Molecular Replacement

Andrey Lebedev
YSBL, York

Modern Molecular Replacement

Model generation

- online resources
- internal databases

Molecular Replacement Search

- Refinement
- Search in the density

Scoring

- Refinement
- Model building

Molecular Replacement in CCP4

Programs

- Molrep
- AMoRe
- Phaser

Pipelines

- Balbes
- MrBUMP
- AMPLE

Molrep

Alexey Vagin

YSBL University of York

Molrep

Features

- Model trimming to the target sequence
- Model surface modification
- Handling pseudo-NMR (ensemble) models

- Anisotropic correction and scaling of the data
- Analysing and handling pseudo-translation
- Low- and high-pass filters accounting for model similarity and completeness

- Conventional and Phased Rotation and Translation functions
- Packing function
- Conventional MR with atomic or map models
- Fitting an atomic model into the electron density map
- Fitting two models

Molrep

http://www.ysbl.york.ac.uk/~alexei/molrep.html

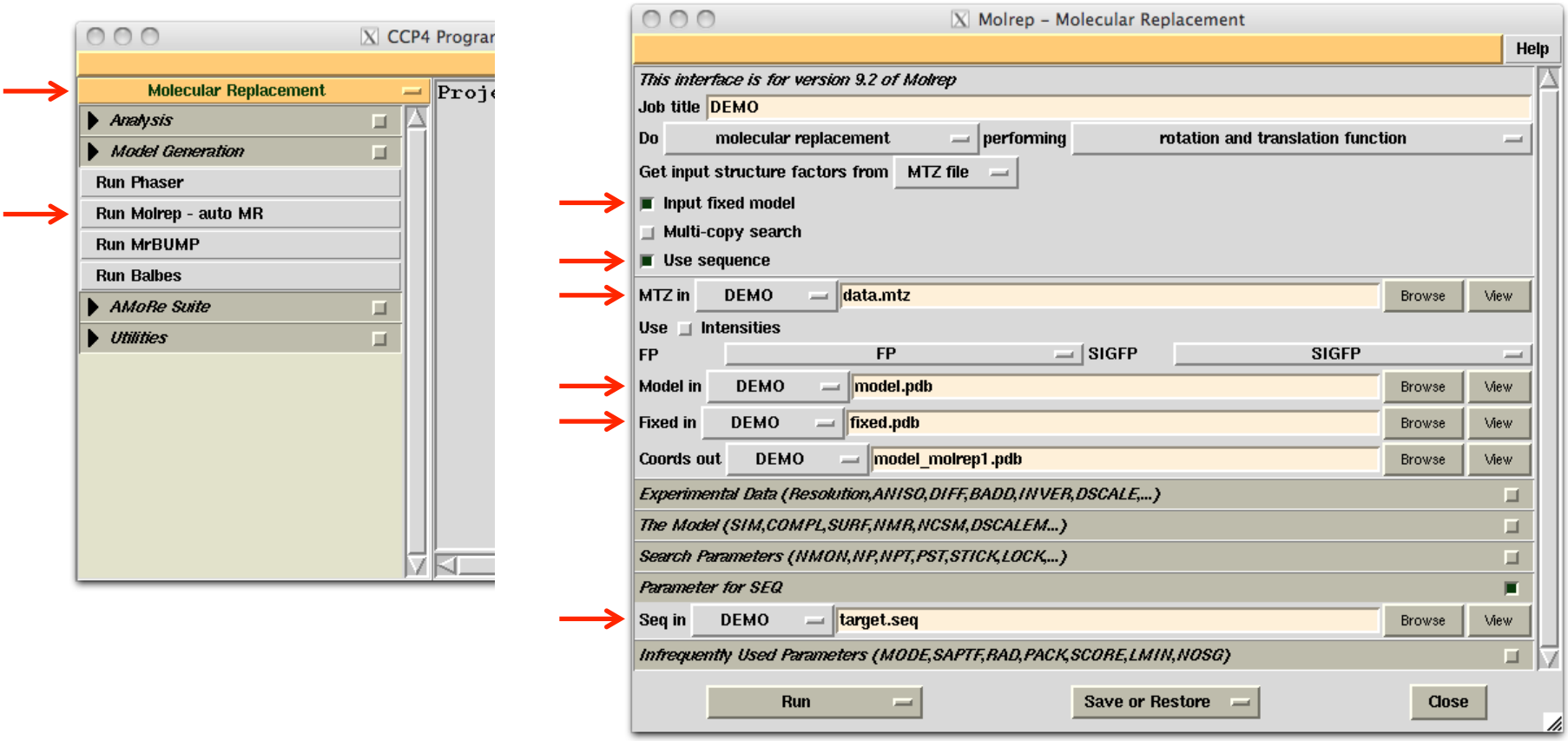
```
molrep -h
```

```
   +----------------------------------------+
   |                                        |
   |             --- MOLREP ---             |
   |       /Vers 11.0.00; 17.06.2010/       |
   |                                        |
   +----------------------------------------+
##
##  You can use program by command string with options:
##
# molrep -f <file_sf_or_map> -m <model_crd_or_map>
#        -mx <fixed model>   -m2 <model_2>
#        -po <path_out>      -ps <path_scrath>
#        -s <file_sequence>  -s2 <file_seq_for_m2>
#        -k <file_keywords>  -doc <y/a/n>
#        -h   -i  -r

                   ........
```

Molrep

```
molrep -f data.mtz -m model.pdb -mx fixed.pdb -s target.seq
```



Default protocol

```
molrep -f data.mtz -m model.pdb -mx fixed.pdb -s target.seq
```

- model correction if sequence provided
- defines the number of molecules per AU
- modification of the model surface
- anisotropic correction of the data
- weighting the data according to model completeness and similarity
- check for pseudotranslation and account for if present
- 30+ peaks in Cross RF for use in TF (accounts for close peaks)
- applied packing function

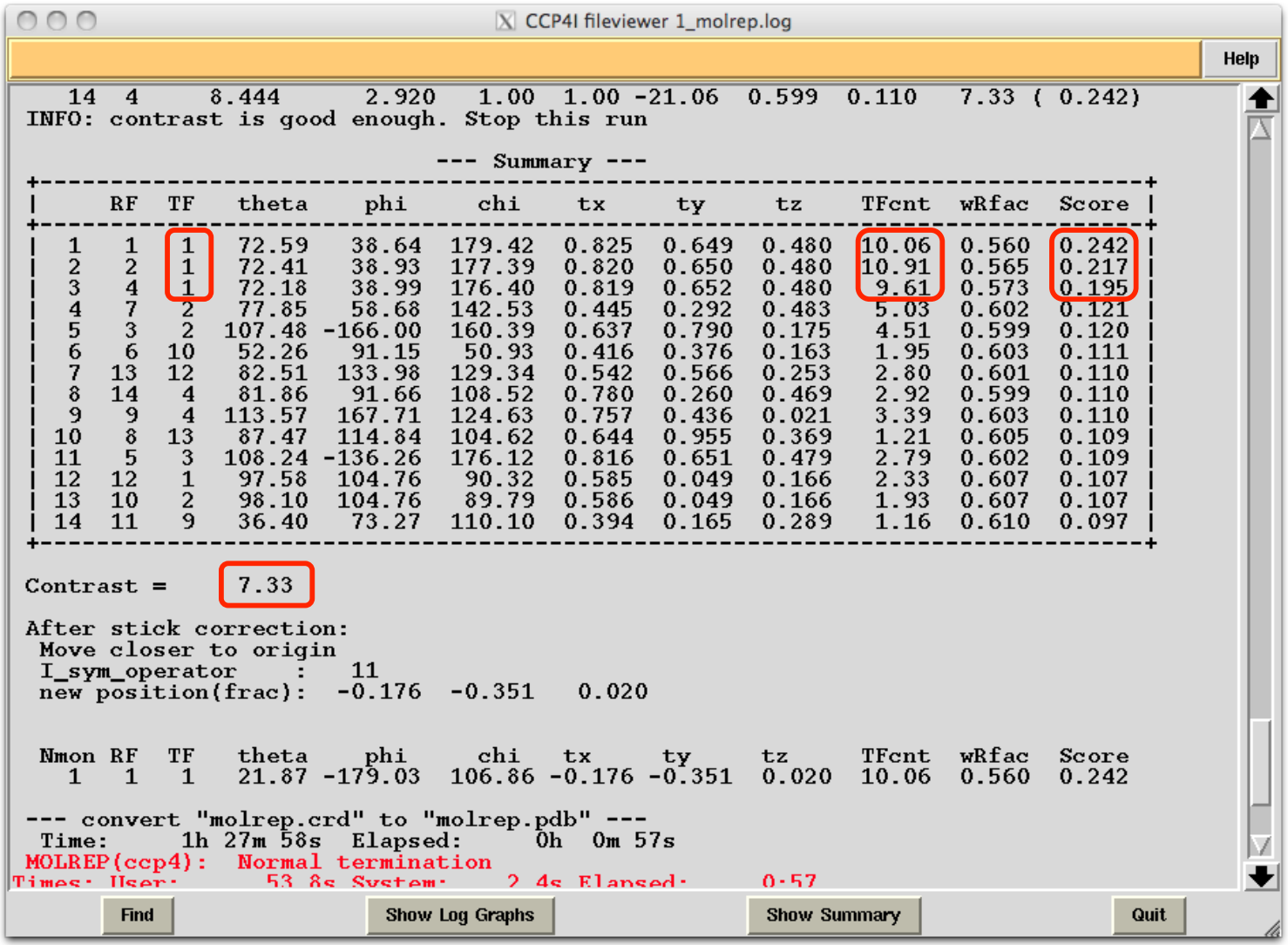
Scripting

```
molrep -f data.mtz -m model.pdb -s target.seq –i <<+
nmon 1
sim 0.33
compl 0.1
np 100
pst N
+
```

