

Automatic MR

Garib Murshudov
MRC-LMB, Cambridge

These slides provided by Ronan Keegan





MR pipelines in CCP4

MrBUMP - Ronan Keegan, Martin Winn

BALBES – Fei Long, Alexei Vagin, Garib Murshudov

**Ample - Ronan Keegan, Martyn Winn, Jaclyn Bibby,
Jens Thomas, Olga Mayans and Daniel Rigden**



Science & Technology
Facilities Council

MrBUMP

Ronan Keegan

(CCP4 STFC RAL)

Martyn Winn

(STFC DL)



Science & Technology
Facilities Council

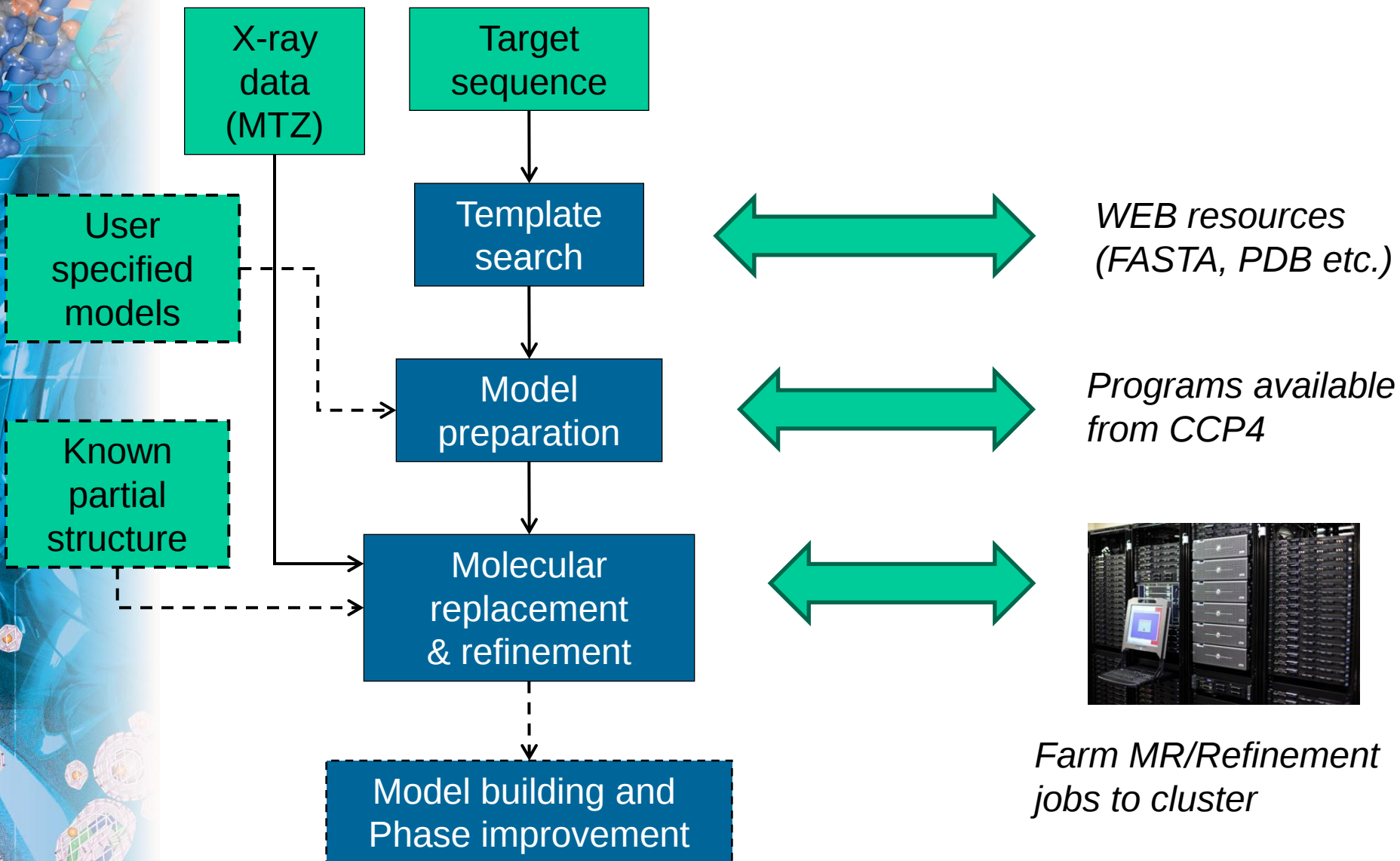
MrBUMP features

- Particular emphasis on generating a **variety of search models**
- Uses a variety of **bioinformatics tools** and **helper applications**
- Uses **on-line databases**

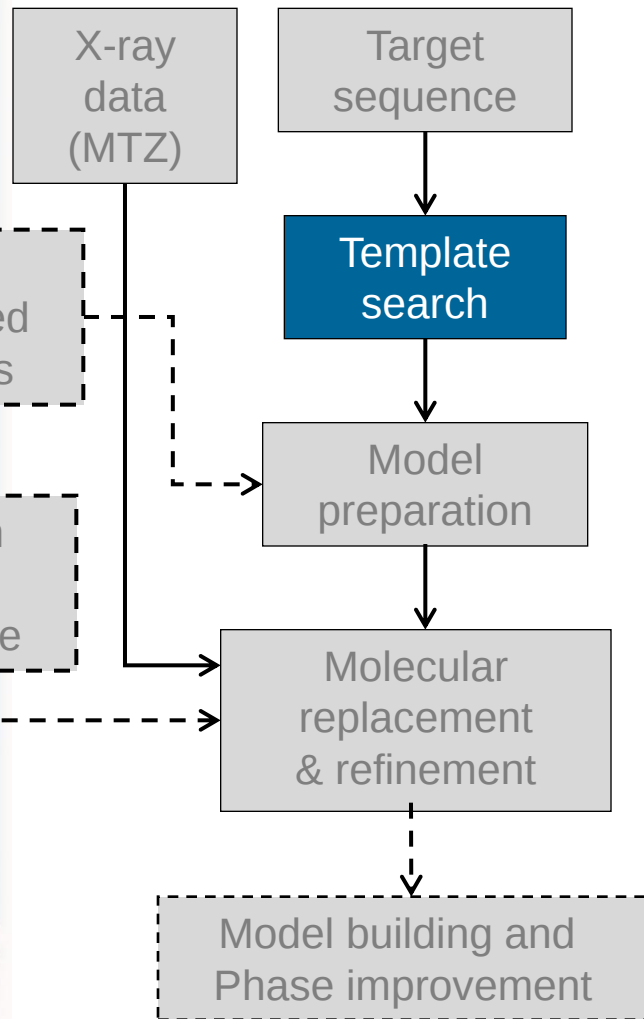
- Favourable cases: gives “one-button” solution
- Complicated cases: lead generation
(suggests likely search models for manual investigation)



