Automated Molecular Replacement with MrBUMP

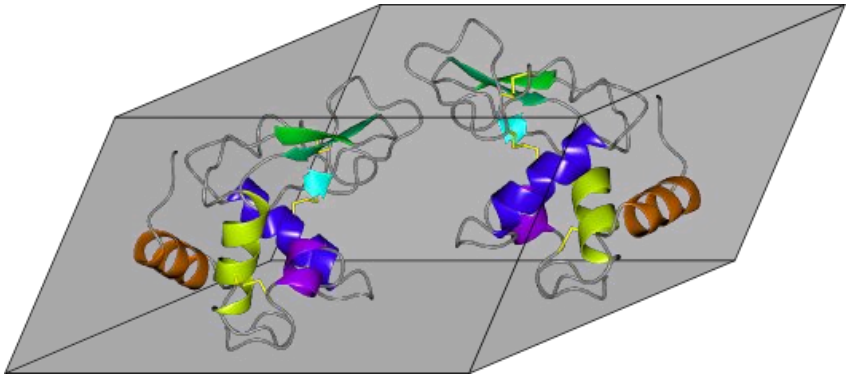
Ronan Keegan

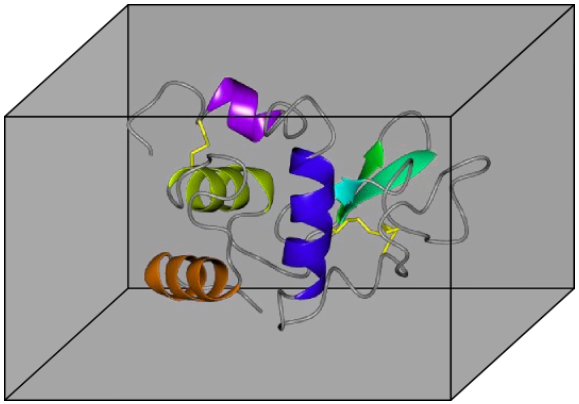
STFC Rutherford Appleton Laboratory

ronan.keegan@stfc.ac.uk

Molecular Replacement: The problem

(Slides courtesy of Gabor Bunkoczi)





Previous crystal form → Current crystal form

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

Complications

Differences between target and search model:

- Deviations in the main chain
- Un-matched Loops
- Different side chains
- Different side chain conformations

Problems stemming from the experimental data:

- Incomplete data
- Anistropy
- Twinning
- Pseudo symmetry

Automation

- Easy cases:
  - provide a one button solution
  - find the best solution and subsequently best starting point for refinement and model completion
- Difficult cases:
  - finding a solution can involve much trial and error
  - success or failure in MR can be highly sensitive to small changes in the search model
  - try many permutations for search model selection, preparation and molecular replacement
- Encapsulate complex procedures such as multiple search model creation and the exploitation of *ab initio* modelling

Ab initio modelling and MR

Deriving structure models based solely on their amino-acid sequence

Has become increasingly successful in recent years and generated models have been shown to work well as molecular replacement search models in the case of smaller proteins (up to 120 residues)

Can be a useful approach where no obvious homologue is available to use as a search model

Can also be used to re-model positioned search models from a likely but poor MR solution and improve the quality of derived phases

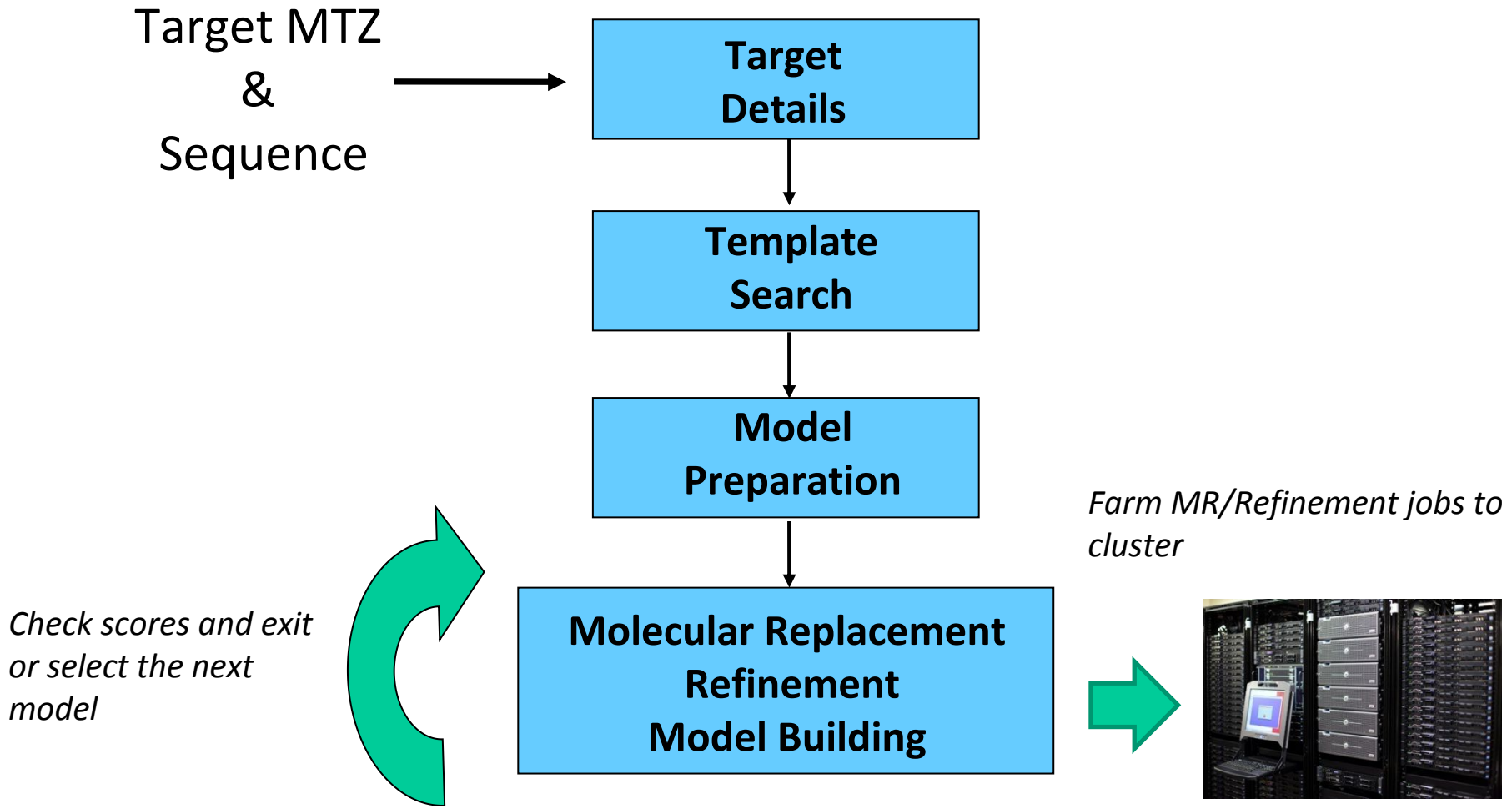
Cluster of aligned ab initio-generated decoy models