You may want to define manually

- the number of copies in the AU, if model is smaller than the target molecule
- similarity (used for weighting), if e.g. the target sequence is not provided
- completeness (used for weighting), to control weighting at low resolution
- the number of top peaks from CRF to be tested by TF
- to switch two-copy search off (switched on by default if pseudotranslation is found).

Log-file


CCP4I fileviewer 1_molrep.log
Help
14 4 8.444 2.920 1.00 1.00 -21.06 0.599 0.110 7.33 (0.242)
INFO: contrast is good enough. Stop this run
--- Summary ---
| RF TF theta phi chi tx ty tz TFcnt wRfac Score |
| 1 1 1 72.59 38.64 179.42 0.825 0.649 0.480 10.06 0.560 0.242 |
| 2 2 1 72.41 38.93 177.39 0.820 0.650 0.480 10.91 0.565 0.217 |
| 3 4 1 72.18 38.99 176.40 0.819 0.652 0.480 9.61 0.573 0.195 |
| 4 7 2 77.85 58.68 142.53 0.445 0.292 0.483 5.03 0.602 0.121 |
| 5 3 2 107.48 -166.00 160.39 0.637 0.790 0.175 4.51 0.599 0.120 |
| 6 6 10 52.26 91.15 50.93 0.416 0.376 0.163 1.95 0.603 0.111 |
| 7 13 12 82.51 133.98 129.34 0.542 0.566 0.253 2.80 0.601 0.110 |
| 8 14 4 81.86 91.66 108.52 0.780 0.260 0.469 2.92 0.599 0.110 |
| 9 9 4 113.57 167.71 124.63 0.757 0.436 0.021 3.39 0.603 0.110 |
| 10 8 13 87.47 114.84 104.62 0.644 0.955 0.369 1.21 0.605 0.109 |
| 11 5 3 108.24 -136.26 176.12 0.816 0.651 0.479 2.79 0.602 0.109 |
| 12 12 1 97.58 104.76 90.32 0.585 0.049 0.166 2.33 0.607 0.107 |
| 13 10 2 98.10 104.76 89.79 0.586 0.049 0.166 1.93 0.607 0.107 |
| 14 11 9 36.40 73.27 110.10 0.394 0.165 0.289 1.16 0.610 0.097 |
Contrast = 7.33
After stick correction:
Move closer to origin
I_sym_operator : 11
new position(frac): -0.176 -0.351 0.020
Nmon RF TF theta phi chi tx ty tz TFcnt wRfac Score
1 1 1 21.87 -179.03 106.86 -0.176 -0.351 0.020 10.06 0.560 0.242
--- convert "molrep.crd" to "molrep.pdb" ---
Time: 1h 27m 58s Elapsed: 0h 0m 57s
MOLREP(ccp4): Normal termination
Times: User: 53.8s System: 2.4s Elapsed: 0:57
Find
Show Log Graphs
Show Summary
Quit

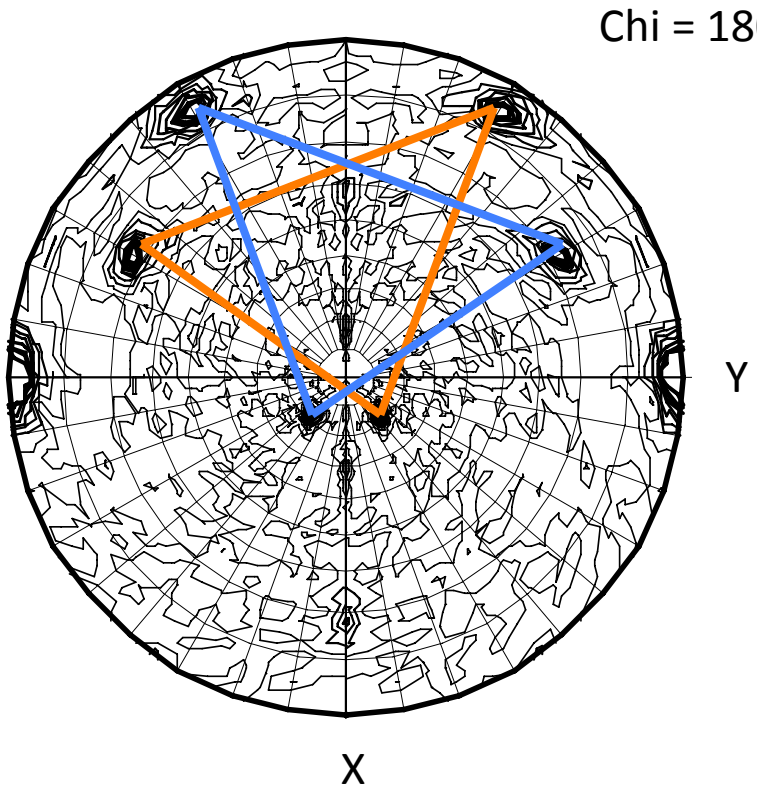
Self Rotation Function (SRF)

```
molrep -f data.mtz
```

```
molrep -f data.mtz -i <<+
rad 20
resmax 2.5
resmin 8.0
lmin 6
+
```

Example

- Space group P21
- One 222-tetramer in the AU



Output

- molrep_rf.ps
- molrep_srf.tab

LMIN < 0 (default)

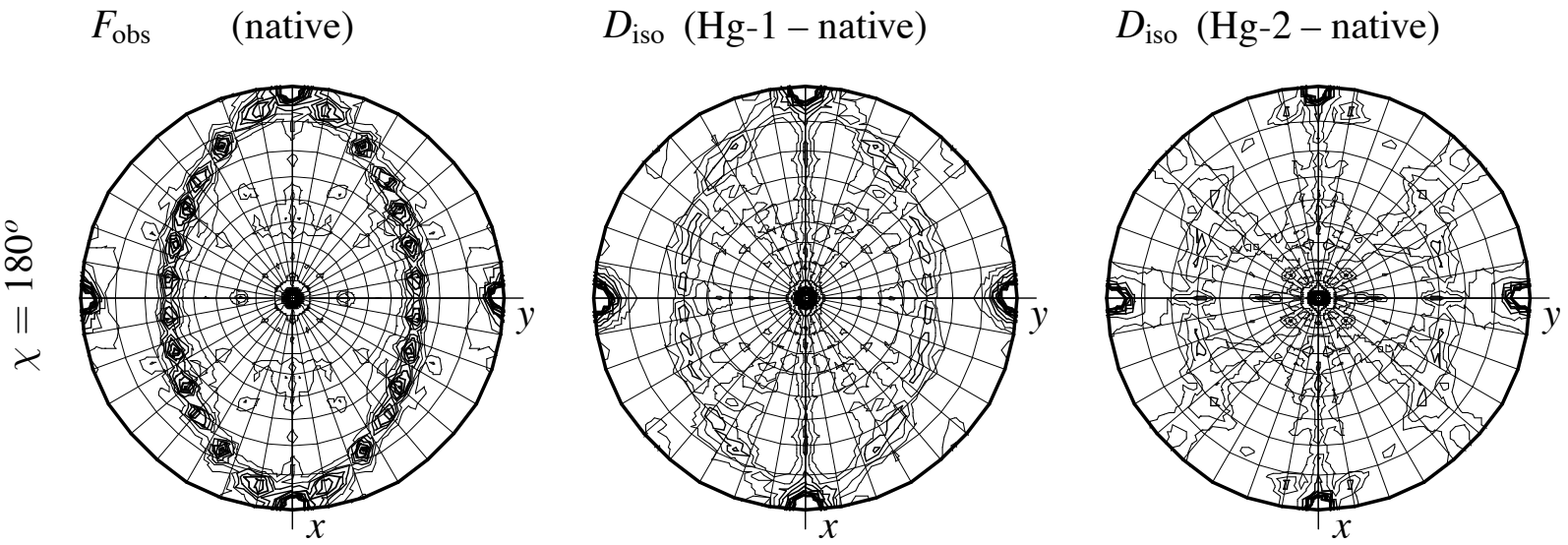
- harmonics with L=2 are removed
- harmonics with L=4 are downweighted