MrBUMP Pipeline



MrBUMP Pipeline



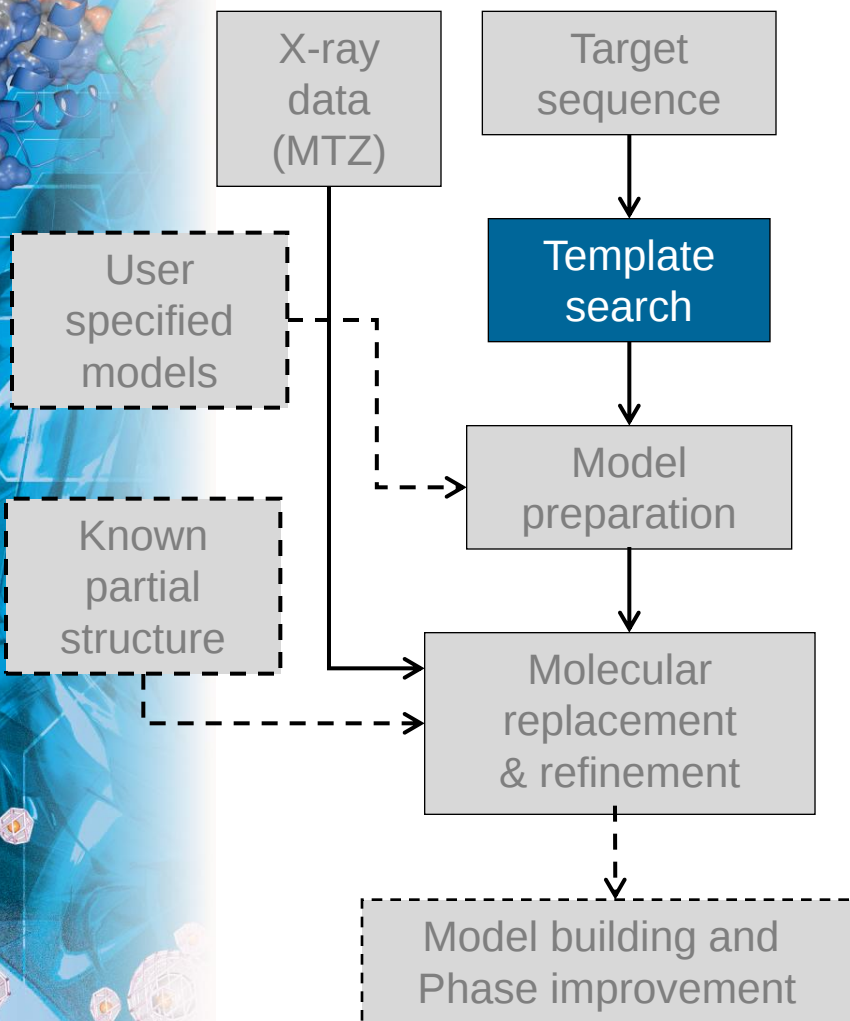
*Initial list of **model templates***

- Sequence based search using target sequence
 - FASTA search of the PDB

*Model template **scoring***

- Based on multiple sequence alignment
 - ClustalW, MAFFT coming with CCP4
 - Probcons, T-coffee optionally installed

