



Collaborative Computational  
Project Number 4  
[www.ccp4.ac.uk](http://www.ccp4.ac.uk)



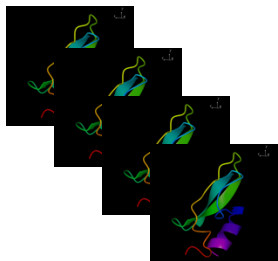
# *MrBUMP – Automated Molecular Replacement*

***Ronan Keegan and Martyn Winn***  
***STFC Daresbury Laboratory***



Science & Technology  
Facilities Council

# The aim of MrBUMP



- An automation framework for Molecular Replacement.
- Particular emphasis on generating a variety of search models.

Wraps **Phaser** and/or **Molrep**.

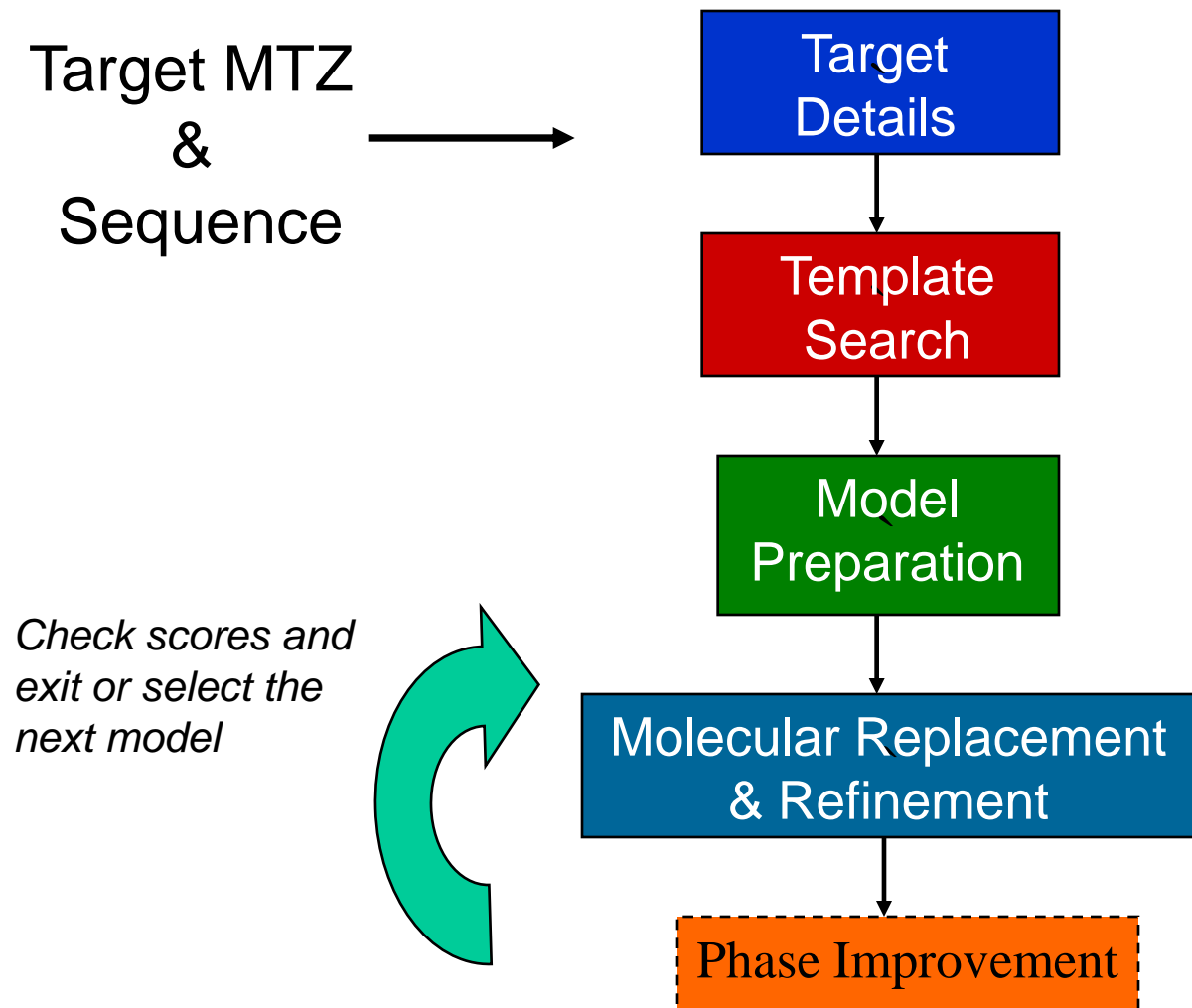
- Also uses a variety of helper applications (e.g. Chainsaw) and bioinformatics tools (e.g. Fasta, Mafft)
- Uses on-line databases (e.g. PDB, Scop)



- In favourable cases, gives “one-button” solution
- In Complicated Cases, will suggest likely search models for manual investigation (lead generation)