SRF: preliminary analysis of X-ray data

Oligomeric state of the protein in crystal

- Content of the asymmetric unit
  - Does crystal contain the right protein?
    - » Be careful: artifacts in SRF, twinning, pseudo-translation
- Selection of oligomeric search model for MR
    - » Be careful, oligomers with the same symmetry can be different

Rare use: check for isomorphous or anomalous signal

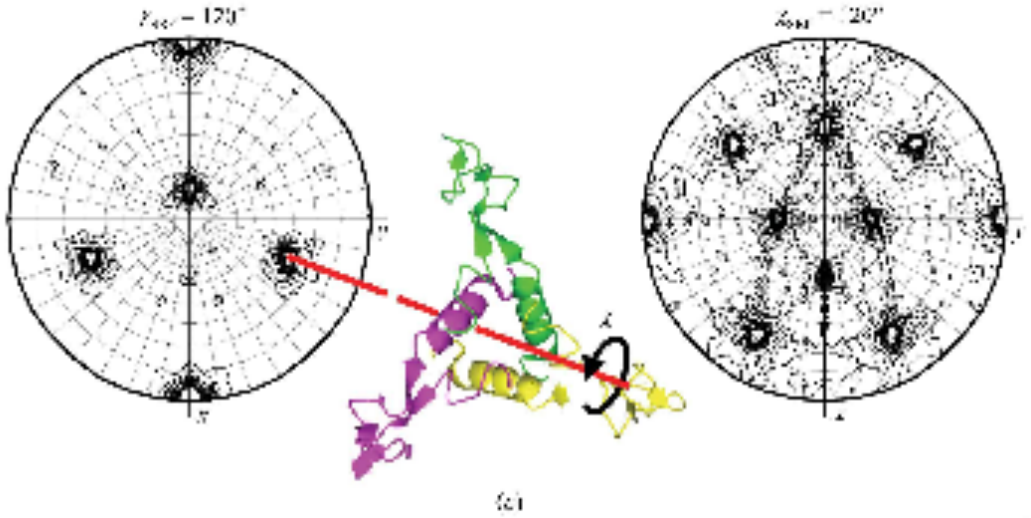


SRF: use for structure solution

Locked Rotation Function

```
molrep -f s100.mtz -m monomer.pdb -s s100.seq -i <<+
lock y
file_tsrf molrep_srf.tab
nsrf 1
+
```

Helps restrict dimensionality in an exhaustive search



Search in the density

- Completion of model
  addition of smaller domain(s) < molrep tutorial
  NCS: "the last subunit" problem
  - high temperature factors in one of the subunits
  - subunit in a special crystallographic environment

- Experimental phasing
  Both experimental phases and model are poor
  Low resolution X-ray data

- Interpretation of EM reconstruction

Exhaustive search in the electron density

FFFear: Fast Fourier Feature Recognition

Clever 6-dimensional search by Kevin Cowtan

1. Sample the 3-dimensional space of rotations
   - For example, for orthorhombic space group, search step 6.0° requires 14098 orientations (a couple of hours)

2. Find the best position(s) for each orientation
   The fast Phased Translation Function

3. Sort solution and find the overall best model

Modified Rotation Function

Why not using Rotation Function?

1. Find orientation:
   Cross-rotation function (no phases used)

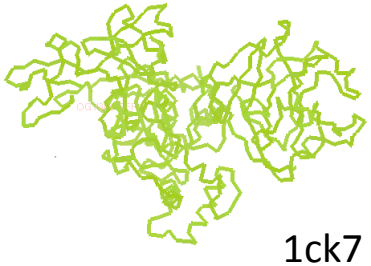
2. Find position:
   Phased translation function

Model completion:

- Small domains or subunits to be added
- Therefore the Rotation Function may fail
  - » No peaks for the domains or subunits of interest

Example

Templates:



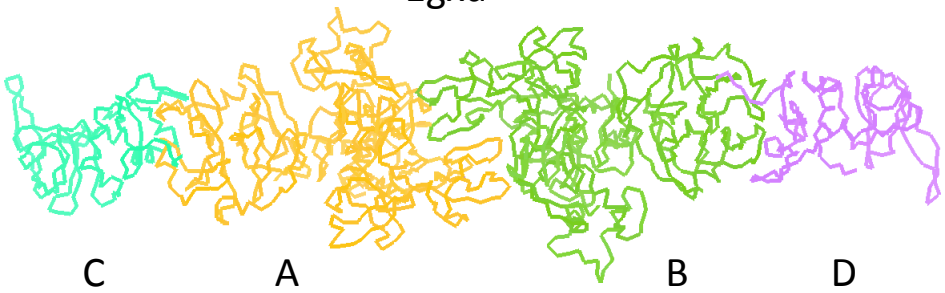
1ck7

1br9

Target structure:

- Matrix metalloproteinase-2 with its inhibitor
  - » Morgunova et al. (2002) PNAS 99, 7414
- resolution 3.1 A

1gxd



Solution:

- A, B: by conventional MR
- C, D: using FFFear
  also can be solved by iterating refinement and search in the density

Modified Rotation Function

- Refine partial model
- Calculate map coefficients (2-1 or 1-1)

```
refmac5 ... hklout AB.mtz xyzout AB.pdb ...
```

- Flatten the map corresponding to the known substructure
- Calculate structure amplitudes from this map
- Use them in Cross-Rotation Function
- And finally – Phased TF

```
molrep -f AB.mtz -mx AB.pdb -m model.pdb -i <<+
labin F=FWT PH=PHWT
sim -1
nmon 1
np 100
diff m
+
```

Example

Search for C in the density from refined A+B:

--- Summary ---

	RF	TF	theta	phi	chi	tx	ty	tz	TFcnt	wRfac	Score
1	38	2	88.09	-107.50	4.93	0.763	0.000	0.200	9.00	0.661	0.090
2	33	2	83.41	-96.71	5.51	0.763	0.000	0.200	9.38	0.661	0.090
3	31	2	177.53	-175.94	179.16	0.236	0.000	0.699	9.49	0.661	0.089
4	27	2	167.32	-104.44	51.93	0.850	0.000	0.388	2.57	0.662	0.082

Search for D in the density from refined A+B+C:

--- Summary ---

	RF	TF	theta	phi	chi	tx	ty	tz	TFcnt	wRfac	Score
1	88	1	172.00	-133.61	173.03	0.609	0.511	0.139	20.89	0.650	0.096
2	86	1	171.51	-130.07	173.56	0.108	0.011	0.140	16.55	0.650	0.095
3	87	1	172.85	-130.98	175.04	0.109	0.011	0.140	14.27	0.650	0.095
4	59	1	165.81	-139.35	167.51	0.125	0.010	0.143	9.97	0.650	0.093

Modified Rotation Function

Useful rules

- Add one domain at time, "NMON 1"
- Use "SIM –1" (Refinement has already weighted the map coefficients)
- Use many picks of RF, e.g. "NP 100"
- The second copy of a domain is sometimes easier to find using its refined copy found previously (a correct solution of the first copy)