MrBUMP Pipeline



Additional model templates

- **Monomers**

- Search **PDBeFold** (**SSM**) for top hits from the initial FASTA search

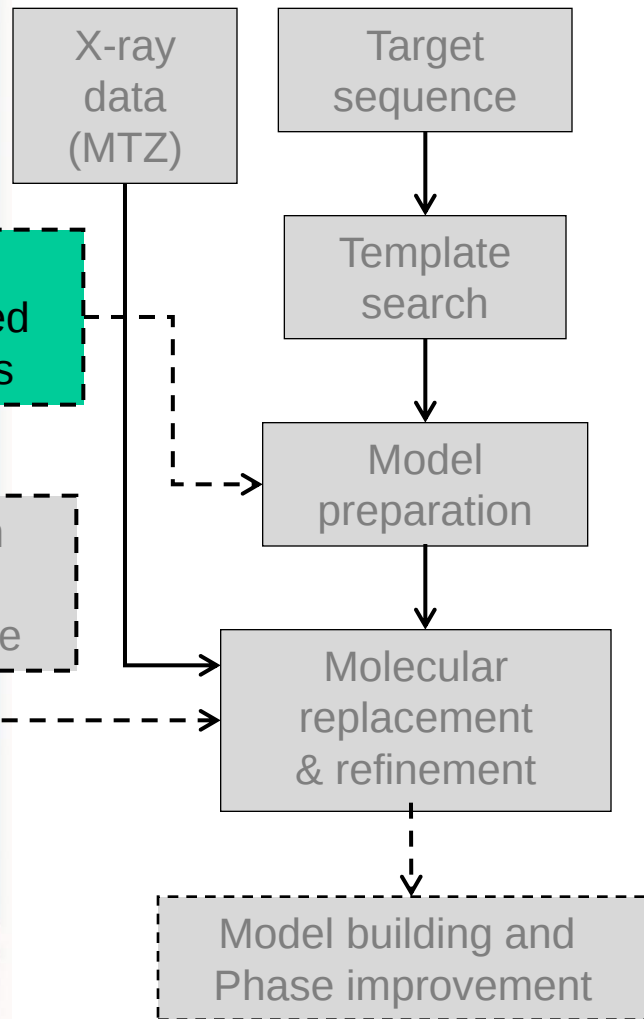
- **Domains if possible**

- Domain definitions from **SCOP** for each initial template

- **Homo-multimers if possible**

- Multimer definitions manually generated using **PISA**

MrBUMP Pipeline

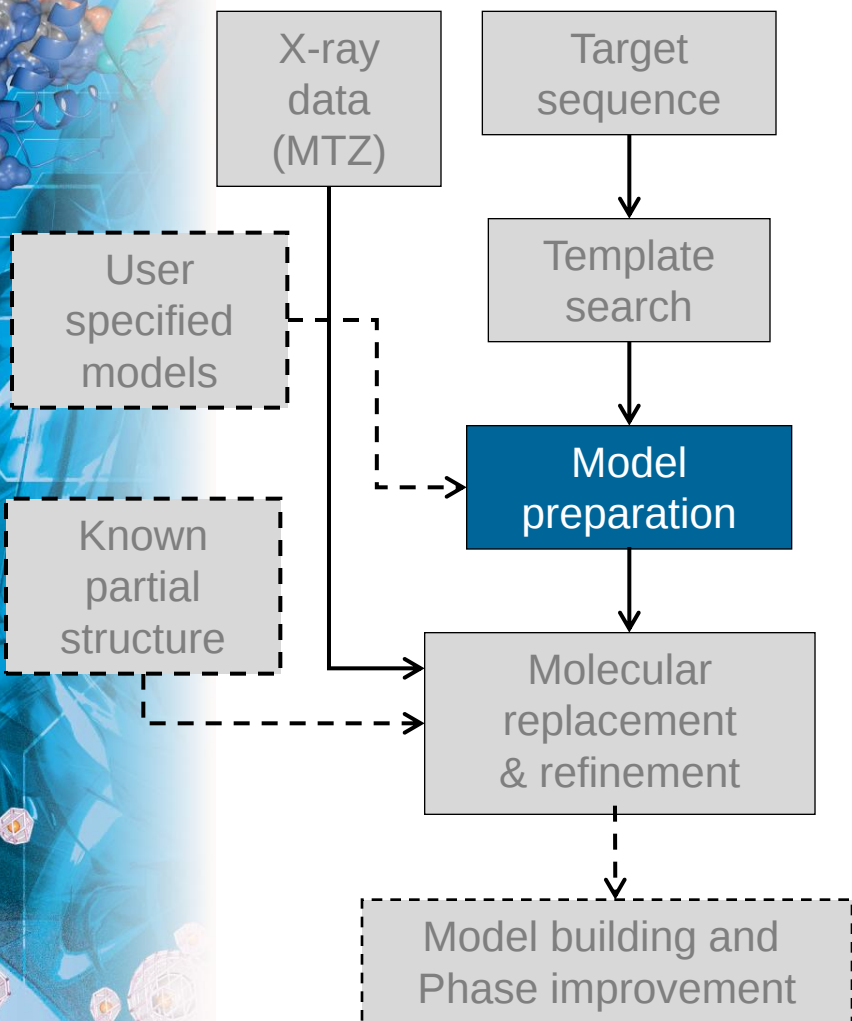


Additional model templates

- **Monomers**

- PDB codes and chain IDs provided by a user
 - » e.g. from **FFAS** or **psiBLAST** searches
- PDB files provided by a user
 - » e.g. unpublished structures

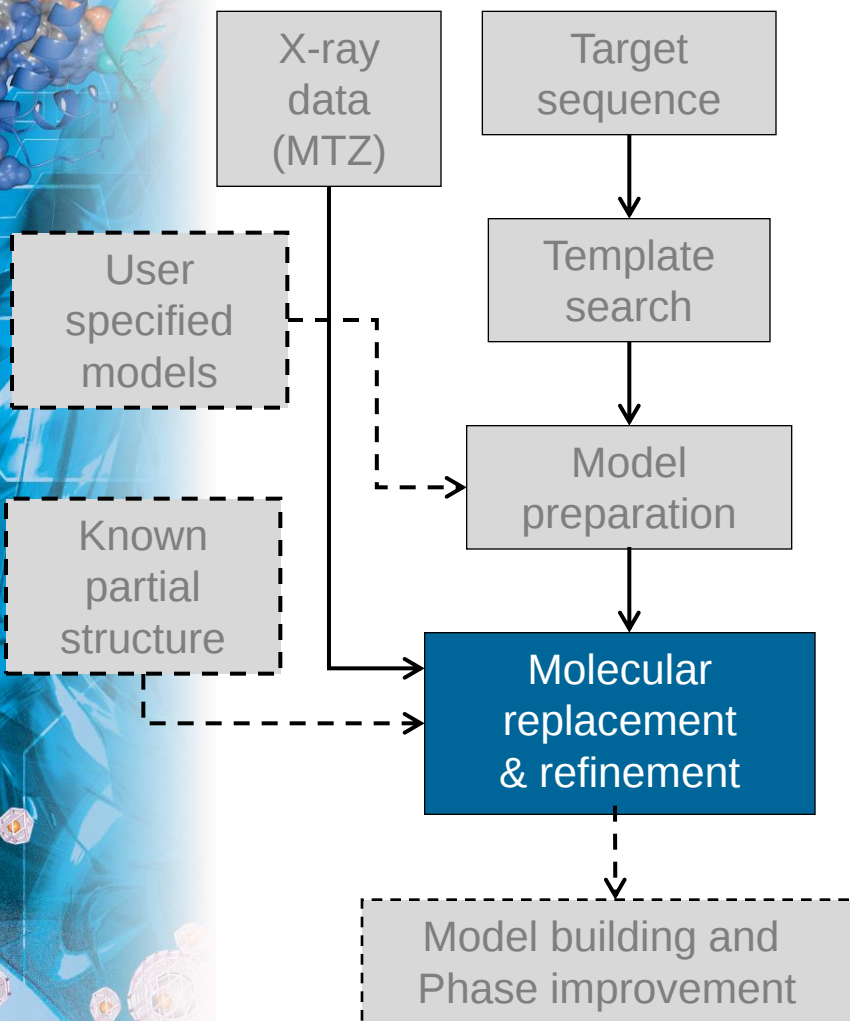
MrBUMP Pipeline



Search Model Preparation

- **Several options** for each template
 - PDBclip (unmodified)
 - Molrep
 - Chainsaw
 - Sculptor
 - polyalanine
- **Ensembles** if requested
 - generated using **Superpose** from top score search models

MrBUMP Pipeline



Molecular Replacement

- **Molrep** or **Phaser** or both for each search model
- Test for enantiomorphic space groups if requested by a user

Restrained Refinement

- Using **Refmac** for each MR result
 - Refinement results are used for **scoring** MR solutions



Scoring MR solutions

final $R_{\text{free}} < 0.35$ or
final $R_{\text{free}} < 0.5$ and dropped by 20%



“good”

final $R_{\text{free}} < 0.48$ or
final $R_{\text{free}} < 0.52$ and dropped by 5%



“marginal”