MR Pipelines in CCP4

- MrBUMP & BALBES
  - Automated model search and preparation, molecular replacement, refinement and model building
- AMPLE
  - Using ab initio modeling to generate search models for molecular replacement

MrBUMP – Automated Molecular Replacement

MrBUMP Pipeline

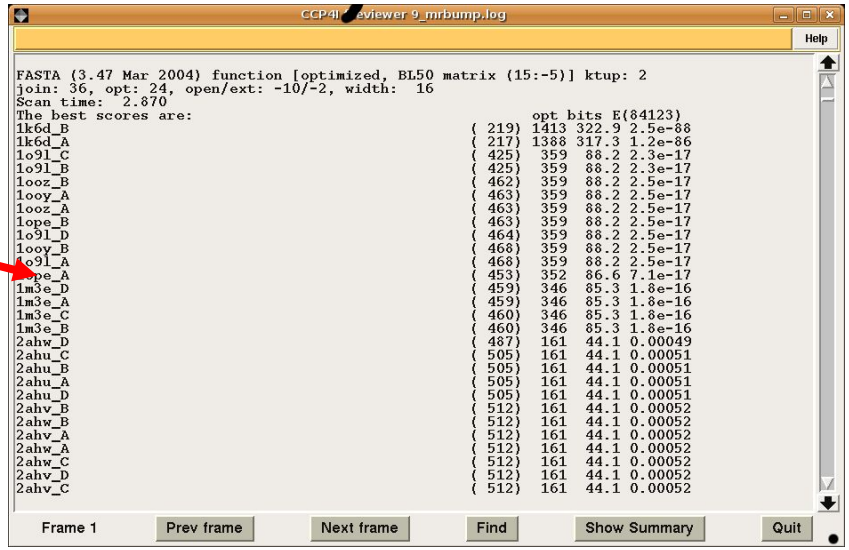


Search for model templates

Phmmer/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:

Can add additional PDB id codes to the list, e.g. from external HHPred or psiBLAST searches

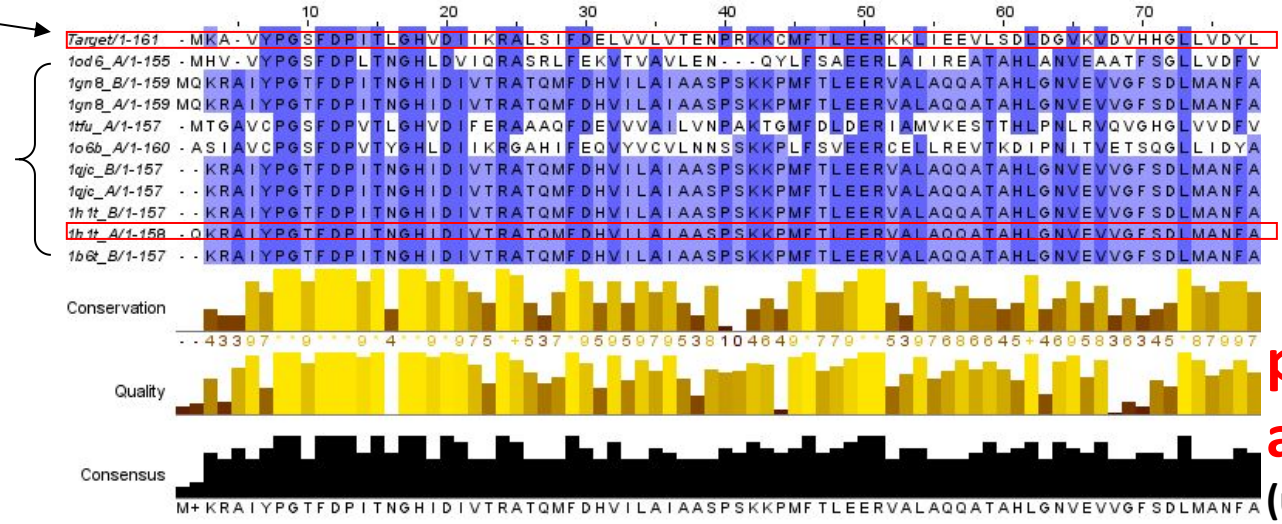
New version calls HHPred directly to do search

Can also add local PDB files

Multiple Alignment step

All model templates need to be aligned for scoring and later model preparation step

target



pairwise alignment **(used later in Chainsaw & Sculptor)**

Jalview 2.08.1 Barton group, Dundee

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

Model template scoring: score = sequence identity X alignment quality

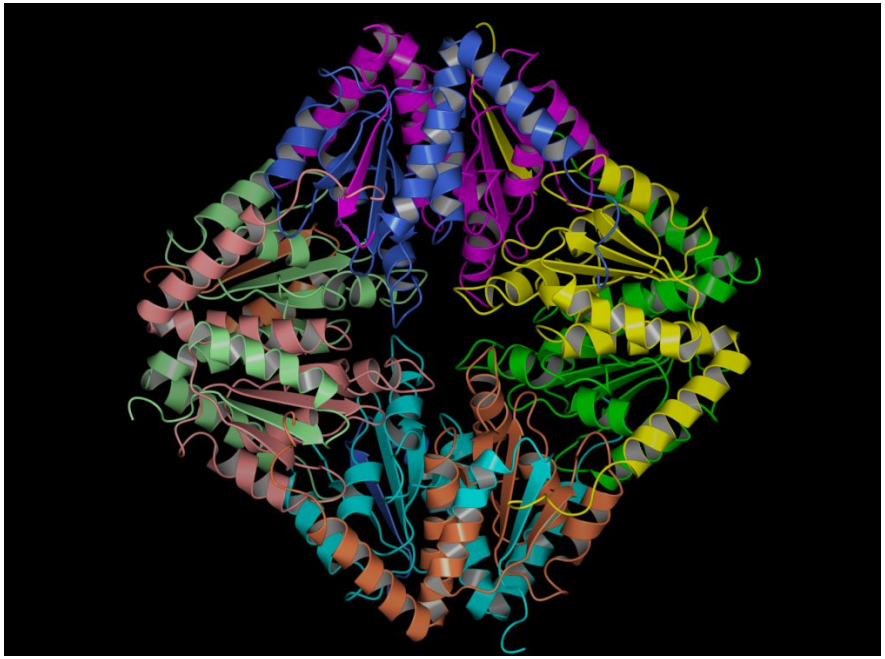
Where *HHPred* is used scoring and alignment come from *HHPred* search results