Compared to the likelihood based RF

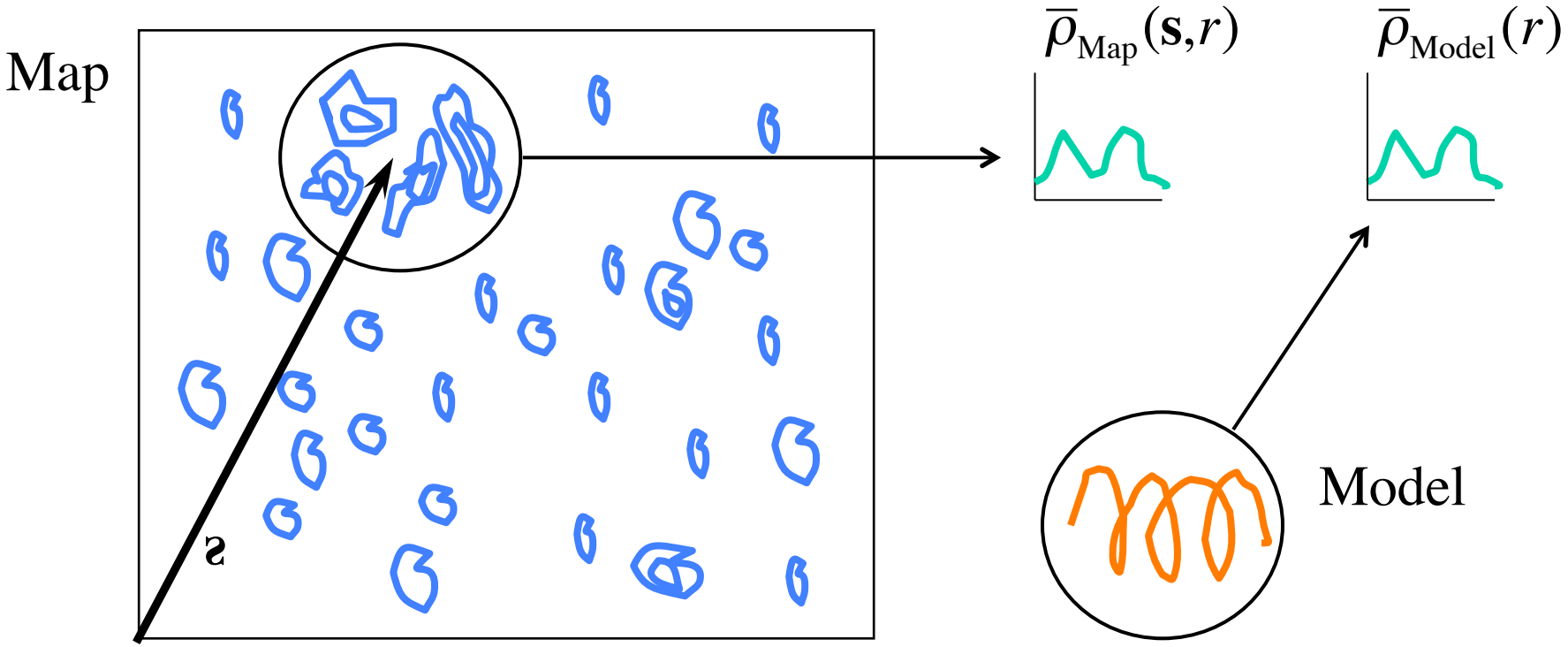
- The likelihood estimates for map coefficients are obtained from refinement
- In addition, the known substructure is improved before next search
- In addition, the noise in the map from known substructure is removed

This method is implemented in the MR pipeline Balbes

SAPTF

Spherically Averaged Phased Translation Function
(FFT based algorithm)

$$\mathrm{SAPTF}(\mathbf{s}) = \int \overline{\rho}_{\mathrm{Map}}(\mathbf{s},r)\ \overline{\rho}_{\mathrm{Model}}(r)\ r^2\,\mathrm{d}r$$



MR with SAPTF

1. Find approximate position:
   Spherically Averaged Phased Translation Function

2. Find orientation:
   Phased Rotation Function
   - Local search of the orientation in the density

3. Verify and adjust position:
   Phased Translation Function

SAPTF Example

X-ray data:

- Crystal of cyanobacterial sucrose-phosphatase

PDB code 1tj3

Resolution, 2.8 Å

Model:

- Identity to the target 100%
- Different conformation

PDB code 1s2o

Derived models:

- domain 1
  172 residues (1-77, 159-244)
- domain 2
  72 residues (88-159)

SAPTF Example

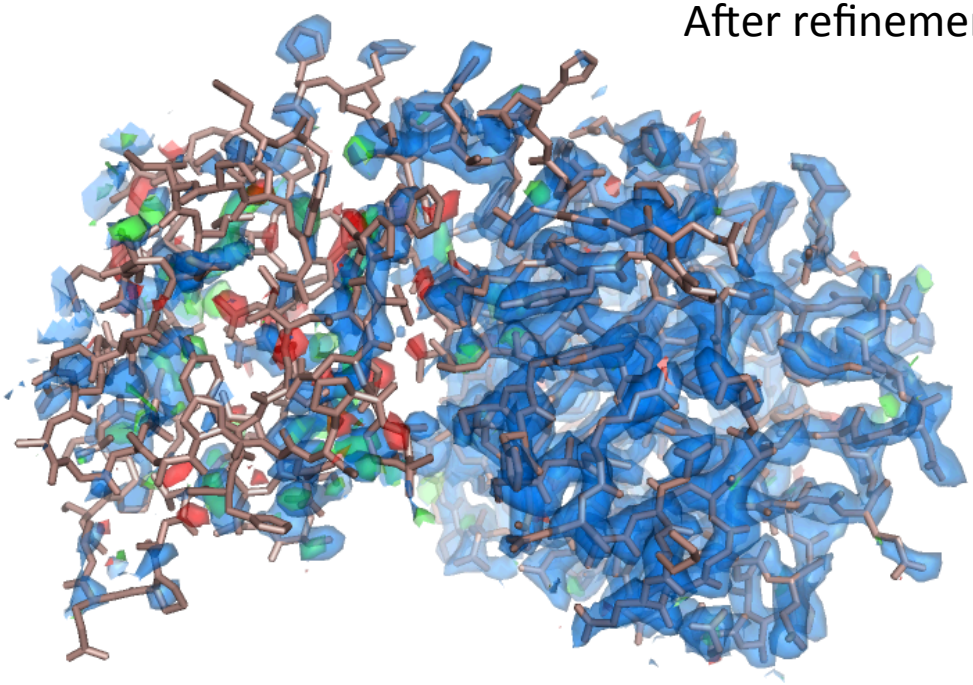
Attempt to find the complete search model (Conventional RF + TF protocol)

```
molrep -f 1tj3.mtz -m 1s2oA.pdb
```

Input:

- X-ray data
- search model

After refinement



SAPTF Example

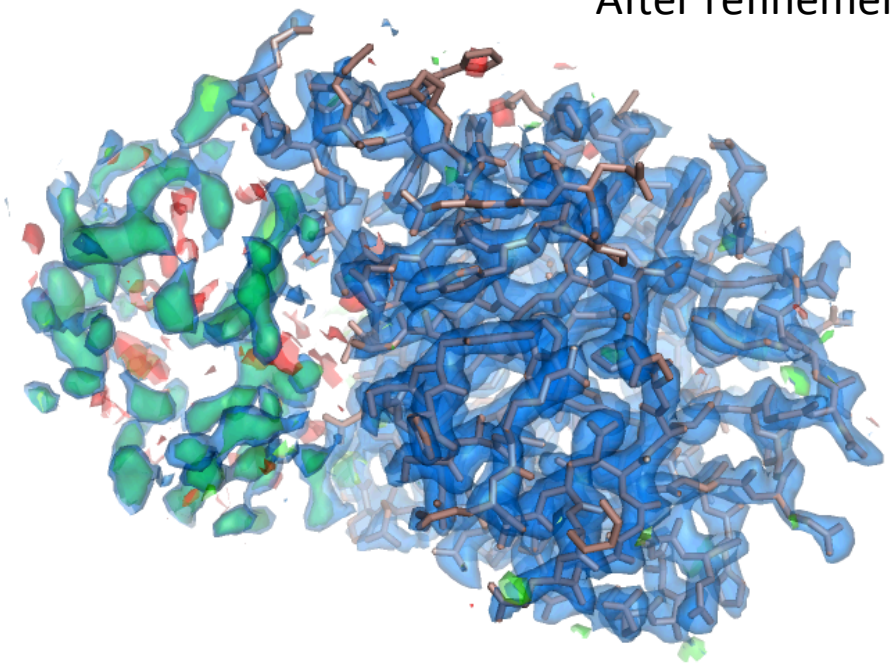
Search for the large domain (Conventional RF + TF protocol)

```
molrep -f 1tj3.mtz -m 1s2oA_dom1.pdb
```

Input:

- X-ray data
- search model

After refinement



SAPTF Example

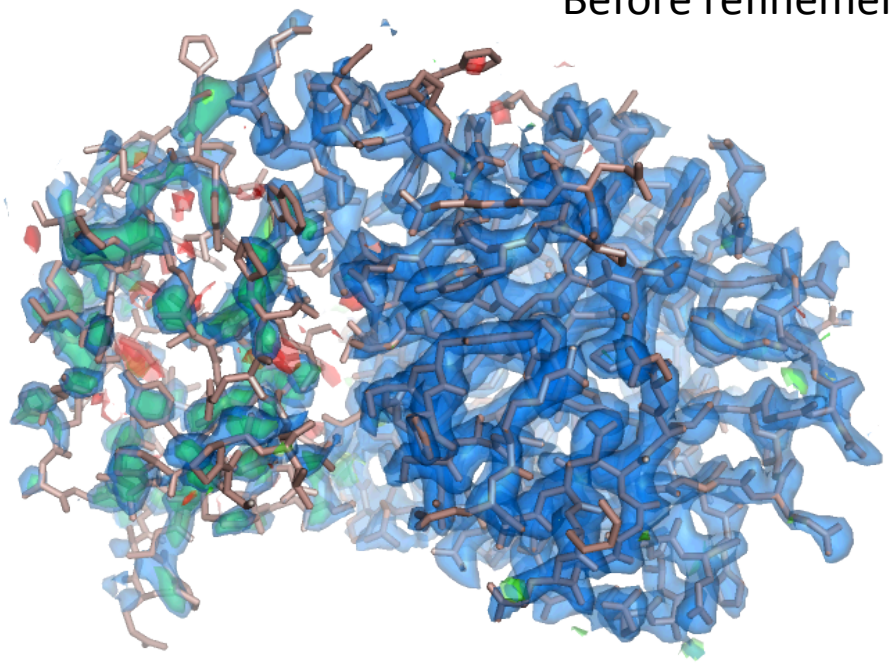
Search for the small domain (SAPTF + Phased RF + Phased TF)

```
molrep -f data.mtz -m 1s2oA_dom2.pdb -i <<+
diff M
labin FP=FWT PHIC=PHIWT
prf Y
sim -1
+
```

Input:

- Map coefficients
- Search model
- Partial structure
  - used as a mask
  - used for Packing Function
  - passed to output PDB-file

Before refinement



SAPTF Example

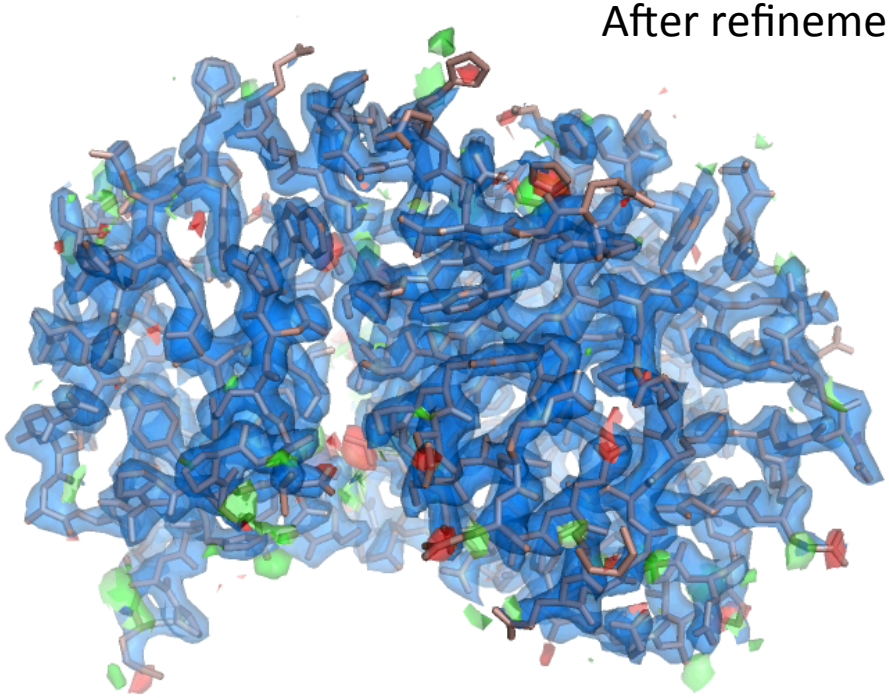
Search for the small domain (SAPTF + Phased RF + Phased TF)

- Tutorial data (typo in tutorial materials)
  http://www.ysbl.york.ac.uk/~alexei/downloads/tutorial_MR.tar.gz

- CCP4I
  - this workshop materials

- Command line:
  - tutorial data
    - » tutorial.pdf, section 2

After refinement



Alternative SAPTF protocol

- SAPTF estimate of the position is not very precise
- Passed RF is sensitive to eccentricity of the model in its map

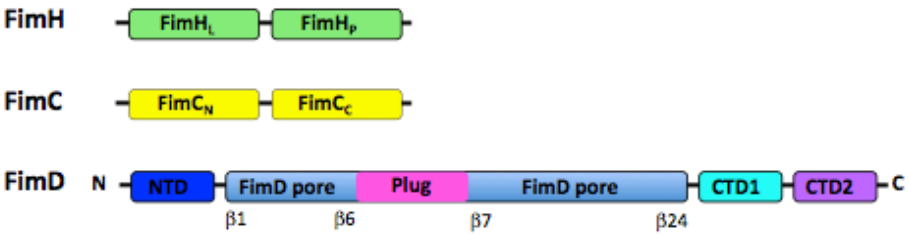
Possible treatment (see also molrep tutorial)

1. Find approximate position:
   Spherically Averaged Phased Translation Function

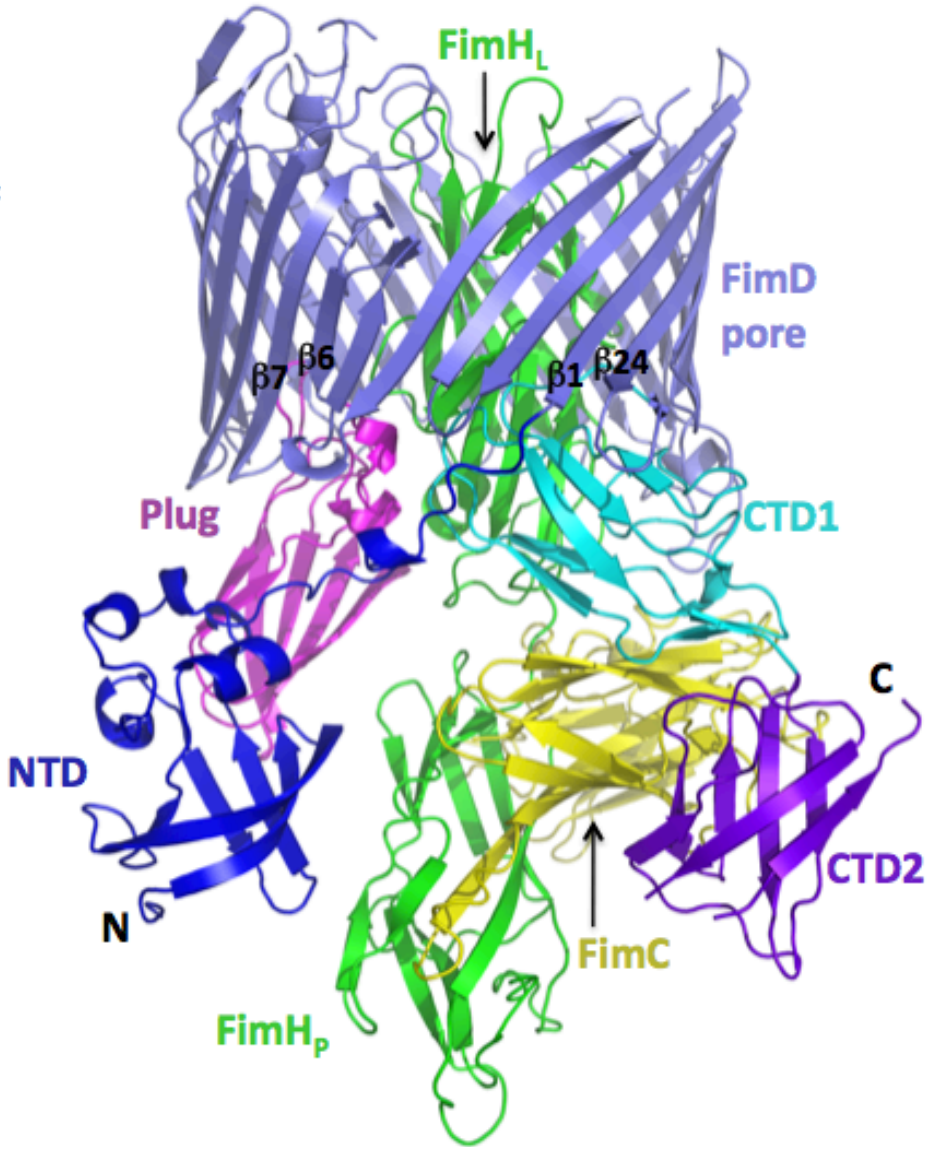
2. Find orientation:
   Local ~~Phased~~ Rotation Function
   - The sphere used in SAPTF is used again, this time as a mask
   - Structure amplitudes from the density in the same sphere

