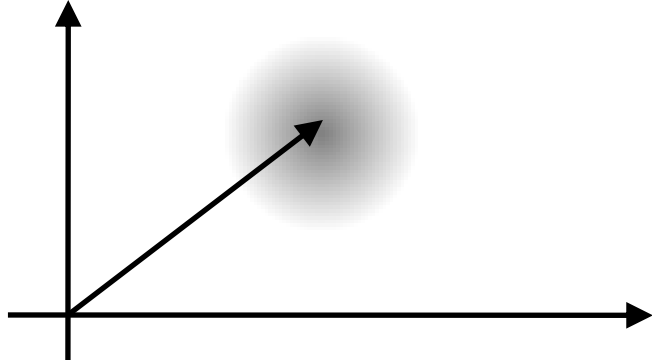


Reflection amplitude likelihood





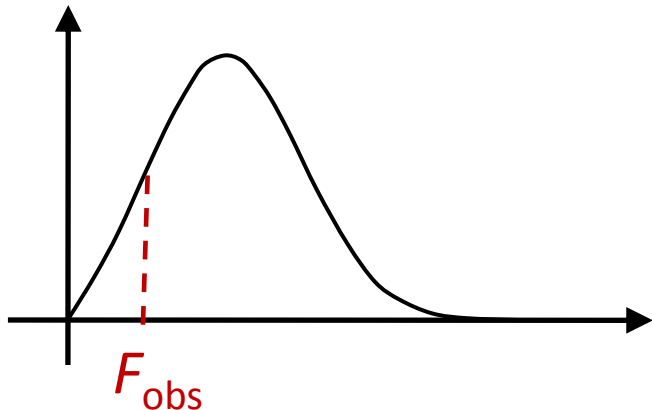
Rotation and translation functions



Identical mathematical forms!

Translation function:

Reflection amplitude likelihood



Rotation function:

Reflection amplitude likelihood
with σ_{Δ} is increased to account
for unknown translation