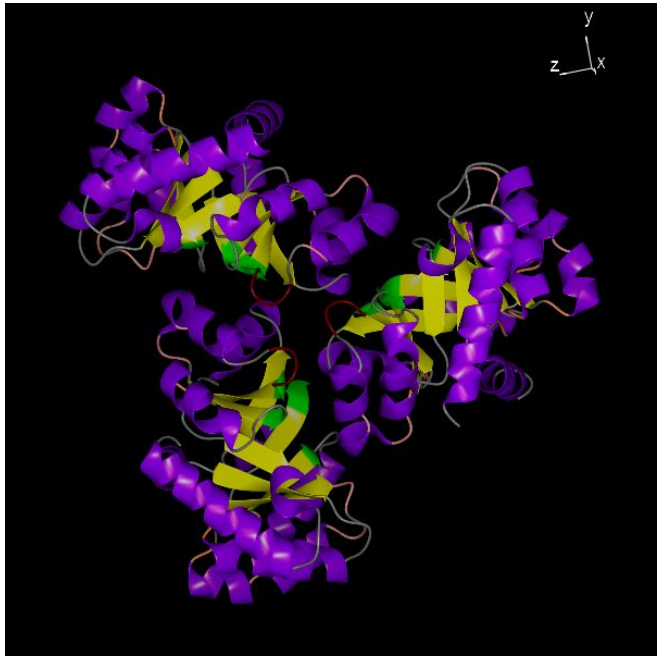
Domains

- Suitable templates for target domains may exist in isolation in the 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 (to be replaced with BALBES database)
- 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 but will also be replaced by BALBES database)
- Better signal-to-noise ratio than monomer, ***if*** assembly is correct for the target.
- Biologically relevant multimers more likely transferable
- Can be determined manually using PISA





Search Model Preparation

Search models are prepared in six different ways:

more side chain truncation (*uses alignments from multiple alignment step or from HHPred*)

Unmod
– Input model template is not modified

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.

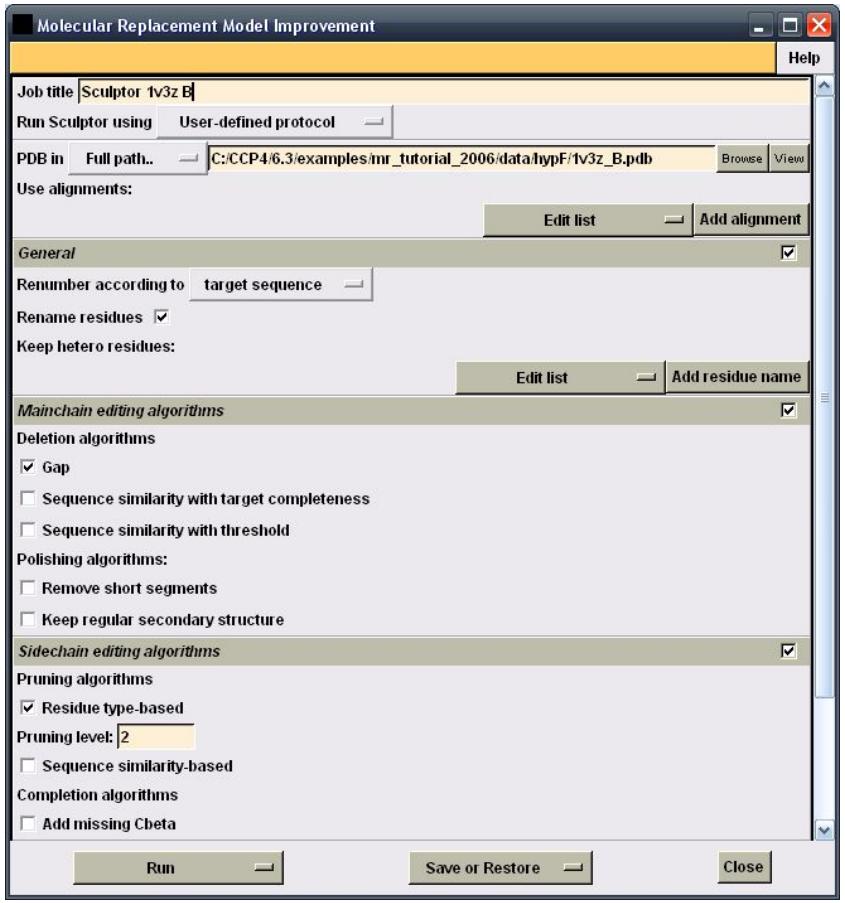
Sculptor
– Similar to chainsaw but different protocols for side chain truncation and alignment

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

deal with deletions

Phaser.Sculptor

- Multi-protocol search model preparation
- Allows for fine-grain control of the parameters used to decide on search model truncation
- Takes into account
  - Surface accessibility
  - Weighted sequence alignment



Ensemble models

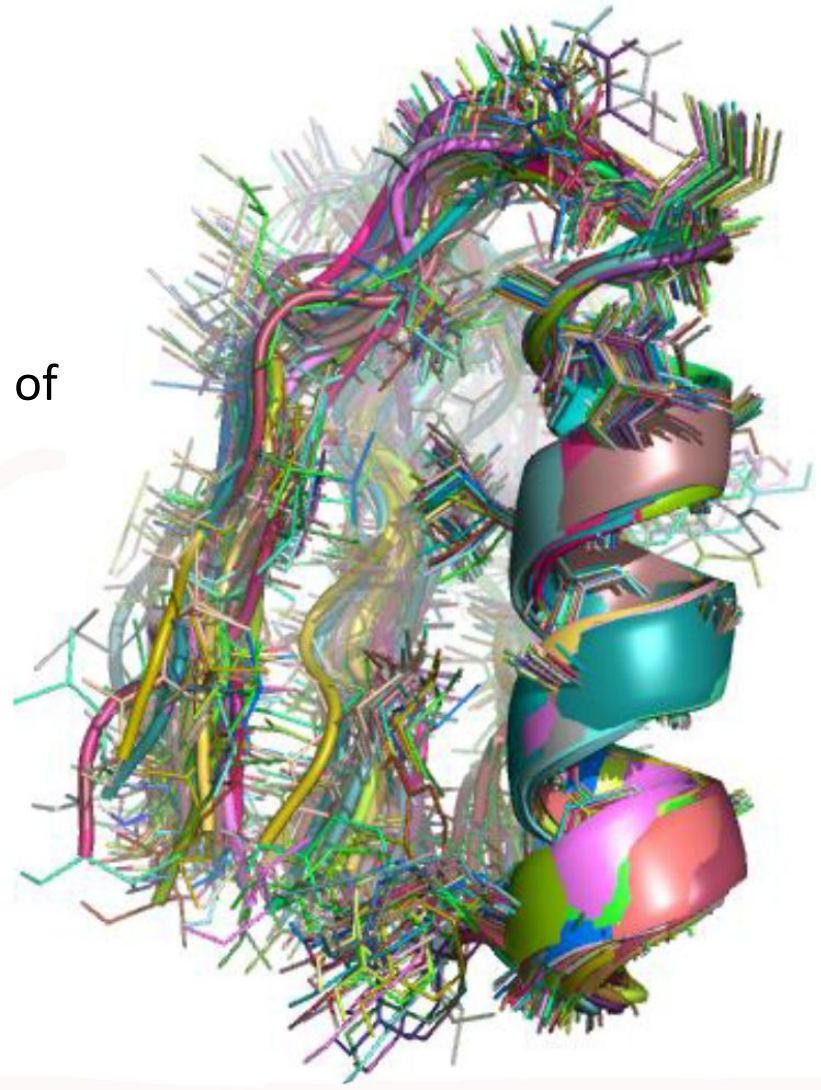
Create ensembles of top search models, for use in additional runs of Phaser and Molrep

