MrBUMP – Automated Molecular Replacement

Ronan Keegan and Martyn Winn
STFC Rutherford Appleton Laboratory & STFC Daresbury Laboratory
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

Target MTZ & Sequence

Target Details

Template Search

Model Preparation

Molecular Replacement & Refinement

Phase Improvement

Check scores and exit or select the next model
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

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

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

Model template scoring: \( \text{score} = \text{sequence identity X alignment quality} \)
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

Prepared with CCP4mg
**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 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

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 Refmac.
Testing enantiomorphically spacegroups

- 11 pairs of enantiomorphically spacegroups containing screw axes of opposite handedness, e.g. P4₁ and P4₃.
- Usually both need to be tested in MR.
- Correct spacegroup indicated by TF and packing.

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 R_free value during refinement is used as a rough estimate of how good the resulting model is.

\[
\begin{align*}
\text{final } R_{\text{free}} &< 0.35 \text{ or } \\
\text{final } R_{\text{free}} &< 0.5 \text{ and dropped by 20%} \\
\text{final } R_{\text{free}} &< 0.48 \text{ or } \\
\text{final } R_{\text{free}} &< 0.52 \text{ and dropped by 5%} \\
\text{otherwise} &
\end{align*}
\]

\[\Rightarrow \quad \text{“good”} \]

\[\Rightarrow \quad \text{“marginal”} \]

\[\Rightarrow \quad \text{“poor”} \]

*conservative*  

---

[Logo: CCP4]
**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
**Inclusion of fixed models**

- MrBUMP will now accept one or more positioned models.
- These are included as fixed models in all MR jobs.
- Thus, solve complexes through consecutive runs of MrBUMP.
- Automation of this in progress ....
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)

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 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
MrBUMP Output

- Log file gives summary of models tried and results of MR
- May get several putative solutions
- Ease of subsequent model rebuilding, model completion may depend on choice of solution
- Worth checking “failed” solutions
- Top solution available from ccp4i
Output files

- 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.
  - Eleanor Dodson, YSBL, York University
  - Geoff Barton, Charlie Bond, University of Dundee
  - Randy Read, Airlie McCoy, Gabor Bunkozci, Cambridge
- Funding:
  - BBSRC (e-HTPX, CCP4)

http://www ccp4 ac uk/ MrBuMP