# Pipeline



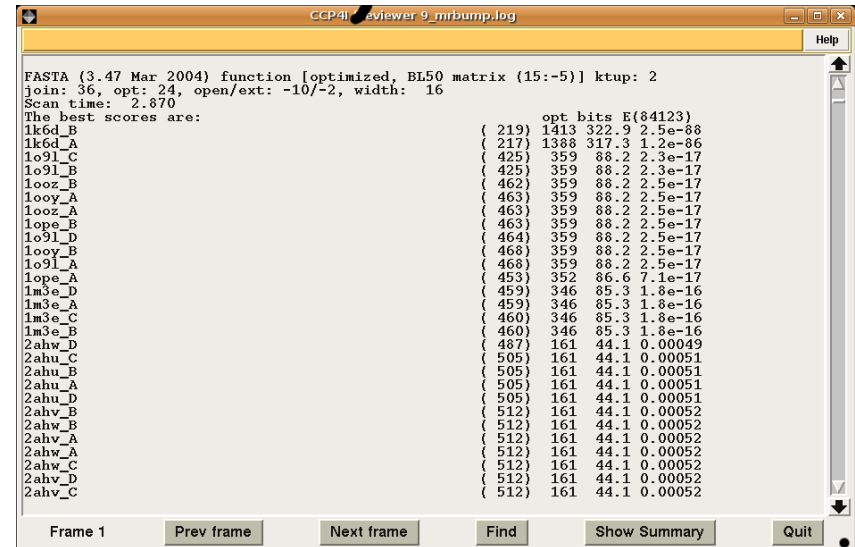
# Search for model templates

FASTA search of PDB

- **Sequence based search** using sequence of target structure

All of the resulting PDB **id codes** are added to a list

These structures are called model **templates**



```
FASTA (3.47 Mar 2004) function [optimized, BL50 matrix {15:-5}] ktup: 2
join: 36, opt: 24, open/ext: -10/-2, width: 16
Scan time: 2.870
The best scores are:
PDB ID      (  )  opt bits E(84123)
1k6d_B      ( 219) 1413 322.9 2.5e-88
1k6d_A      ( 217) 1388 317.3 1.2e-86
1o9l_C      ( 425) 359  88.2 2.3e-17
1o9l_B      ( 425) 359  88.2 2.3e-17
1ooz_B      ( 462) 359  88.2 2.5e-17
1ooz_A      ( 463) 359  88.2 2.5e-17
1ooz_A      ( 463) 359  88.2 2.5e-17
1ope_B      ( 463) 359  88.2 2.5e-17
1o9l_D      ( 464) 359  88.2 2.5e-17
1ooz_B      ( 468) 359  88.2 2.5e-17
1o9l_A      ( 468) 359  88.2 2.5e-17
1ope_A      ( 453) 352  86.6 7.1e-17
1m3e_D      ( 459) 346  85.3 1.8e-16
1m3e_A      ( 459) 346  85.3 1.8e-16
1m3e_C      ( 460) 346  85.3 1.8e-16
1m3e_B      ( 460) 346  85.3 1.8e-16
2ahw_D      ( 487) 161  44.1 0.00049
2ahu_C      ( 505) 161  44.1 0.00051
2ahu_B      ( 505) 161  44.1 0.00051
2ahu_A      ( 505) 161  44.1 0.00051
2ahu_D      ( 505) 161  44.1 0.00051
2ahv_B      ( 512) 161  44.1 0.00052
2ahv_B      ( 512) 161  44.1 0.00052
2ahv_A      ( 512) 161  44.1 0.00052
2ahv_A      ( 512) 161  44.1 0.00052
2ahw_C      ( 512) 161  44.1 0.00052
2ahv_D      ( 512) 161  44.1 0.00052
2ahv_C      ( 512) 161  44.1 0.00052
```

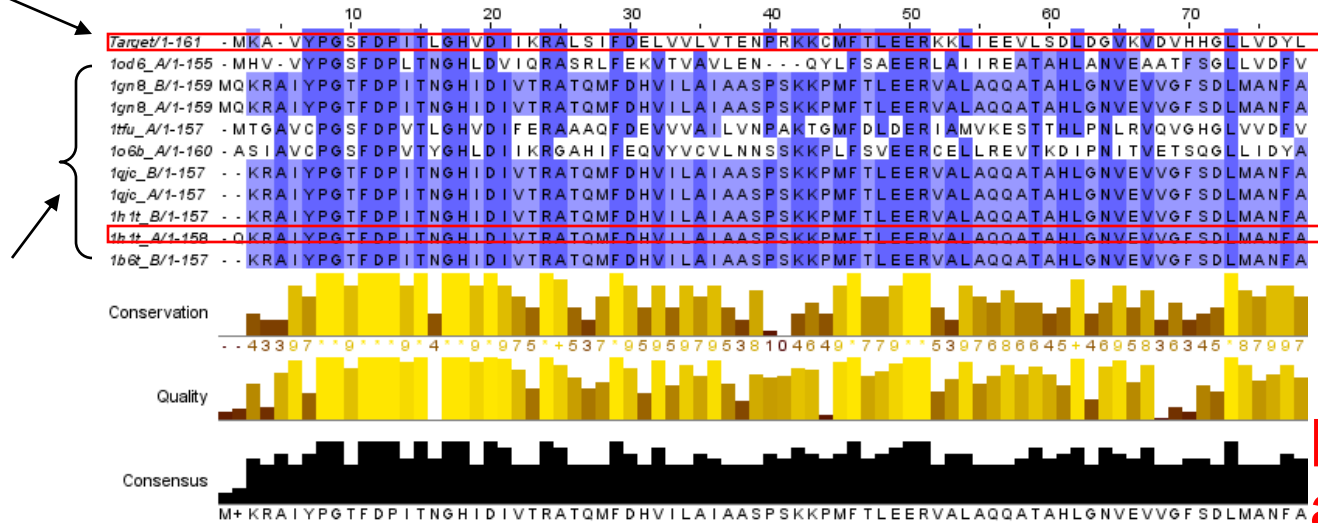
Other templates from:

- SSM search using top hit from the FASTA search
- Can add additional PDB id codes to the list, e.g. from **FFAS** or **psiBLAST** searches
- Can add local PDB files



# Multiple Alignment step

target



Jalview 2.08.1 Barton group, Dundee

pairwise  
alignment  
(used in  
Chainsaw)

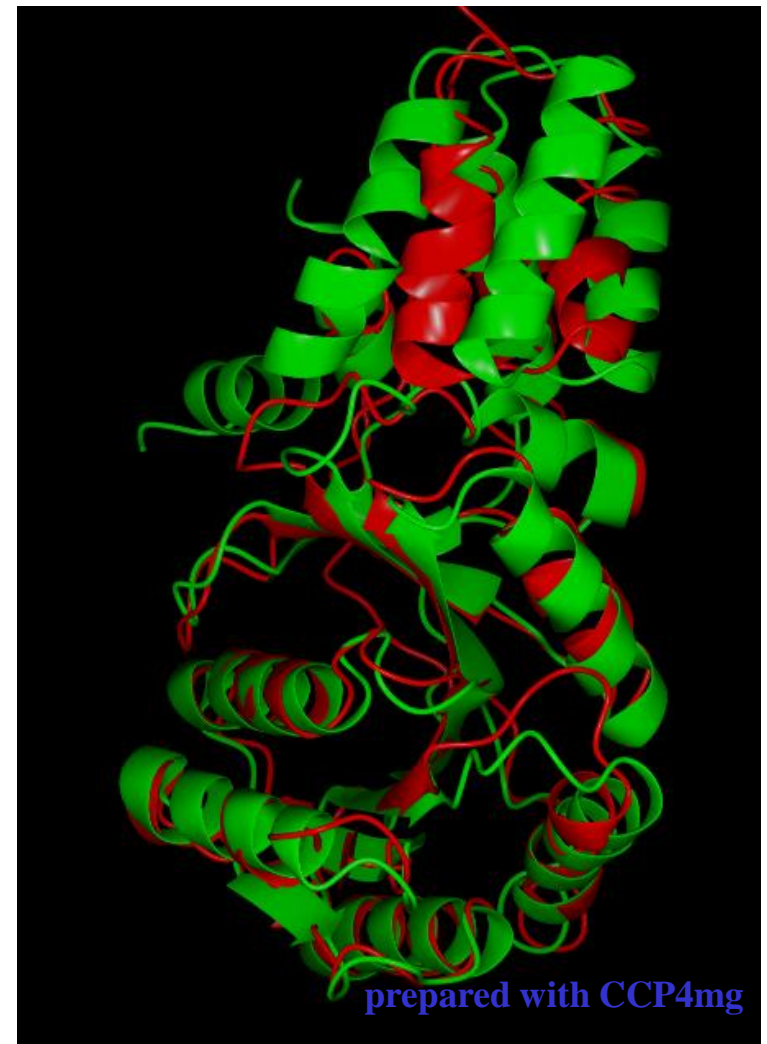
currently support ClustalW, MAFFT, probcons or T-coffee for multiple alignment