MR programs will weight up regions of low alignment variance and down weight regions with more variability

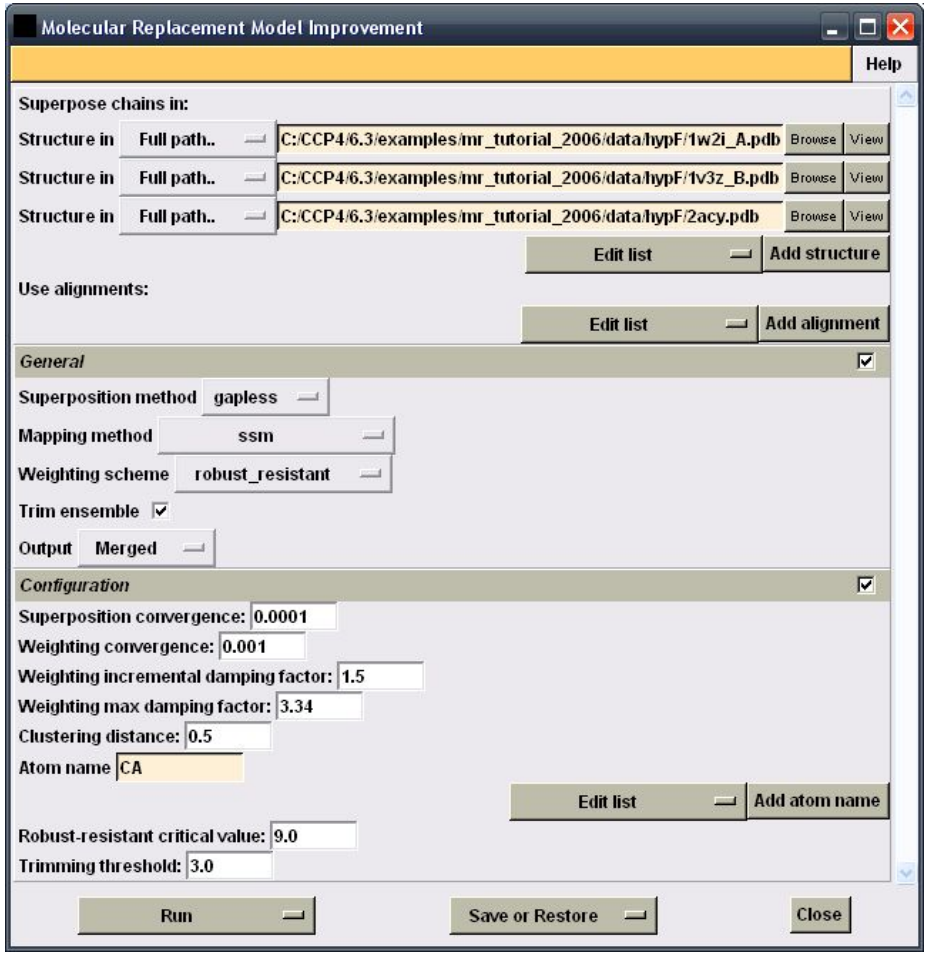
Can give solutions in cases where individual models fail

Models must be sufficiently similar (MW and rmsd)

Use Gesamt or Superpose to align structures



Phaser.Ensembler



- Generate ensembles from set of similar homologues
- Several protocols:
  - Gapped or non-gapped
  - Different alignment methods including manual input of alignment
  - Trimming based on RMSD threshold

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

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"

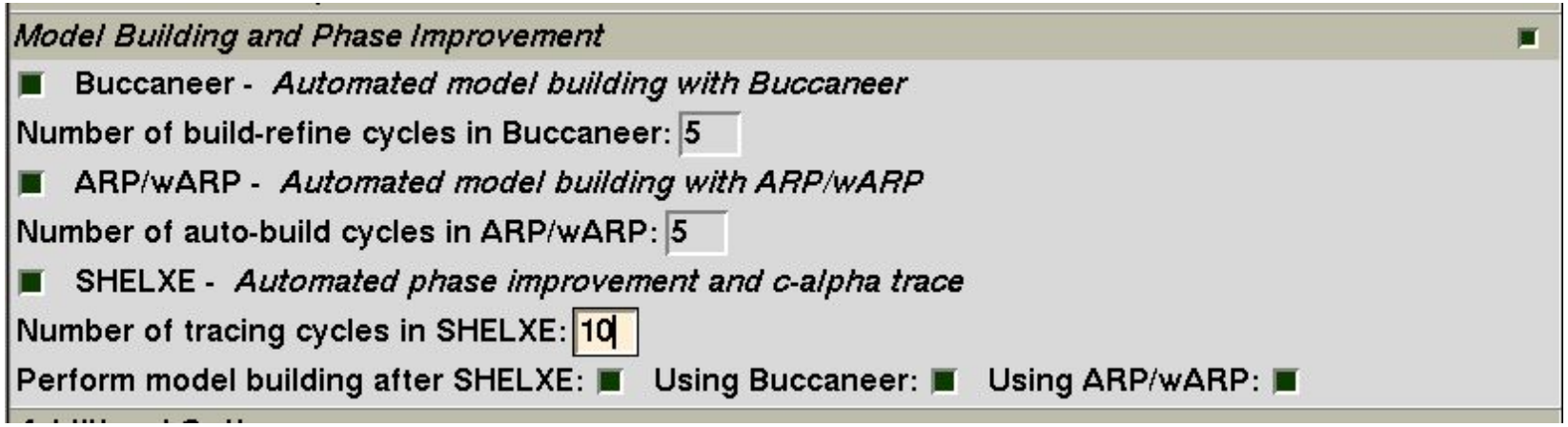
conservative .....

Model Building

Model building programs can be invoked post-MR to give a better assessment of whether or not the MR solution is correct

Options:

- Buccaneer
- ARP/wARP
- SHELXE
- SHELXE can also be combined with Buccaneer/ARPw/ARP to bring model closer to completion



Assessing Solutions with SHELXE

- Attempt to rebuild the structures
  - SHELXE: partial CC of >25% & average fragment length of 10 or more
  - Further confirmation provided through building with ARP/wARP and Buccaneer.