3. Verify and adjust position:
   Phased Translation Function

More complicated example



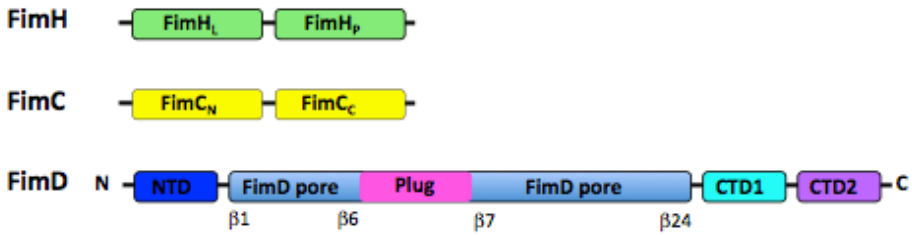
- Asymmetric unit two copies
- Resolution 2.8 Å



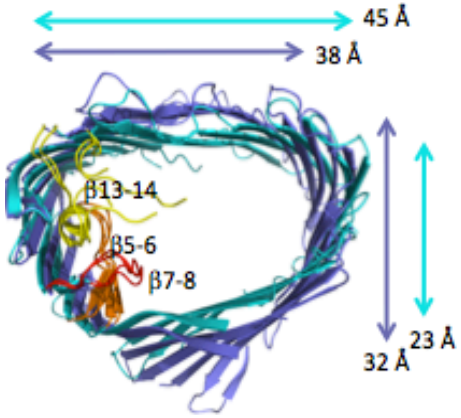
Phane et. al (2011) Nature, 474, 50-53

Usher complex structure solution

1. Conventional MR
   - FimC-N + FimC-C
   - FimH-L + FimH-P
   - FimD-Pore



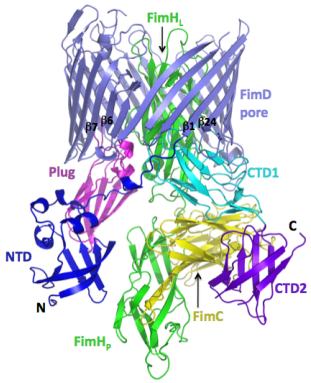
2. Jelly body refinement (Refmac)
   - FimD-Pore

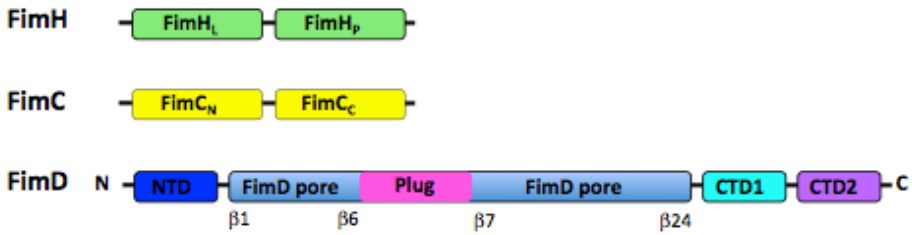


3. <u>Fitting into the electron density</u>
   - FimD-Plug
   - FimD-NTD
   - FimD-CTD-2

4. Manual building
   - FimD-CTD-1

Performance of fitting methods





	search model	sequence identity	"Masked" RF PTF *prf n*	SAPTF PRF PTF *prf y*	SAPTF Local RF PTF *prf s*
FimD-Plug	3fip_A	38.5%	2 (2)	– (–)	1 (2)
FimD-NTD	1ze3_D	100%	2 (2)	1 (2)	2 (2)
FimD-CTD-2	3l48_A	33.3%	– (–)	2 (2)	– (–)

Trying several methods is a good practice (also because of cross-validation)

NCS copy in a special position

False origin solution:

- Is an artifact of Molecular replacement in the presence of pseudotranslation

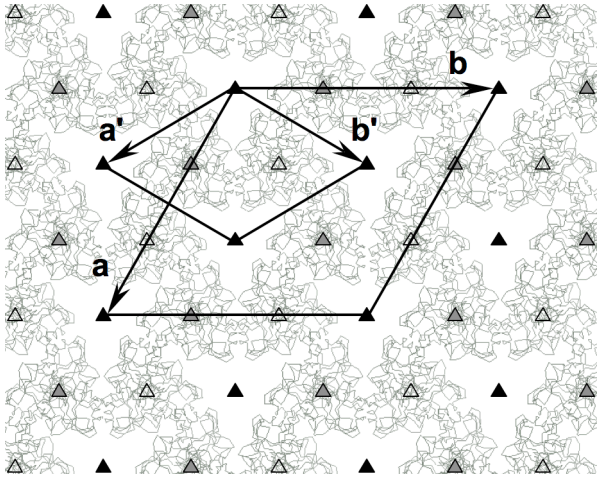
The most recent example:

- Twinning + three alternative origins in a substructure:

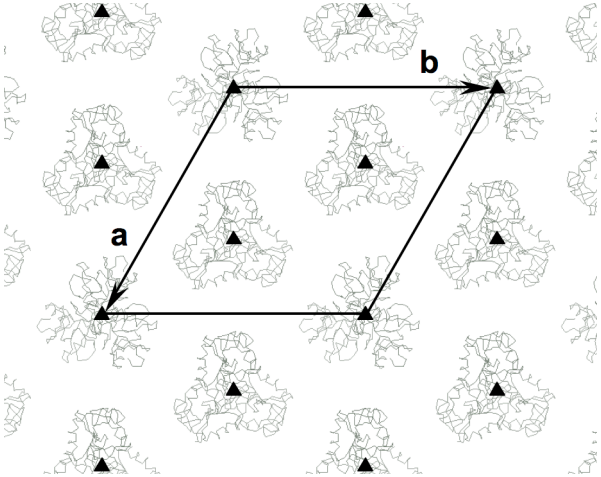
▲ True origins

△ △ False origins

All origins are equivalent in the small cell



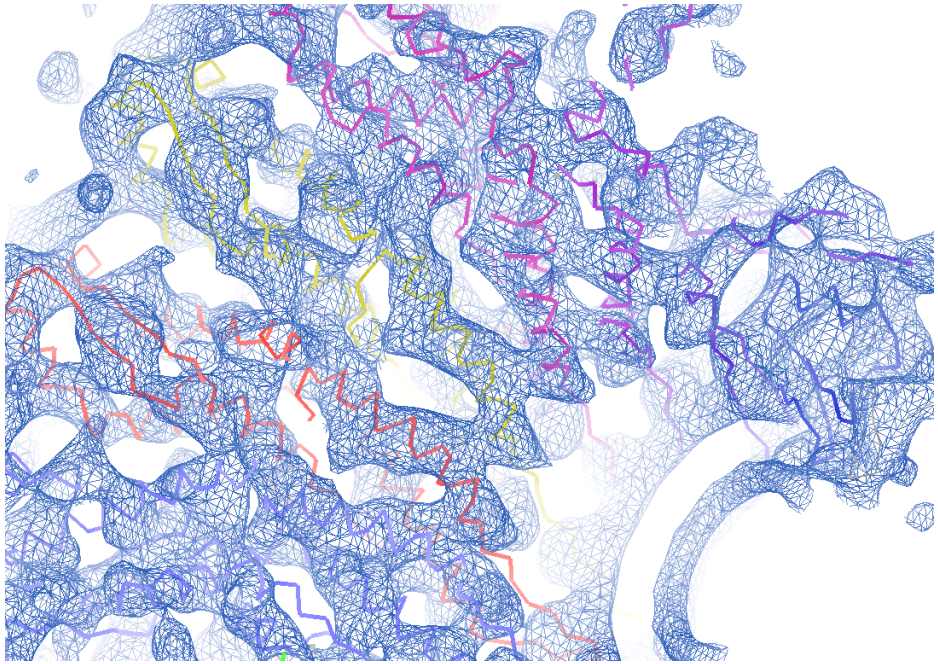
$P3_121$ (a' b' c)