Model template scoring: score = sequence identity X alignment quality



# Domains

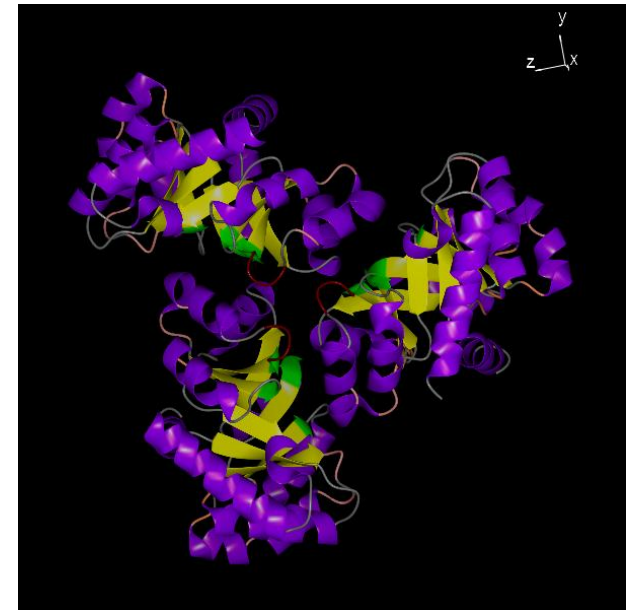
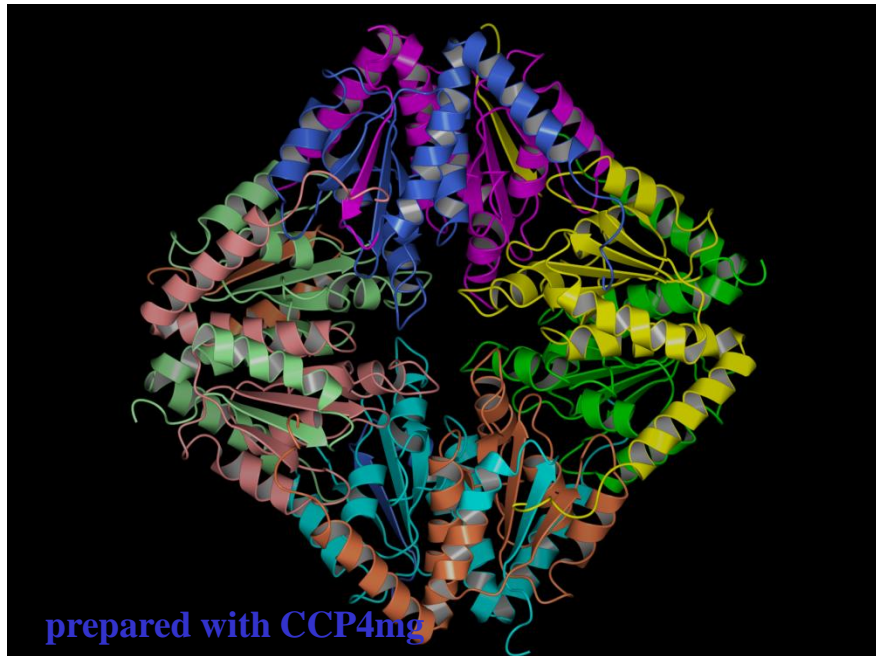
- Suitable templates for target domains may exist in isolation in PDB, or in combination with dissimilar domains
- In case of relative domain motion, may want to solve domains separately
- SCOP database is scanned to see if domains exist for each of the PDBs in the list of templates
- Domains are then extracted from the parent PDB structure file and added to the list of template models as additional search models for MR.





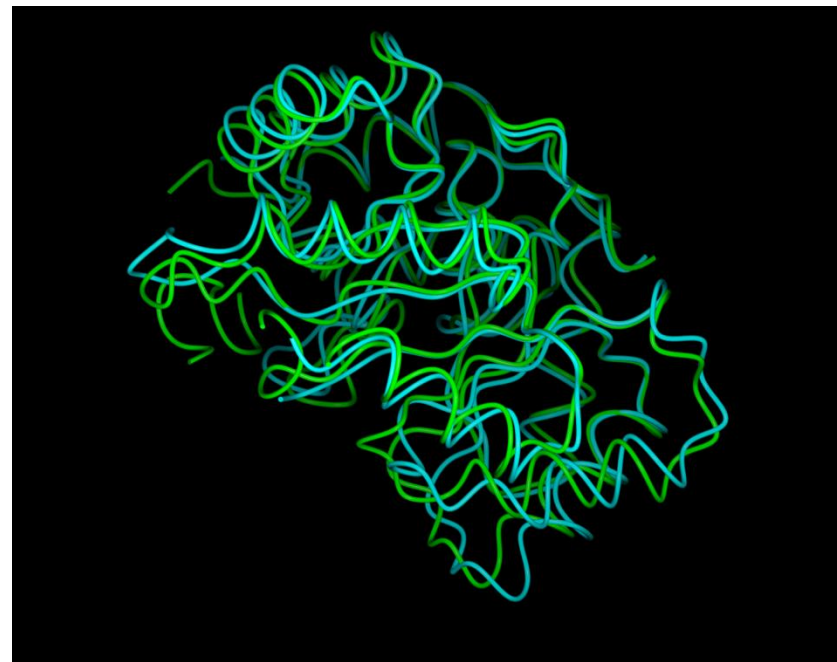
# Multimers

- Use template multimer as model for target multimer (currently uses PQS, will use PISA)
- Better signal-to-noise ratio than monomer, *if* assembly is correct for the target.
- Biologically relevant multimers more likely transferable



# *Ensemble model*

- Create ensemble of top search models, for use in additional run of **Phaser**.
- Models must be sufficiently similar (MW and rmsd)
- **Molrep** can also use ensembles (not yet implemented)





# Search Model Preparation

Search models prepared in four ways:

## PDBclip

- original PDB with waters removed, most probable conformations selected and format tidied (e.g. chain ID added)

## Molrep

- Molrep contains a model preparation function which will align the template sequence with the target sequence and prune the non-conserved side chains accordingly.

## Chainsaw

- Can be given any alignment between the target and template sequences. Non-conserved residues are pruned back to the gamma atom.