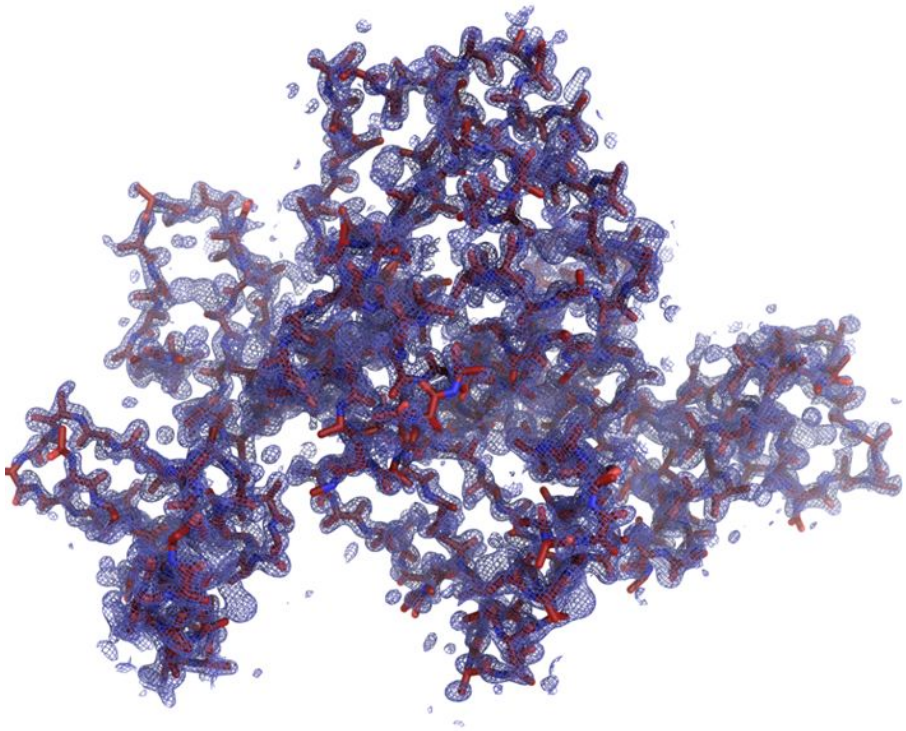
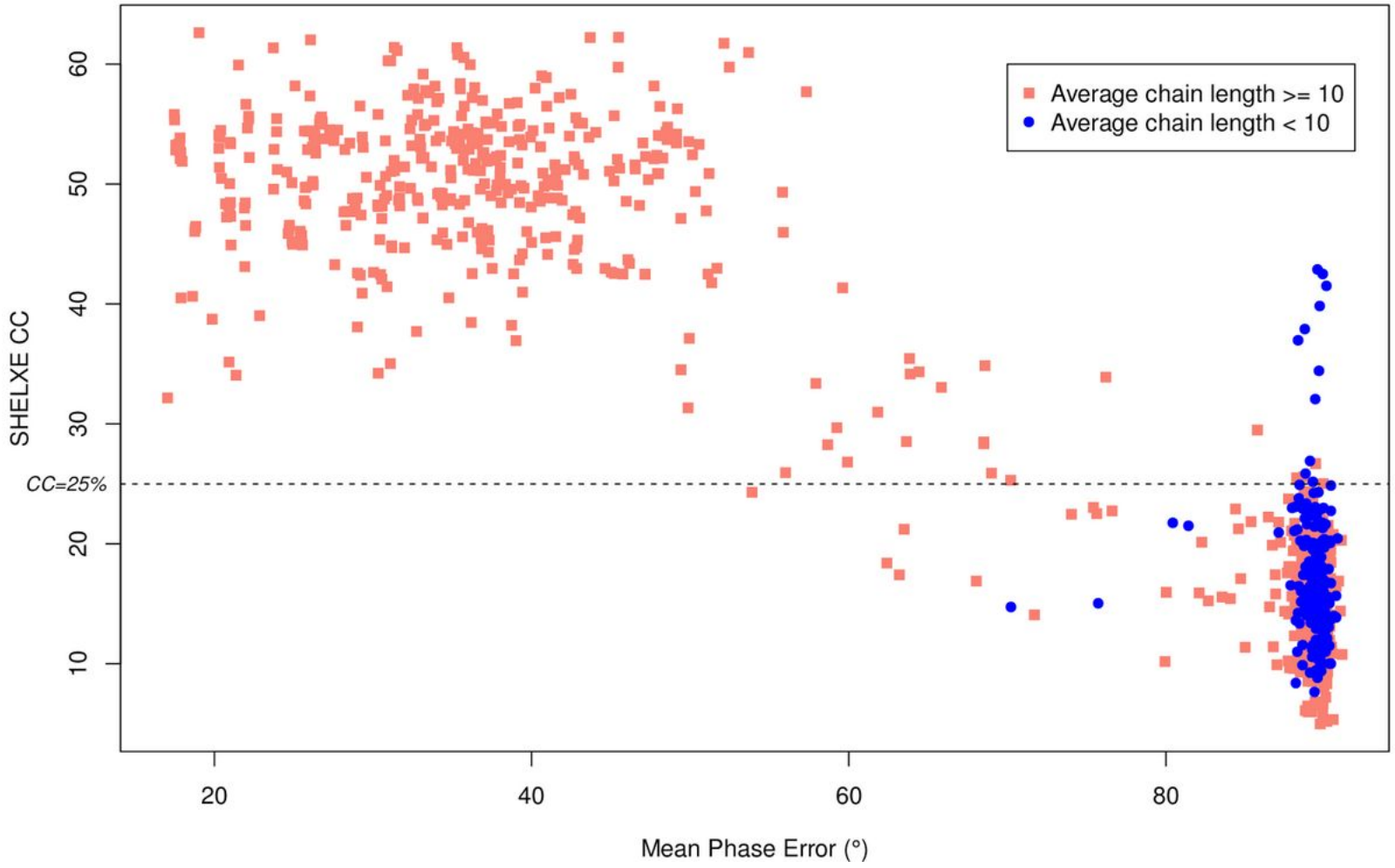