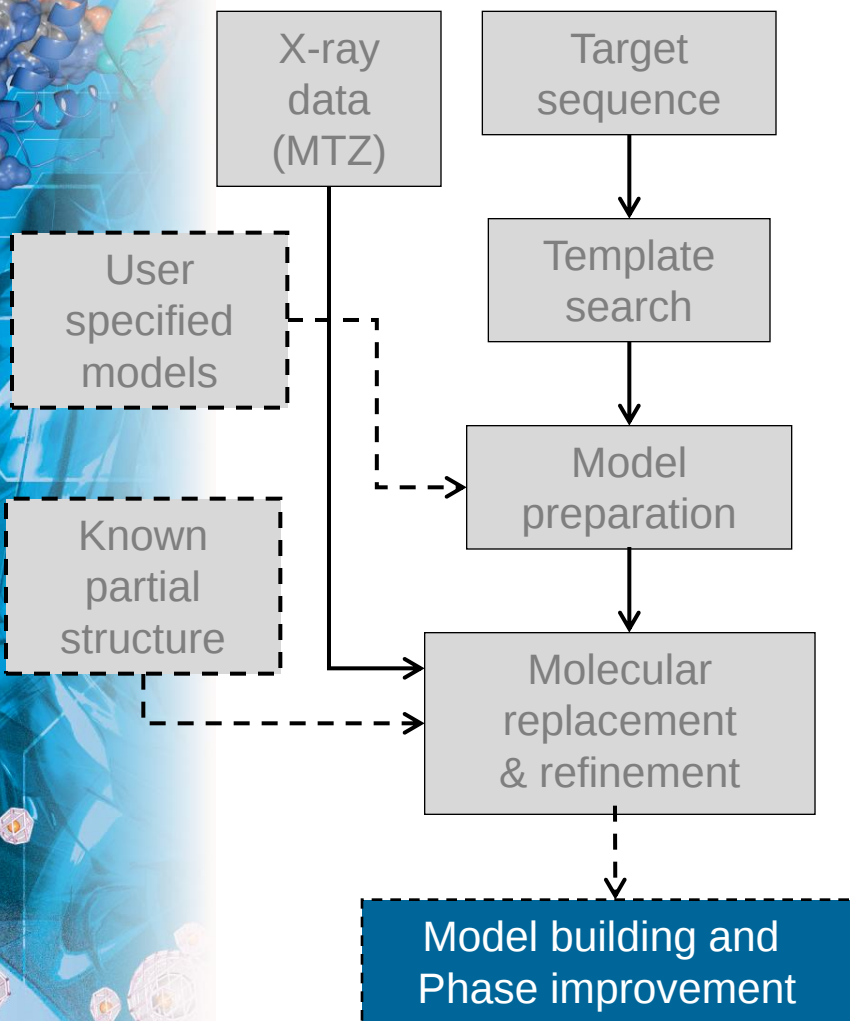
otherwise



“poor”

*conservative approach,
a “poor” solution can still be a solution*

MrBUMP Pipeline



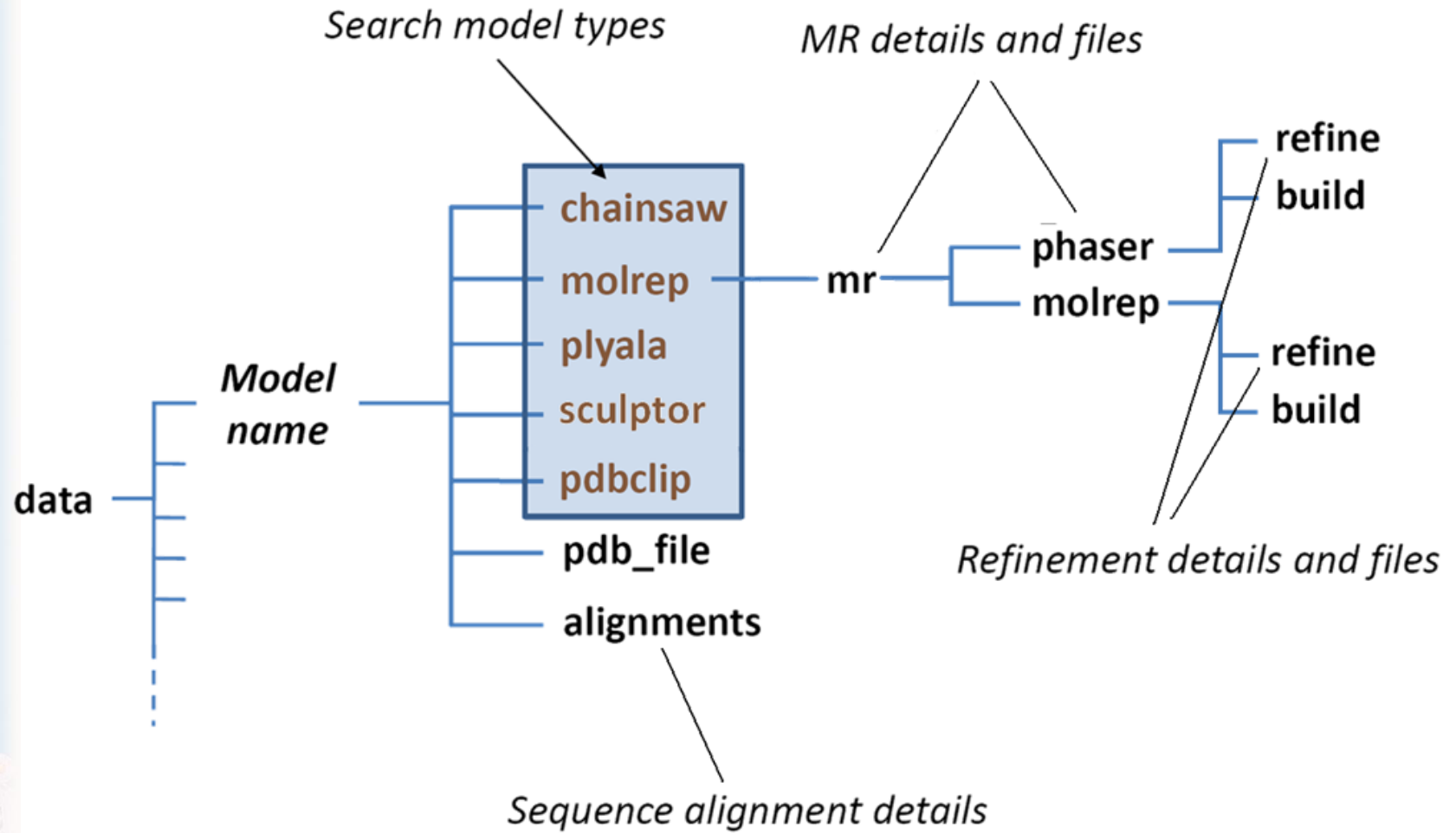
Model building

- **Buccaneer**, **ARP/wARP** or **SHELXE** if requested
- Model completeness is used for **scoring** MR solutions