## Polyalanine

- Created by excluding all of the side chain atoms beyond the CB atom using the Pdbset program

more side  
chain  
truncation



deal with  
deletions



Also create an ensemble model based on top 5 models



# *Molecular Replacement Step*

## *Running MR*

- For each search model, MR done with **Molrep** or **Phaser** or both.
- MR programs run mostly with defaults
- MrBUMP provides LABIN columns, MW of target, sequence identity of search model, number of copies to search for, number of clashes tolerated
- Allow Molrep / Phaser to set resolution limits and weights

## *MR output*

- MR scores and un-refined models available for later inspection  
⇒ assess quality of solution, extent of model bias
- MrBUMP doesn't use MR scores, but checks for output file with positioned model, and passes to ....



# Testing enantiomorphic spacegroups

- 11 pairs of enantiomorphic spacegroups containing screw axes of opposite handedness, e.g.  $P4_1$  and  $P4_3$ )
- usually both need to be tested in MR
- correct spacegroup indicated by TF and packing

Spacegroup from MTZ file: 'P 31 2 1'

Do MR using enantiomorphic spacegroup as well ☒

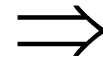
- **MrBUMP** can test both in **Molrep** and/or **Phaser**.
- For each search model, best MR results used to fix spacegroup for subsequent steps.
- Discrimination good for good search model + correct MR solution



# Restrained Refinement Step

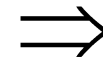
- The resulting models from molecular replacement are passed to **Refmac** for restrained refinement.
- The change in the Rfree value during refinement is used as rough estimate of how good the resulting model is.

final Rfree < 0.35 or  
final Rfree < 0.5 and dropped by 20%



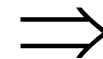
“good”

final Rfree < 0.48 or  
final Rfree < 0.52 and dropped by 5%



“marginal”

otherwise



“poor”

*conservative .....*



# *Phase improvement*

If resolution better than 1.7Å use Acorn procedure:  
initial phase set from refined MR solution  
artificial phase extension to 1.0Å  
dynamic density modification

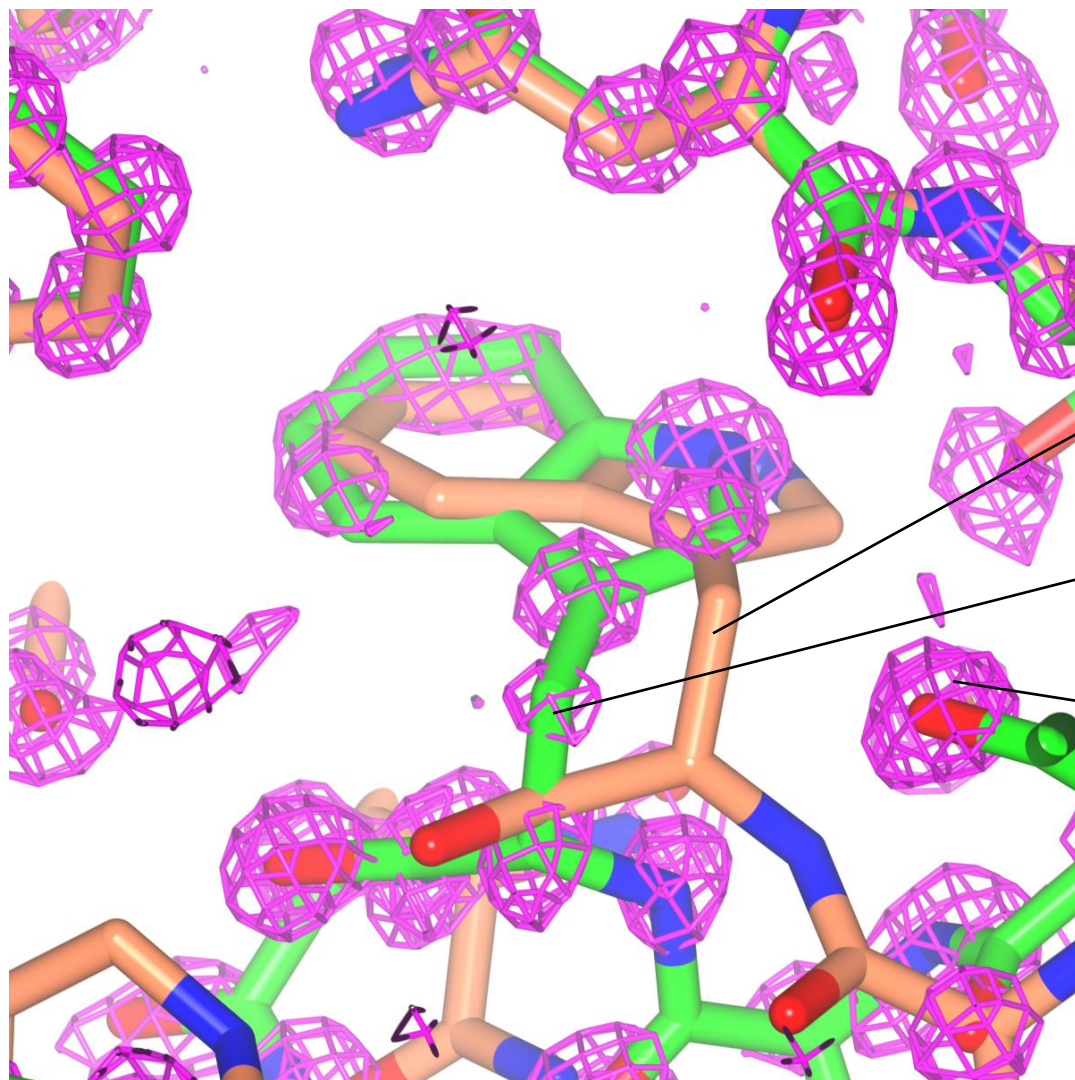
Result:

CC for medium Es good indicator of solution  
Use E-maps for re-building

At lower resolutions:

Use parrot (todo!)





dUTPase from *C.jejuni*  
data to 1.65Å

positioned/refined  
search model

final model (1w2y)

Acorn map (as  
generated by  
MrBUMP)

CC: 0.078 → 0.156  
ARP/wARP re-builds  
into Acorn map

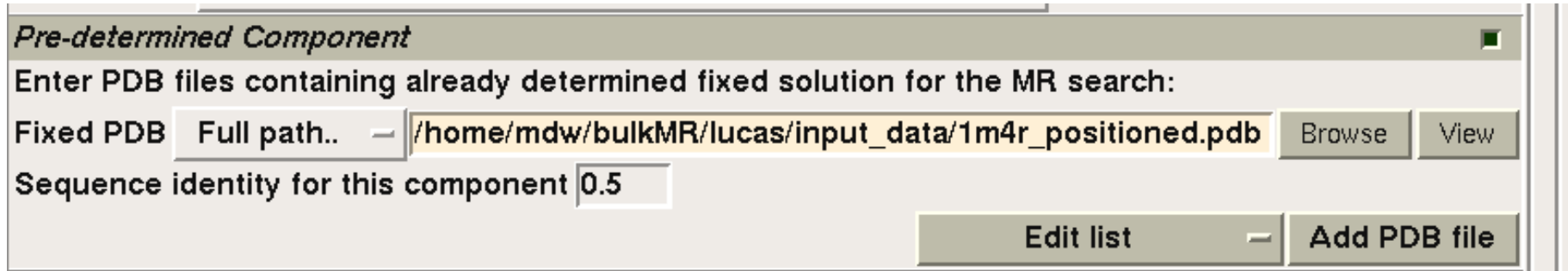
ccp4mg





# *Inclusion of fixed models*

- MrBUMP will now accept one or more positioned models.
- These are included as fixed models in all MR jobs.



The screenshot shows a software window titled "Pre-determined Component" with a green close button in the top right corner. Below the title bar, the text "Enter PDB files containing already determined fixed solution for the MR search:" is displayed. There are two input fields: "Fixed PDB" and "Full path..". The "Full path.." field contains the text "/home/mdw/bulkMR/lucas/input\_data/1m4r\_positioned.pdb". To the right of this field are two buttons: "Browse" and "View". Below these fields is a label "Sequence identity for this component" followed by a text box containing the value "0.5". At the bottom right of the window are two buttons: "Edit list" and "Add PDB file".

- Thus, solve complexes through consecutive runs of MrBUMP.
- Automation of this in progress ....

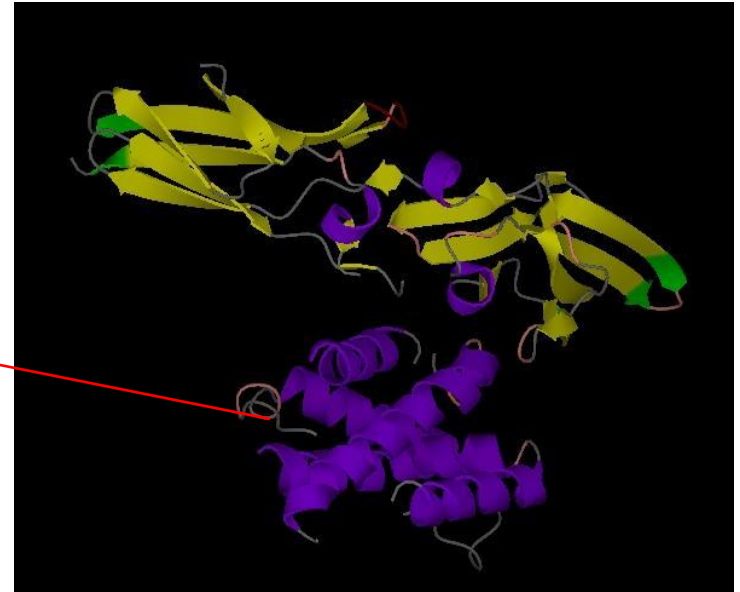
## *Example (with thanks to Lucas Bleicher)*

1:1 complex (1 copy in a.s.u.)  
data to 1.9Å in C2

Small protein (151 res) already  
solved, easy to locate in  
complex.