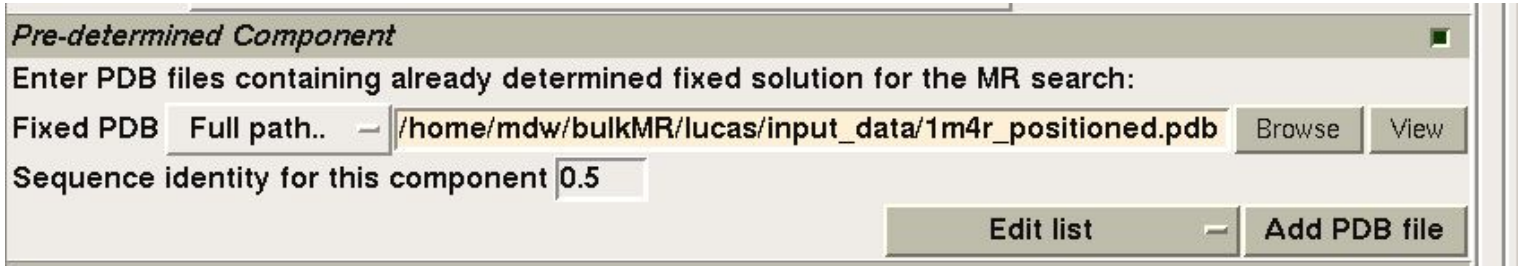
Figure courtesy of Andrea Thorn

SHELXE CC vs. Mean Phase Error



Inclusion of fixed models

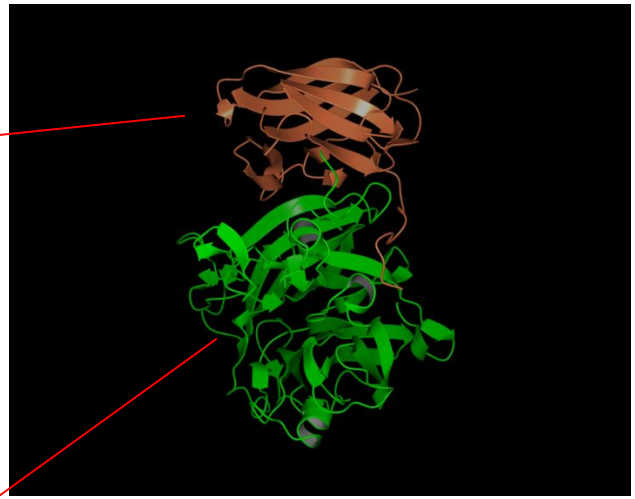
MrBUMP will accept one or more positioned models.
These are included as fixed models in all MR jobs.



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

Example (thanks to Elien Vandermarliere)

Target is an arabinofuranosidase
Data to 1.55Å in P212121



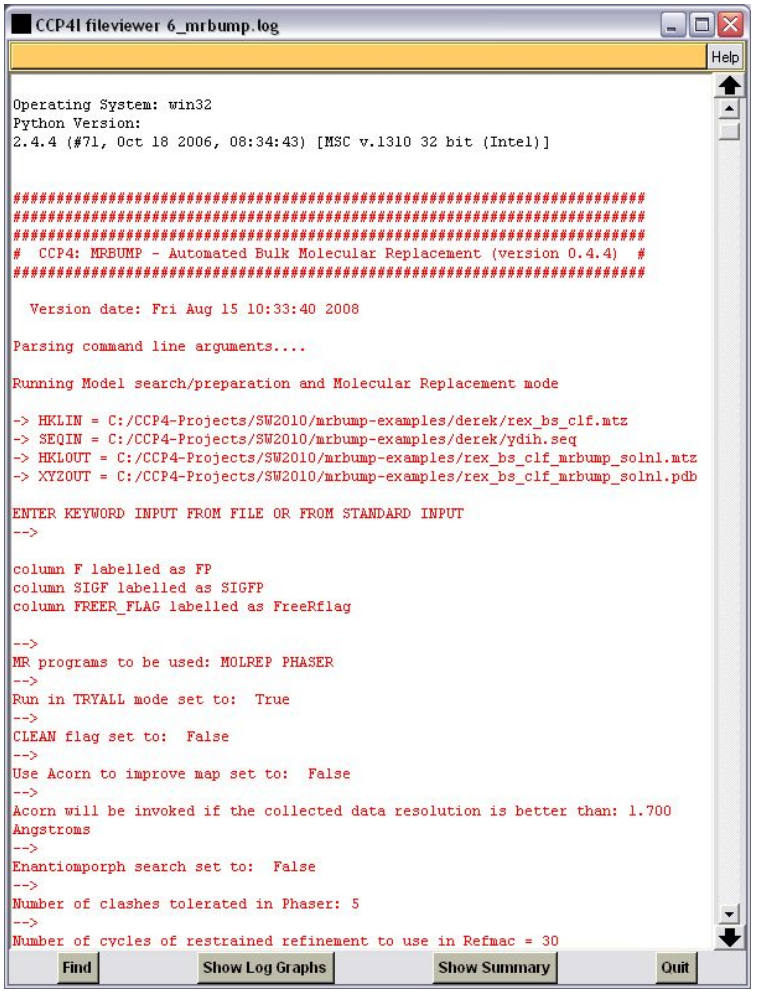
Small C domain (144 res) solved with 34% seq ident model
(1w9t_B_MOLREP best out of 4 solutions)

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

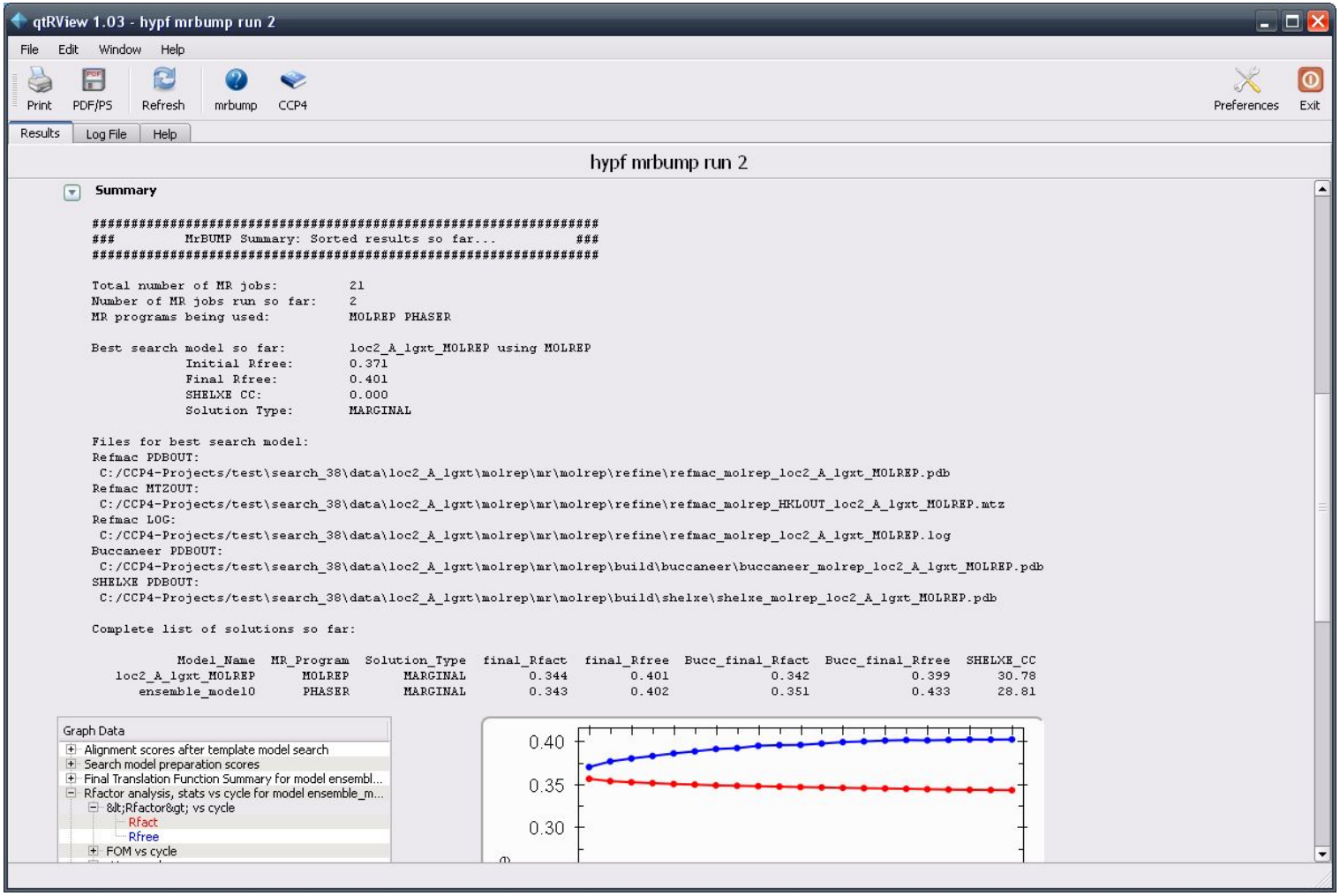
ARP/wARP then builds 457/493 residues to R/Rfree 0.185/0.225