Phase improvement

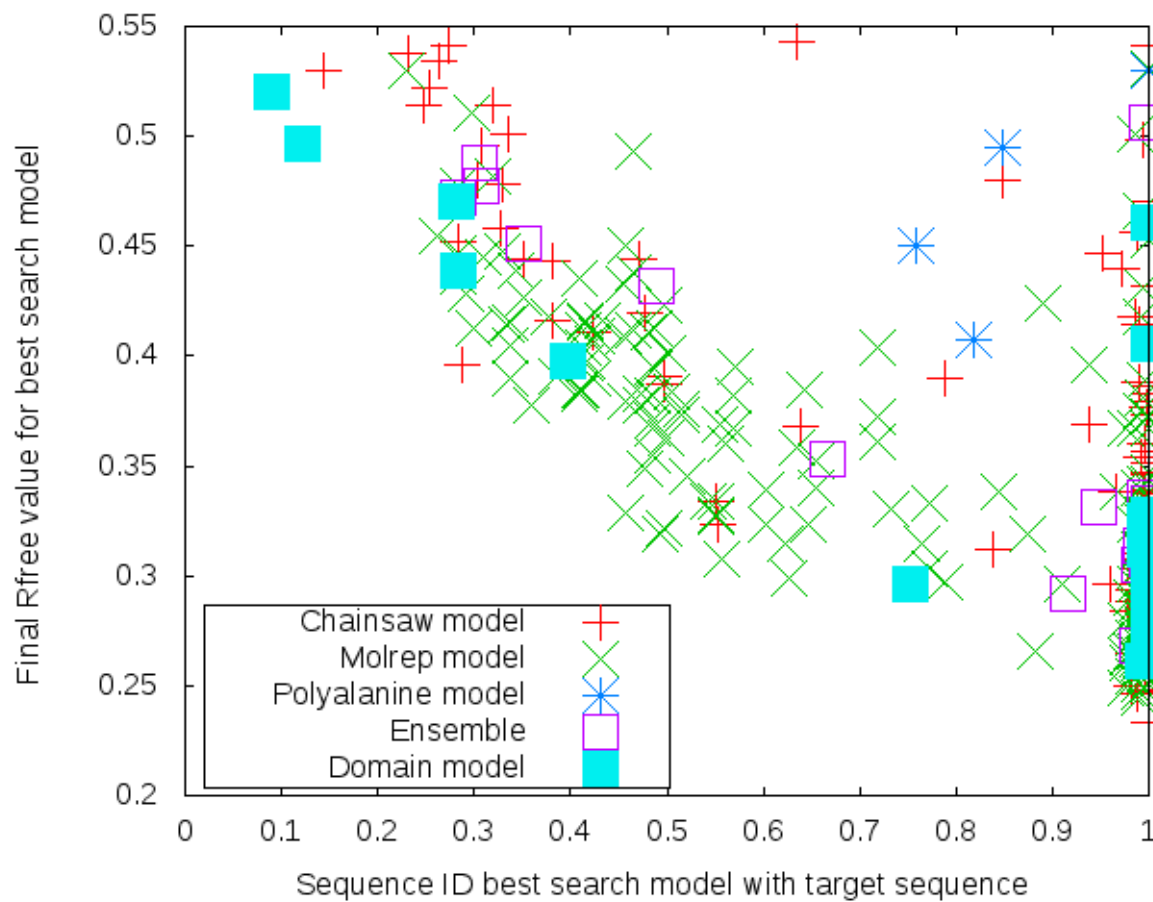
- **Acorn** if resolution is better than 1.7Å
- CC for medium Es is used for **scoring** MR solutions

MrBUMP data directory



MrBUMP Testing Results

Model type for best solution



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]

Free-R: FREE [Browse] [View] Sigma: SIGFP [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: ☒ PDClip ☒ 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 [Add PDB id]

PDB id: 1tom [Add PDB id]

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

Chain id: 1smw_A [Add Chain id]

Chain id: 1smm_B [Add Chain id]

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

[Run] [Save or Restore] [Close]

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

6.1.0 CCP4Interface 2.0.3 running on rmk65sam Project: mrbump

Job	Date	Status	Program	Input	Output
7	22 Jan 09	FINISHED	refmac5	test refma	
6	22 Jan 09	FINISHED	refmac5	test refma	
5	22 Jan 09	FINISHED	refmac5	test refma	
4	22 Jan 09	FINISHED	refmac5	test refma	
3	22 Jan 09	FINISHED	refmac5	test refma	
2	05 Jan 09	FAILED	mrump	felix data	
1	05 Jan 09	FINISHED	mrump	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