Larger protein (217 res), 2  
domains, nearest homologs  
around 26%

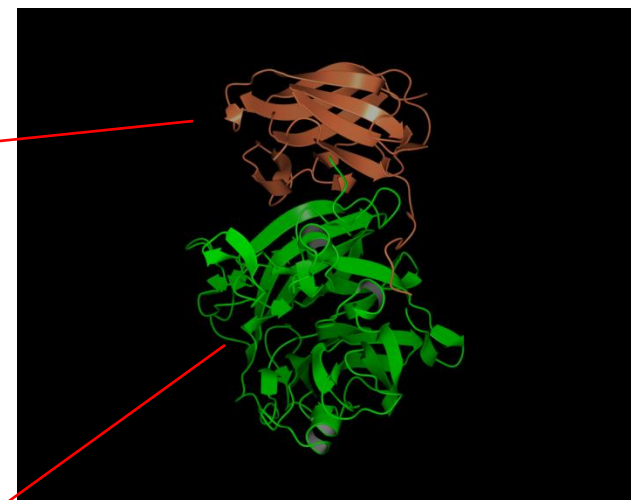
Run MrBUMP with small protein fixed.



# Example (thanks to Elien Vandermarliere)

Target is an arabinofuranosidase  
Data to 1.55Å in P212121

Small C domain (144 res) solved with 34%  
seq ident model  
(1w9t\_B\_MOLREP best out of 4 solutions)



With C domain solution fixed, large N domain (345 res) solved with  
28% seq ident model  
(1gyh\_C\_CHNSAW best out of 7 solutions)

**Not yet solved!**

Acorn: CC increases from 0.04 to 0.18

ARP/wARP then builds 457/493 residues to R/Rfree 0.185/0.225



# MrBUMP in CCP4i

MrBUMP: Automated Model generation and Molecular Replacement

Job title: Example 1nio

Program Mode: Model search and Molecular Replacement

MTZ in: MRB\_INPUT eg3.mtz [Browse] [View]

F: FREE [Browse] [View]

Free-R: FREE [Browse] [View]

Spacegroup from MTZ file: P 21 21 21 *Note that MrBUMP will assume this spacegroup for the MR stage*

SEQ in: MRB\_INPUT eg2.seq [Browse] [View]

MTZ out: MRB\_TEST eg2\_mrbump\_soln1.mtz [Browse] [View]

PDB out: MRB\_TEST eg2\_mrbump\_soln1.pdb [Browse] [View]

Number of molecules in the asu: [Leave blank for automatic calculation]

**Template Search Options**

Multiple alignment program: Mafft [Browse]

E-value for Fasta search: 0.02

☒ Update local copies of search databases

☒ Run the fasta search locally. *Requires fasta34 to be installed*

Search methods to use: ☒ SCOP ☒ PQS ☐ SSM

**Search Model Preparation Options**

Maximum number of search results from which to generate search models for use in MR: 10

Search models to create: ☒ PDBClip ☒ Molrep ☒ Chainsaw ☐ Polyalanine

**Molecular Replacement and Refinement Options**

Molecular Replacement program (first): Phaser [Browse]

Maximum number of prepared models to use in Phaser Ensemble: 5

Number of clashes to tolerate in Phaser: 5

Number of cycles of restrained refinement in Refmac: 30

**Development Options**

Enter PDB id codes to be ignored in the template model search: (e.g. 1nio)

PDB id: 1nio [Browse]

PDB id: 1mom [Browse]

[Edit list] [Add PDB id]

Enter Chain id codes to be included in the template model search: (e.g. 1nio\_A)

Chain id: 1smw\_A [Browse]

Chain id: 1smm\_B [Browse]

[Edit list] [Add Chain id]

☐ Use debug mode. *Gives a more verbose output*

[Run] [Save or Restore] [Close]

- MrBUMP included in CCP4 6.1 series
- Runs on Linux, OSX and Windows.
- Comes with CCP4 GUI .
- Can also be run from the command line with keyword input
- Tutorials available

.0.3 running on rmk65sam Project: mrbump

		Change Project	Help
22 Jan 09	FINISHED	refmac5	test refma
22 Jan 09	FINISHED	refmac5	test refma
22 Jan 09	FINISHED	refmac5	test refma
22 Jan 09	FINISHED	refmac5	test refma
22 Jan 09	FINISHED	refmac5	test refma
05 Jan 09	FAILED	mrbump	felix data
05 Jan 09	FINISHED	mrbump	felix data

Directories&ProjectDir

View Any File

View Files from Job

Search/Sort Database..

Graphical View of Project

Delete/Archive Files..

Kill Job

ReRun Job..