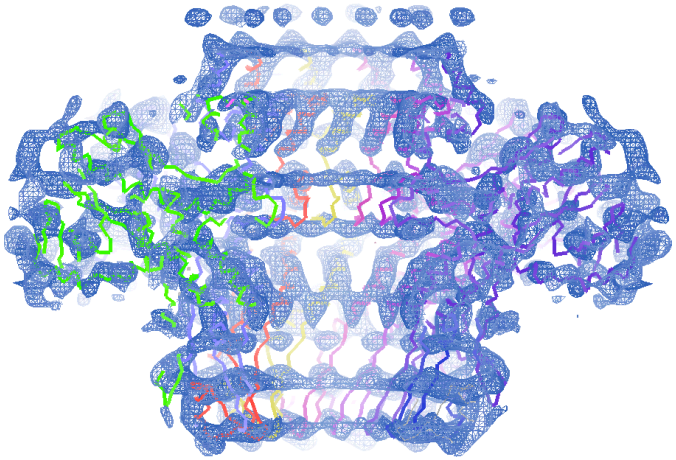


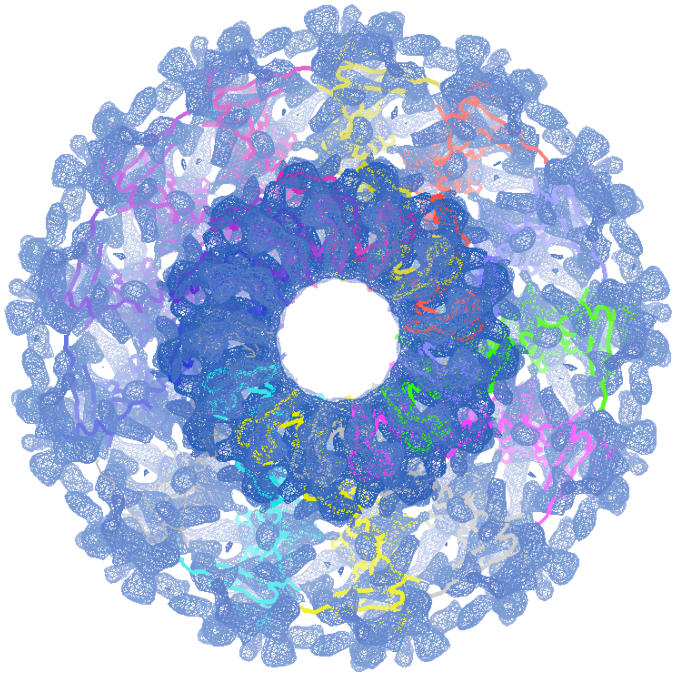
$P3_1$ (a b c)

Watson *et al*. (2011). *JBC*, online pre-publication.

Fitting into EM maps







Balbes

Fei Long
Garib Murshudov

MRC LMB Cambridge

Alexey Vagin

YSBL University of York

Balbes features

Input: mtz-file and sequence

Output: the best solution search models for manual investigation

- Balbes has a database containing preprocessed data from the PDB
  - updated mongthly
  - used for generation of models for given sequence
    - » different types of models
    - » any model is an ensemble if possible

- Can handle complexes
  - several sequences in the input sequence file

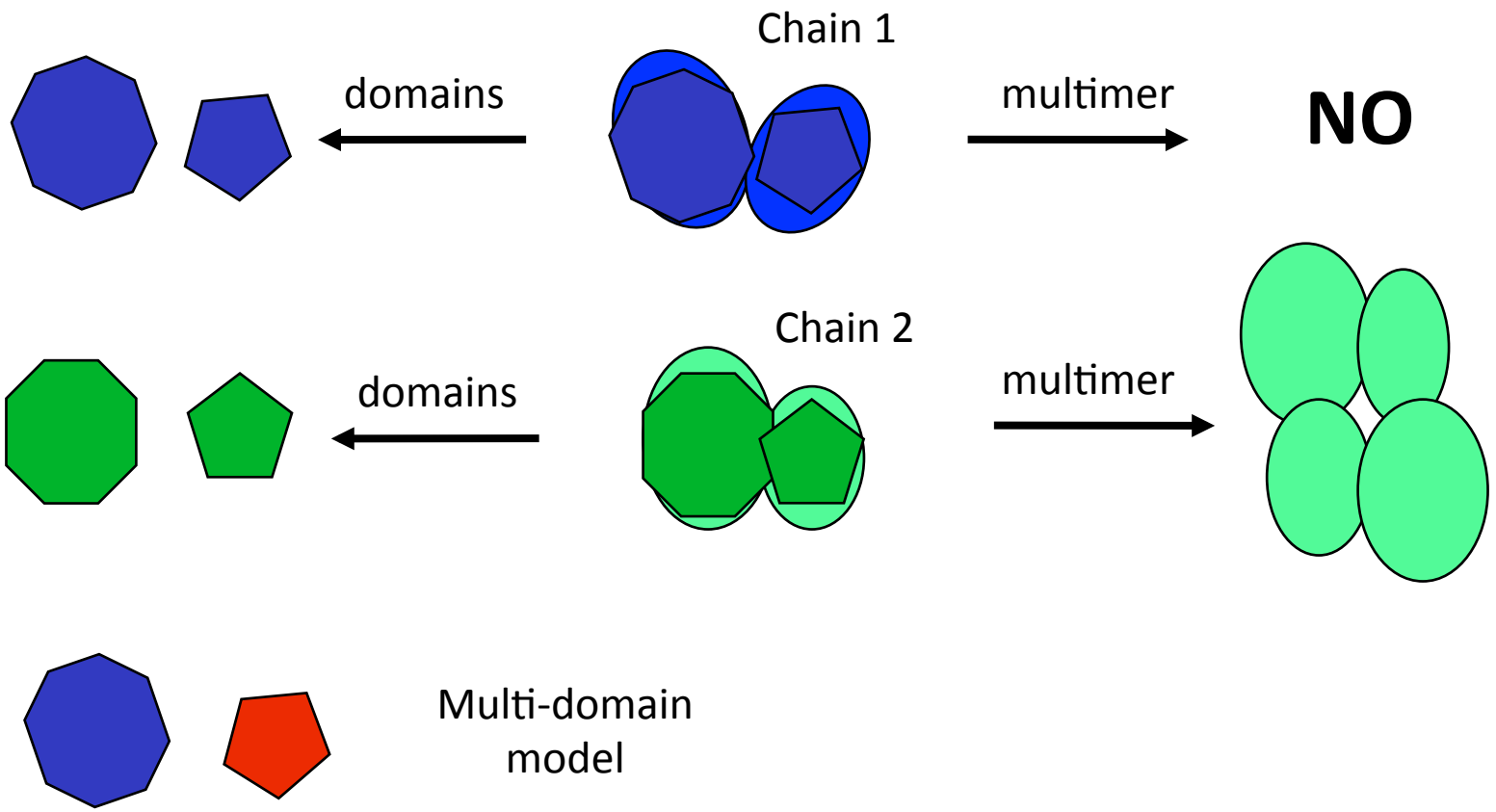
- Strucrure solution using Molrep and Refmac
  - Uses the search in the density for second, third etc component

Database

- Non-redundant chains from the PDB
  - The PDB entries of better resolution are preferred
- More than 30000 domain definitions
  - flexible parts removed
  - hierarchically organized according to 3D similarity
- Multimers
  - generated using PISA definition

Model preparation

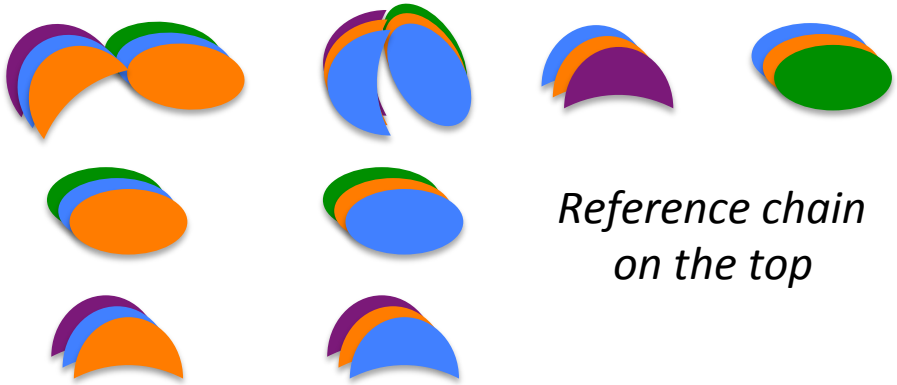
All models are corrected by sequence alignment
and by accessible surface area



Ensemble models from Balbes

Homologues from the Balbes database:

Ensemble search models:



Reference chain on the top

Example:

Ensemble models from Balbes

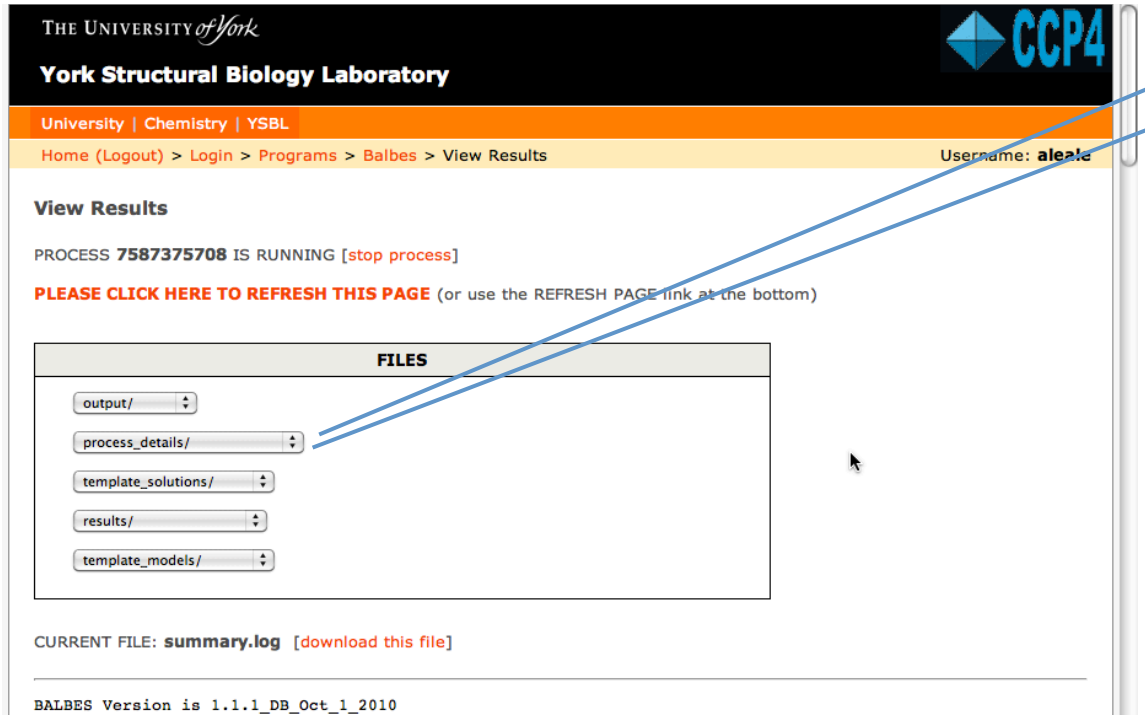
- If ensemble model is possible, it is generated
- Reference chain: occupancy = 1
- Other chains: occupancy < 1
  - » depends on the similarity of this chain with the reference chain

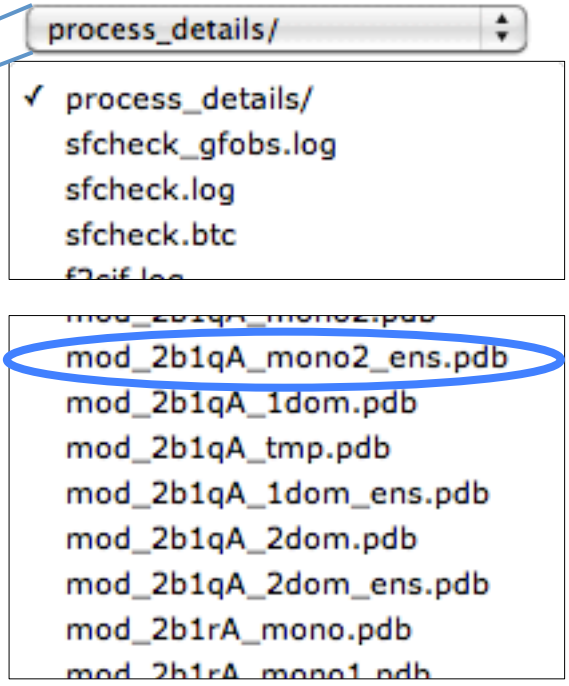
```
MODEL        1     1.000  2b1qA
ATOM      1  N   MET A   1      54.911  30.868  97.738  1.00 28.22 ...
...
MODEL        8     0.000  2b1qA
ATOM      1  N   TYR A   2      62.248  23.657 102.889  0.49 60.31 ...
...
```

- Reference chain is corrected according to the target sequence
- Reference chain is passed to refinement
  - » and the one from the best solution to output

Firstly, try Balbes. If it fails, take generated models and try manual MR

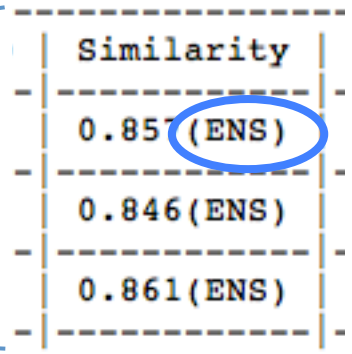
Ensemble models from Balbes



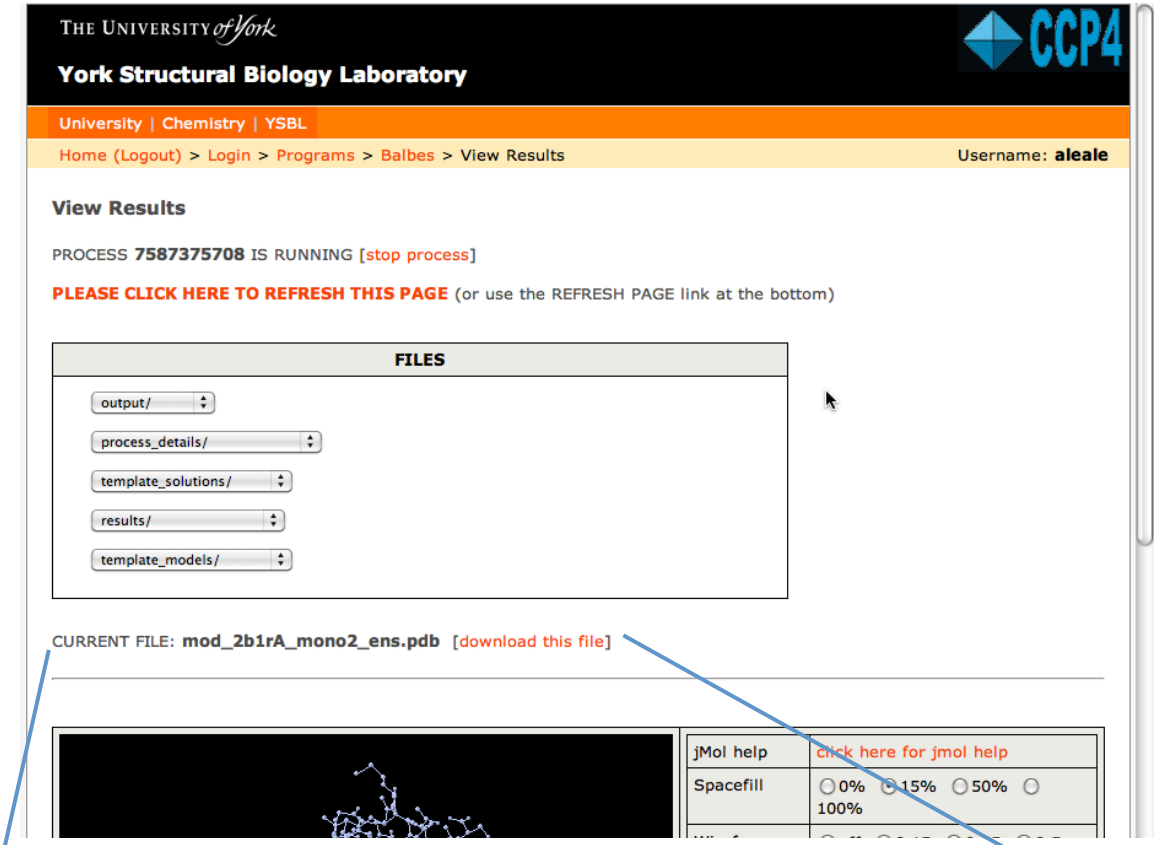


For the structure 1 in sequence 1

PDB_CODE	2b1qA					
NO. of MODEL	3					
Model	Chain ID	Similarity	Residues	Multimer?	Domain?	Monomers(expected)
1	A	0.857(ENS)	244	Monomer	No	1
2	A	0.846(ENS)	162	Monomer	Yes	1
3	A	0.861(ENS)	72	Monomer	Yes	1



Ensemble models from Balbes



CURRENT FILE: **mod_2b1rA_mono2_ens.pdb** [download this file]

All files:

- results > AllFilesIn_7587375708.tar.gz

Acknowledgements

Alexey Vagin	University of York
Fei Long	MRC LMB
Garib Murshudov	MRC LMB