Balbes

Fei Long

Garib Murshudov

MRC LMB Cambridge

Alexey Vagin

YSBL University of York

Balbes features

Input

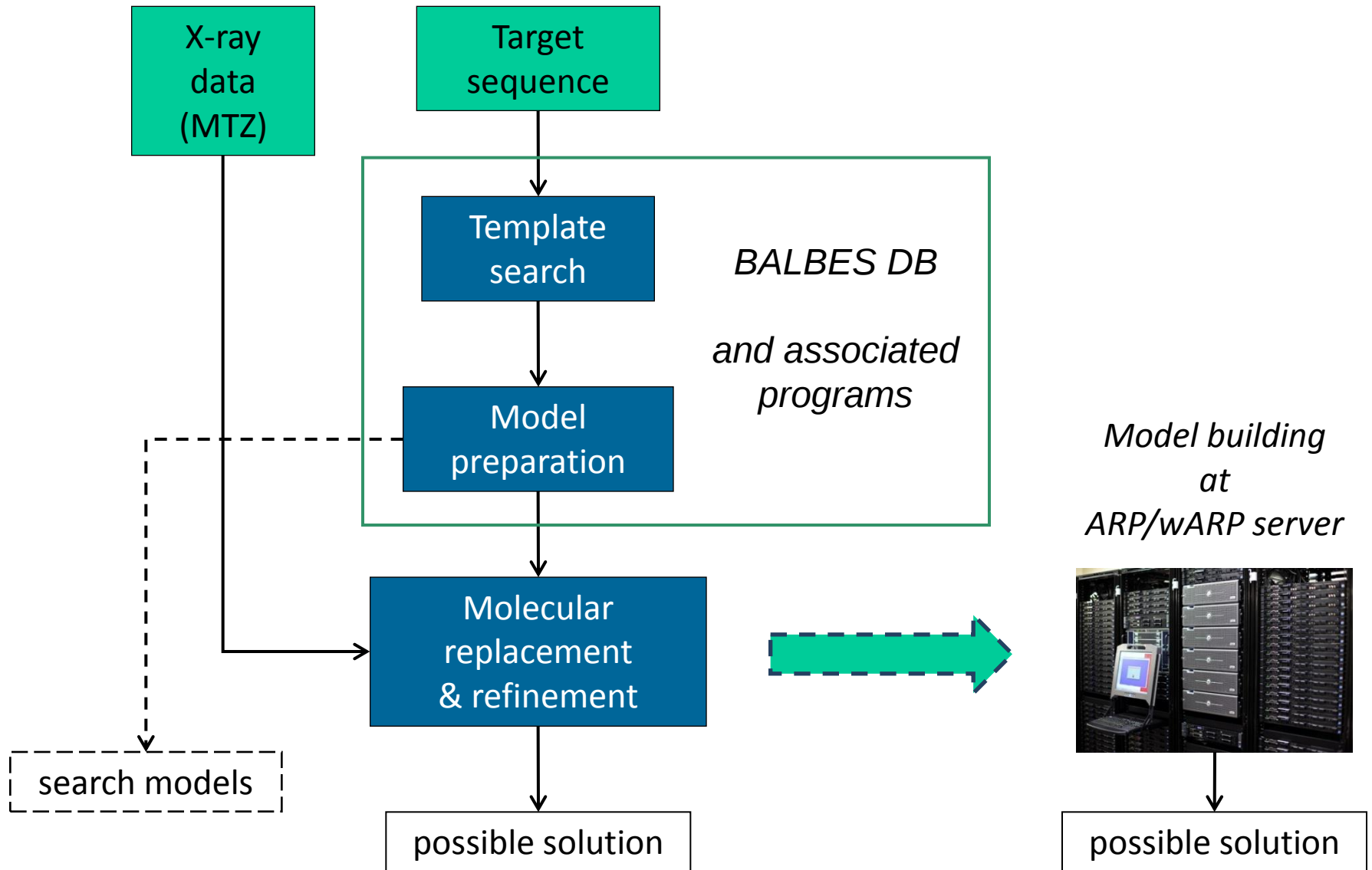
- mtz-file
- sequence

Output

- the best solution
- search models for manual investigation

- Balbes has a **database** containing preprocessed data from the PDB
- Structure solution using **Molrep** and **Refmac**
 - Uses the **search in the density** for a second, third etc. component

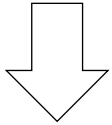
BALBES Pipeline



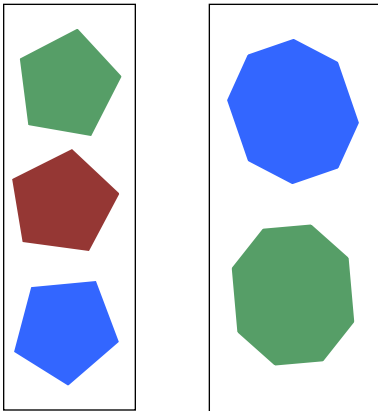
- Non-redundant chains from the PDB
 - The PDB entries of better resolution are preferred
- More than 30000 domain definitions
 - flexible parts removed
 - hierarchically organized according to 3D similarity
- Multimers
 - generated using PISA definition
- Relations
 - domain entry points to original PDB ID, chain ID and residues
 - multimer entry to original PDB ID, chain IDs

Model preparation

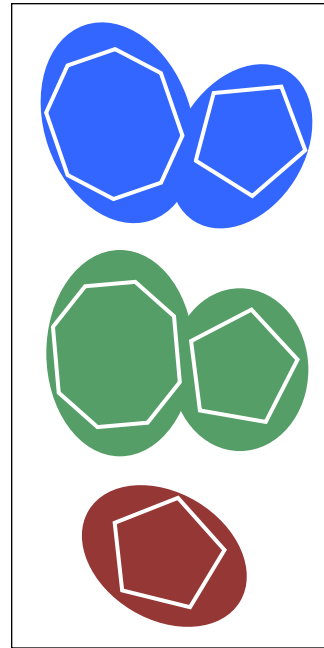
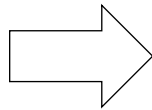
MARPVPEETVATERVKE



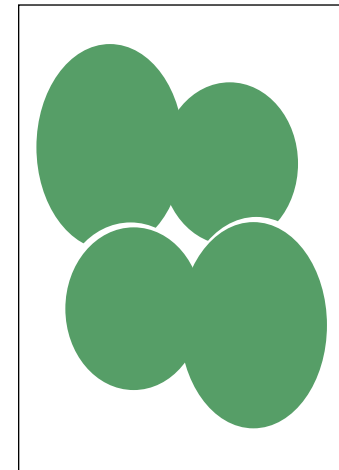
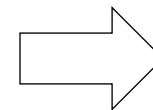
MARPVPEETVATERVKE



Domain models



Chain models



Multimeric models

- All models are corrected
 - by sequence alignment
 - by accessible surface area
- Several input sequences mean protein-protein complex
 - hetero-multimeric models will be generated if possible

Search order

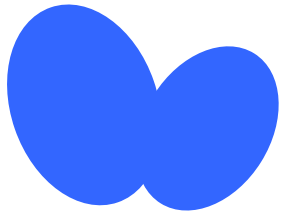


Sequence identity decreases



"Structure 1"

"Model 1"



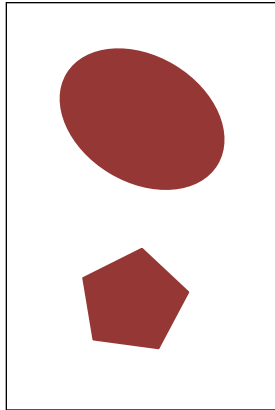
"Model 2"



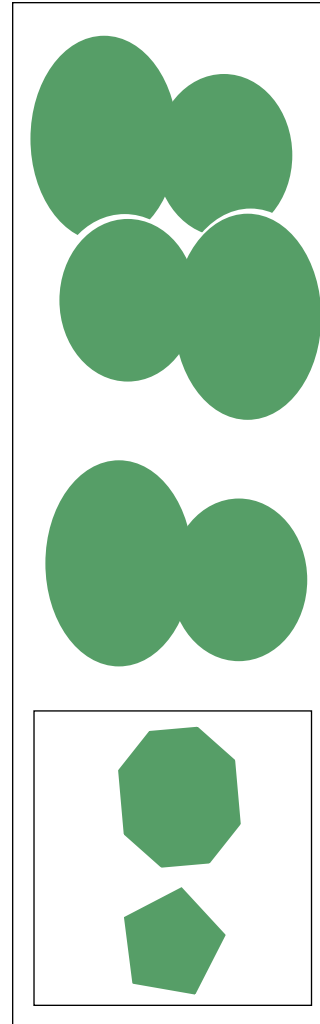
"Model 3"