Edit Job Data

Preferences

System Administration

Mail CCP4

Exit



# MrBUMP output

Log file gives summary of models tried and results of MR

- May get several putative solutions
- *Ease of subsequent model re-building, model completion may depend on choice of solution*
- Worth checking “failed” solutions

Top solution available from ccp4i

Detailed results located in: **<ccp4i project directory>/search\_<job number>**

In this directory, there are a number of subdirectories, including:

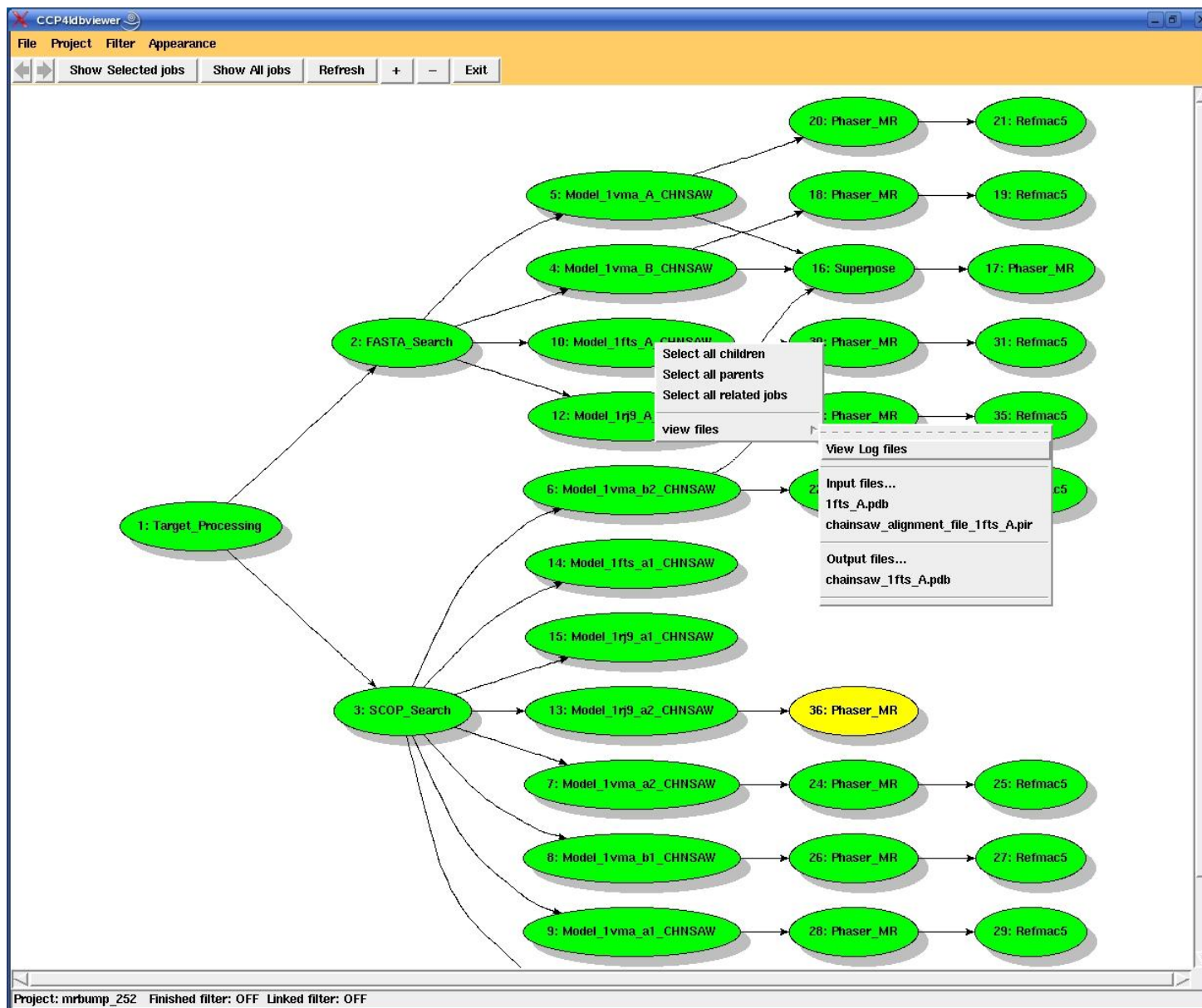
## **data**

Contains the data files and log files from all jobs run. The directory hierarchy is of the form **<template>/<search model>/<pipeline step>**  
e.g. <ccp4i project directory>/search\_55/data/loc0\_A/chainsaw/mr

## **results**

Results from the successful search model are placed into subdirectory "solution". Other results are placed into subdirectory "marginal\_solns".



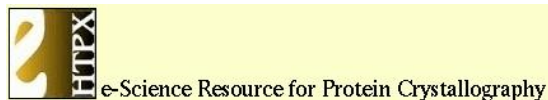




# Acknowledgements

- Martyn Winn CCP4 @ Daresbury
- Thanks to authors of all underlying programs and services  
(see references in MrBUMP log file)
- Other suggestions from:
  - Dave Meredith, Graeme Winter, Daresbury Laboratory.
  - Eugene Krissinel, EBI, Cambridge.
  - Eleanor Dodson, YSBL, York University
  - Geoff Barton, Charlie Bond, University of Dundee
  - Randy Read, Airlie McCoy, Cambridge
- Funding:
  - BBSRC (e-HTPX, CCP4)

<http://www.ccp4.ac.uk/MrBUMP>



Science & Technology  
Facilities Council