MrBUMP output…

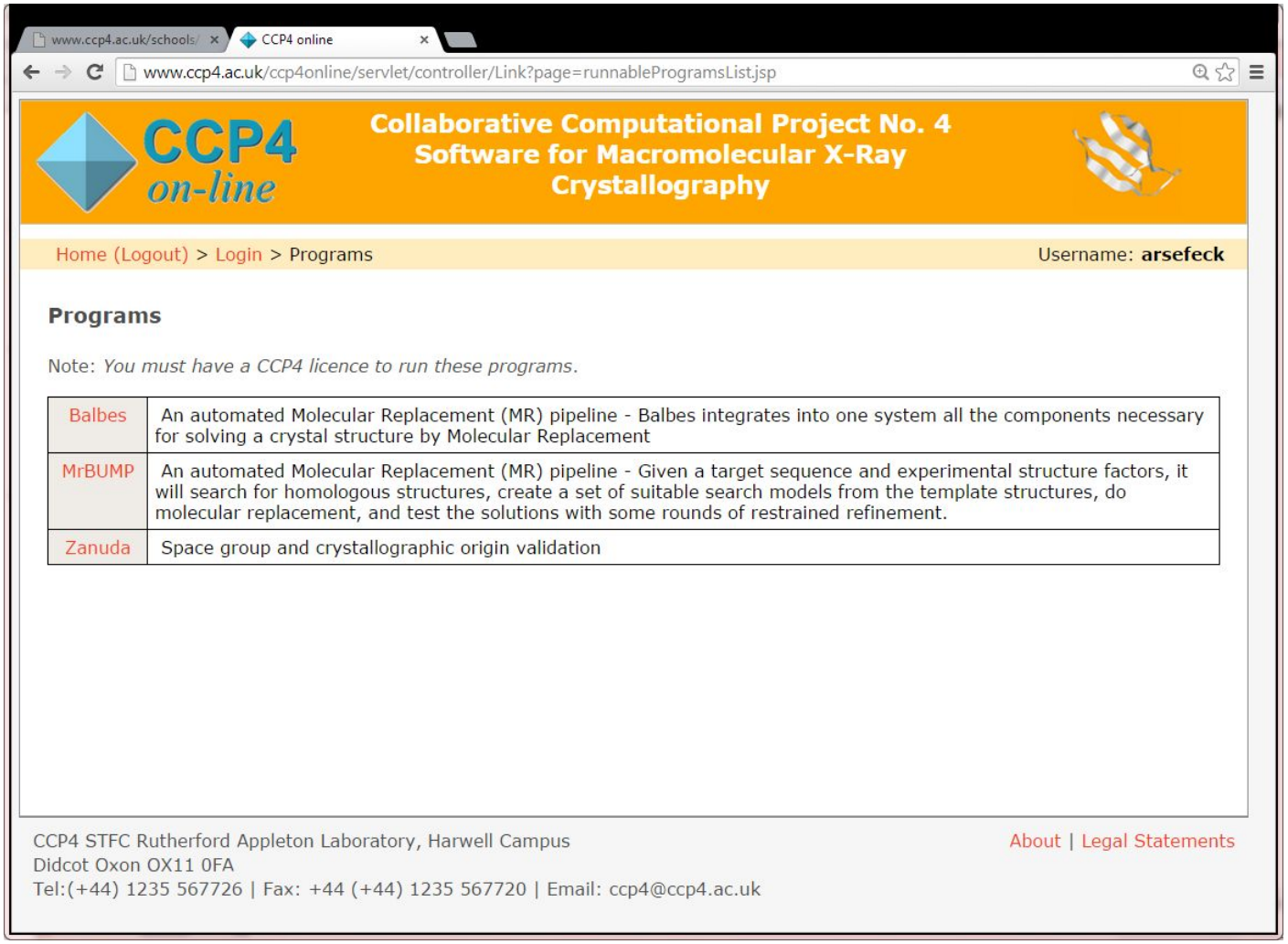
- Log file gives summary of models tried and results of MR
- May get several putative solutions
- Ease of subsequent model re-building, model completion may depend on choice of solution
- Worth checking “poor” solutions
- Top solution available from ccp4i