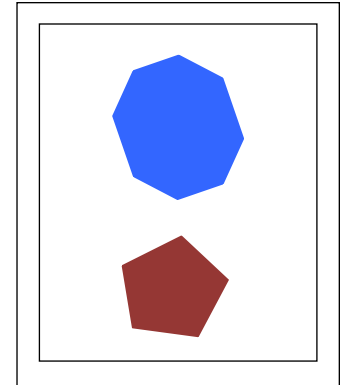
"Structure 2"



"Structure 3"

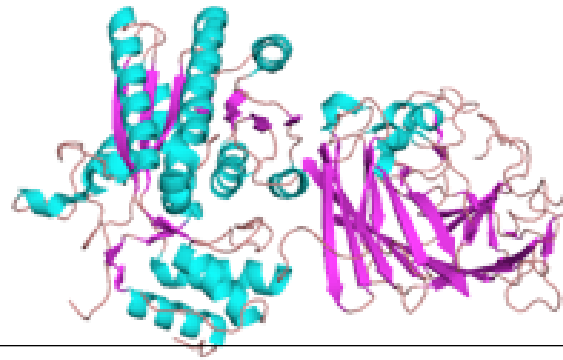


"Multidomain
model"



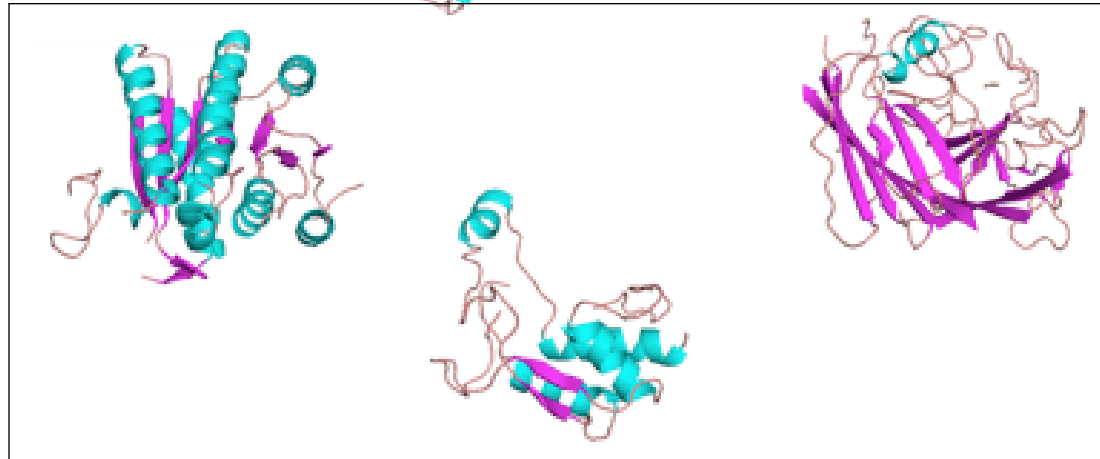
Multi-domain model (test case)

Target



1z45 - isomerase

Multi-domain
model



1yga - domain of
isomerase (51%)

1udc - two domains of
isomerase (49%)

1ek6 - two domains of
isomerase (55%)

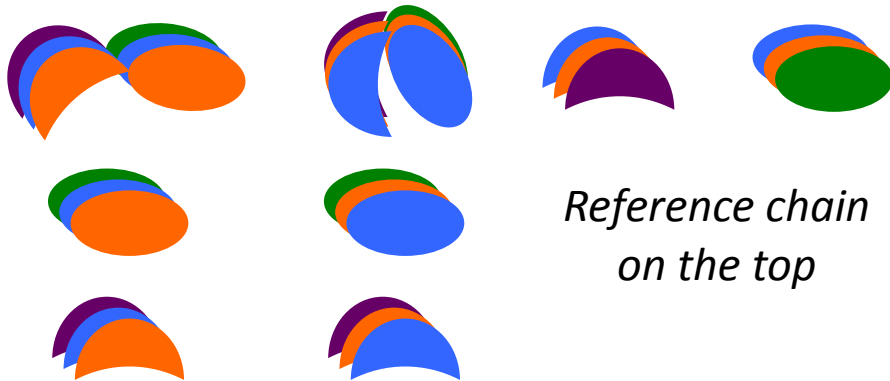
Ensemble models

Ensemble models are generated if possible

Homologues from the Balbes database:



Ensemble search models:



Ensemble models from Balbes

THE UNIVERSITY of York
York Structural Biology Laboratory

University | Chemistry | YSBL

Home (Logout) > Login > Programs > Balbes > View Results Username: alezie

View Results

PROCESS **7587375708** IS RUNNING [[stop process](#)]

PLEASE CLICK HERE TO REFRESH THIS PAGE (or use the REFRESH PAGE link at the bottom)

FILES

output/

process_details/

template_solutions/

results/

template_models/

CURRENT FILE: **summary.log** [[download this file](#)]

BALBES Version is 1.1.1_DB_Oct_1_2010

process_details/

✓ process_details/
 sfcheck_gfobs.log
 sfcheck.log
 sfcheck.btc

mod_2b1qA_mono2.pdb
mod_2b1qA_mono2_ens.pdb
 mod_2b1qA_1dom.pdb
 mod_2b1qA_tmp.pdb
 mod_2b1qA_1dom_ens.pdb
 mod_2b1qA_2dom.pdb
 mod_2b1qA_2dom_ens.pdb
 mod_2b1rA_mono.pdb
 mod_2b1rA_mono1.pdb

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

Similarity

0.857 (ENS)

0.846 (ENS)

0.861 (ENS)

AMPLE

Ab initio Modelling of Proteins for moLEcular replacement

- Uses Rosetta to generate *ab initio* models
- Several new structures has been already solved

*Jaclyn Bibby, Jens Thomas,
Olga Mayans and Daniel Rigden
Institute of Integrative Biology*