MrBUMP output - QTRview

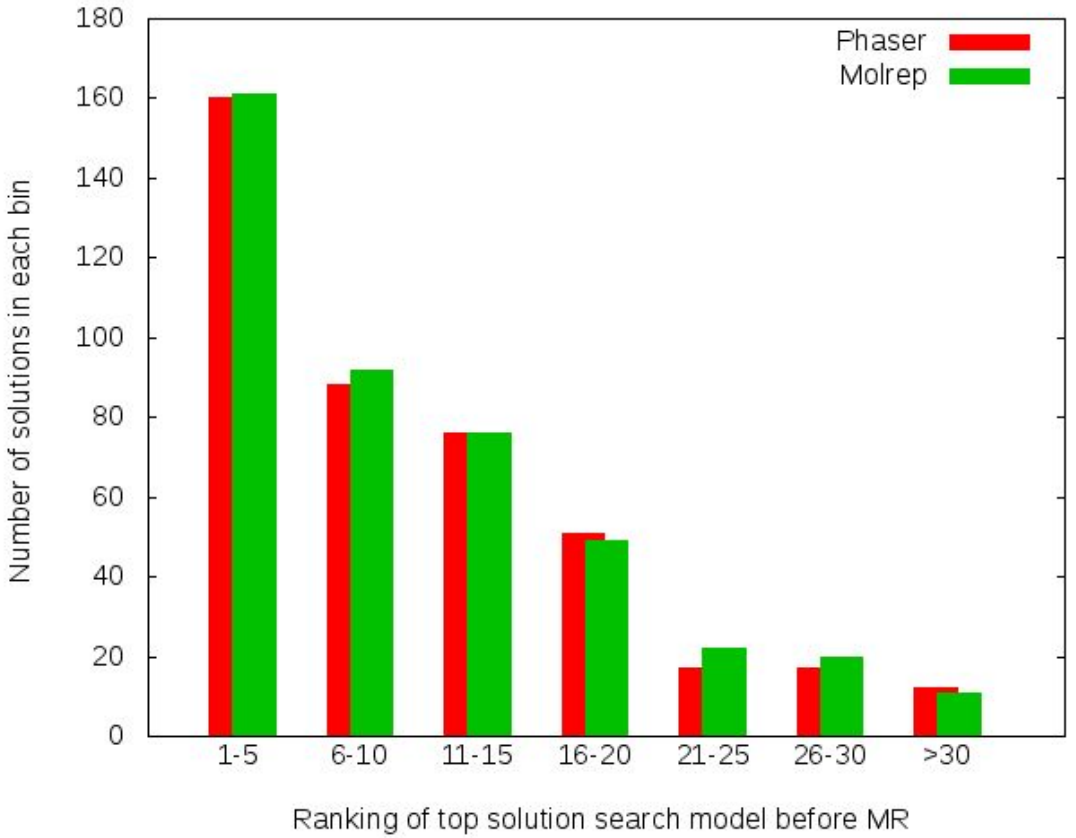


CCP4 Online


www.ccp4.ac.uk/schools/
CCP4 online
www.ccp4.ac.uk/ccp4online/servlet/controller/Link?page=runnableProgramsList.jsp
CCP4 on-line
Collaborative Computational Project No. 4
Software for Macromolecular X-Ray Crystallography
Home (Logout) > Login > Programs
Username: arsefeck
Programs
Note: You must have a CCP4 licence to run these programs.
Balbes
An automated Molecular Replacement (MR) pipeline - Balbes integrates into one system all the components necessary for solving a crystal structure by Molecular Replacement
MrBUMP
An automated Molecular Replacement (MR) pipeline - Given a target sequence and experimental structure factors, it will search for homologous structures, create a set of suitable search models from the template structures, do molecular replacement, and test the solutions with some rounds of restrained refinement.
Zanuda
Space group and crystallographic origin validation
CCP4 STFC Rutherford Appleton Laboratory, Harwell Campus
Didcot Oxon OX11 0FA
Tel:(+44) 1235 567726 | Fax: +44 (+44) 1235 567720 | Email: ccp4@ccp4.ac.uk
About | Legal Statements

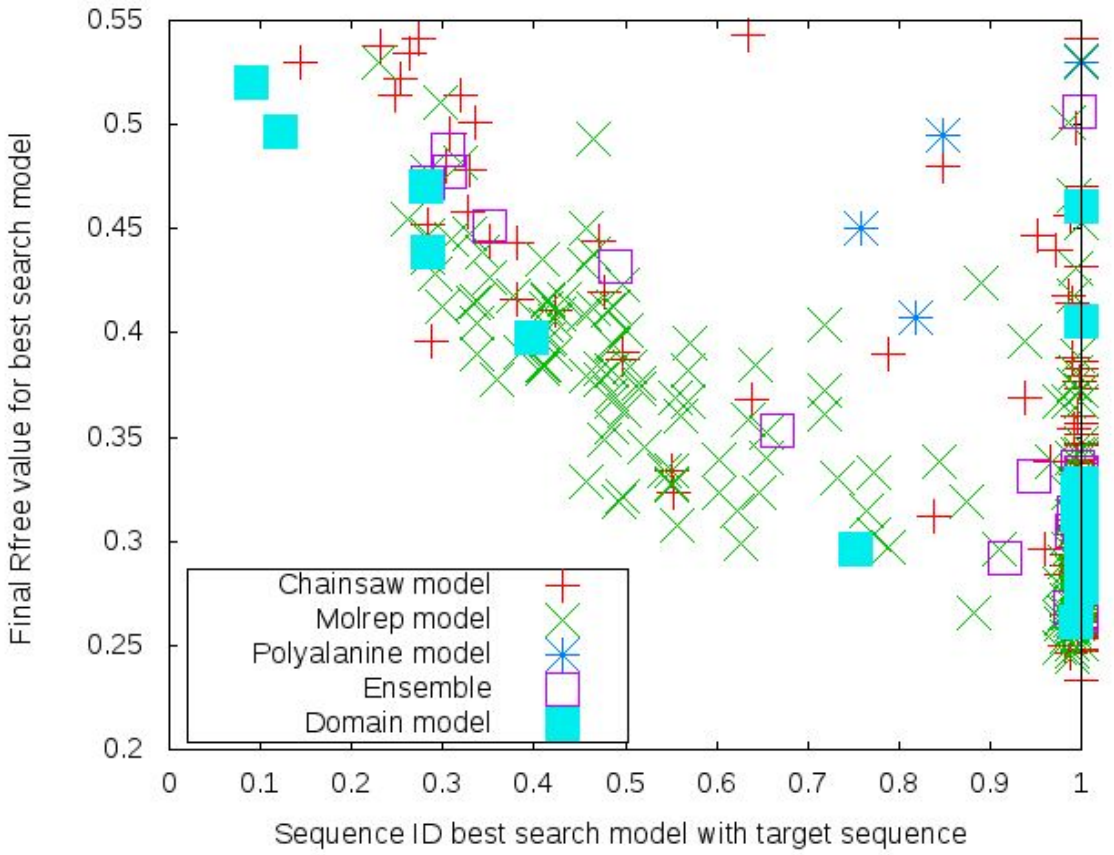
Benefit of processing many models

- Sample set from 2009 PDB depositions processed
- Before MR step all search models are scored based on sequence alignment
- Best MR solutions based against initial scoring of the search model used (results are for successful MR jobs only)



MrBUMP Testing Results

- Model type for best solution (successful MR test cases)



Acknowledgements

- Jaclyn Bibby, Daniel Rigden, Jens Thomas, University of Liverpool
- Olga Mayans, University of Liverpool
- Martyn Winn, Daresbury Laboratory
- Andrea Thorn, Tim Gruene & George Sheldrick (SHELX)
- Developers of Rosetta and Quark
- Refmac: Garib Mushudov, LMB-MRC Cambridge
- Molrep: Alexei Vagin & Andrey Lebedev
- Phaser: Randy Read, Airlie McCoy & Gabor Bunkozci
- Thanks to authors of all underlying programs
- Funding:
    - BBSRC, CCP4
- Support from CCP4 & the Research Complex at Harwell