*Ronan Keegan and Martyn
Winn*

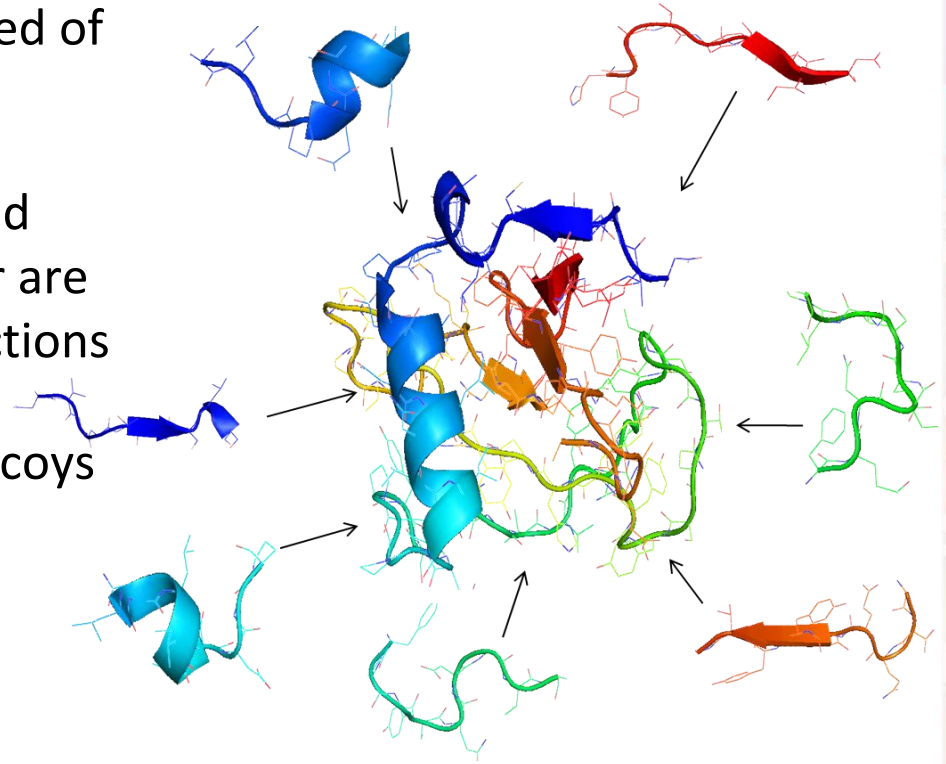
*Collaborative
Computational Project 4
(CCP4)*



Ab initio structure prediction

Modest computing power is needed:

1. 1000's of “Decoys” assembled of fragments from PDB structures
2. Decoys are clustered and centroid representatives of largest cluster are considered candidate fold predictions
3. Side chains added to selected decoys



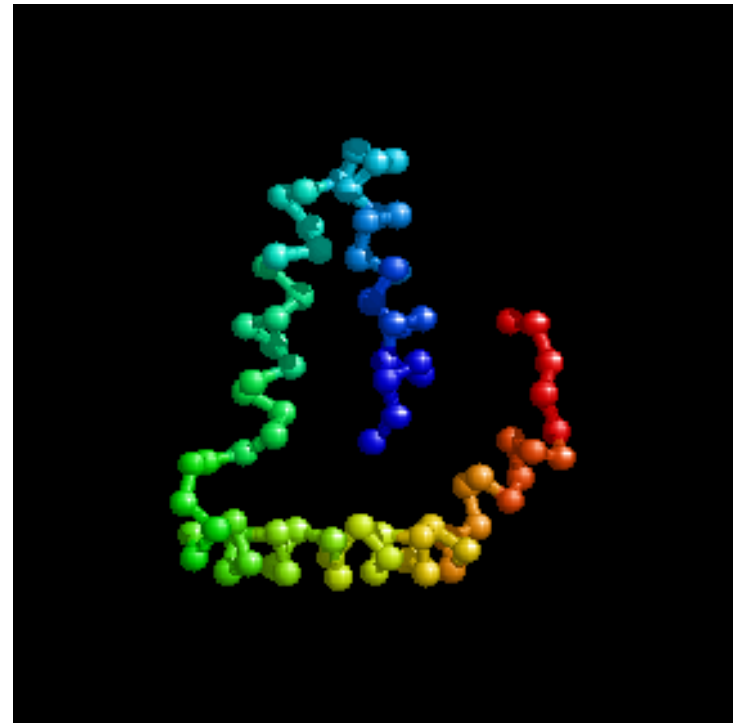
Ab initio structure prediction

Modest computing power is needed:

1. 1000's of “Decoys” assembled of fragments from PDB structures
2. Decoys are clustered and centroid representatives of largest cluster are considered candidate fold predictions
3. Side chains added to selected decoys

Supercomputing resources are needed:

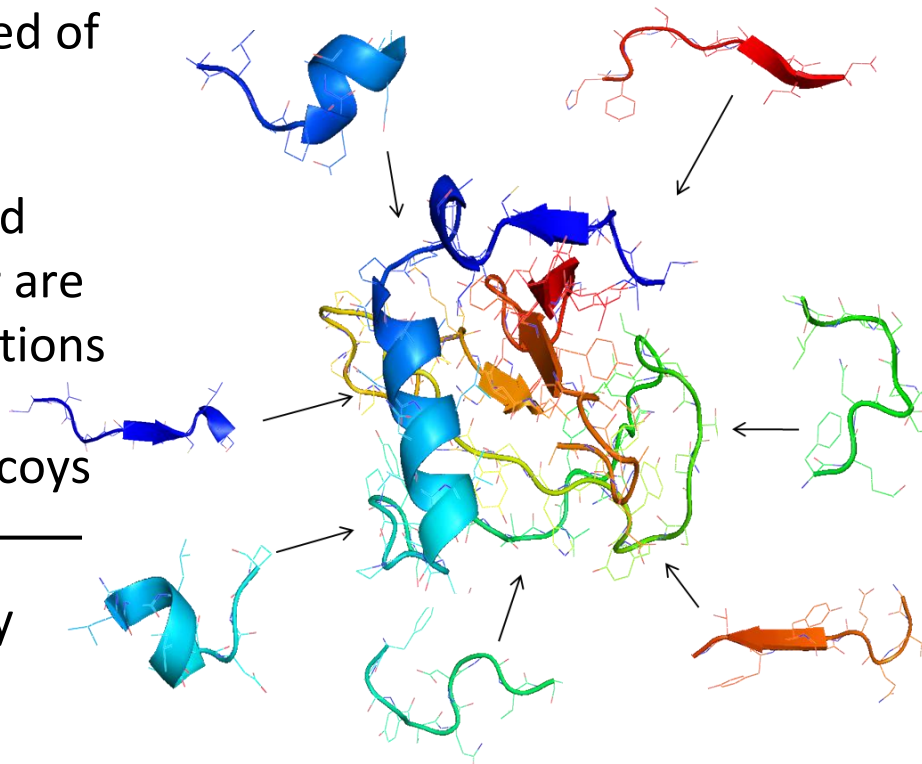
4. Force field refinement

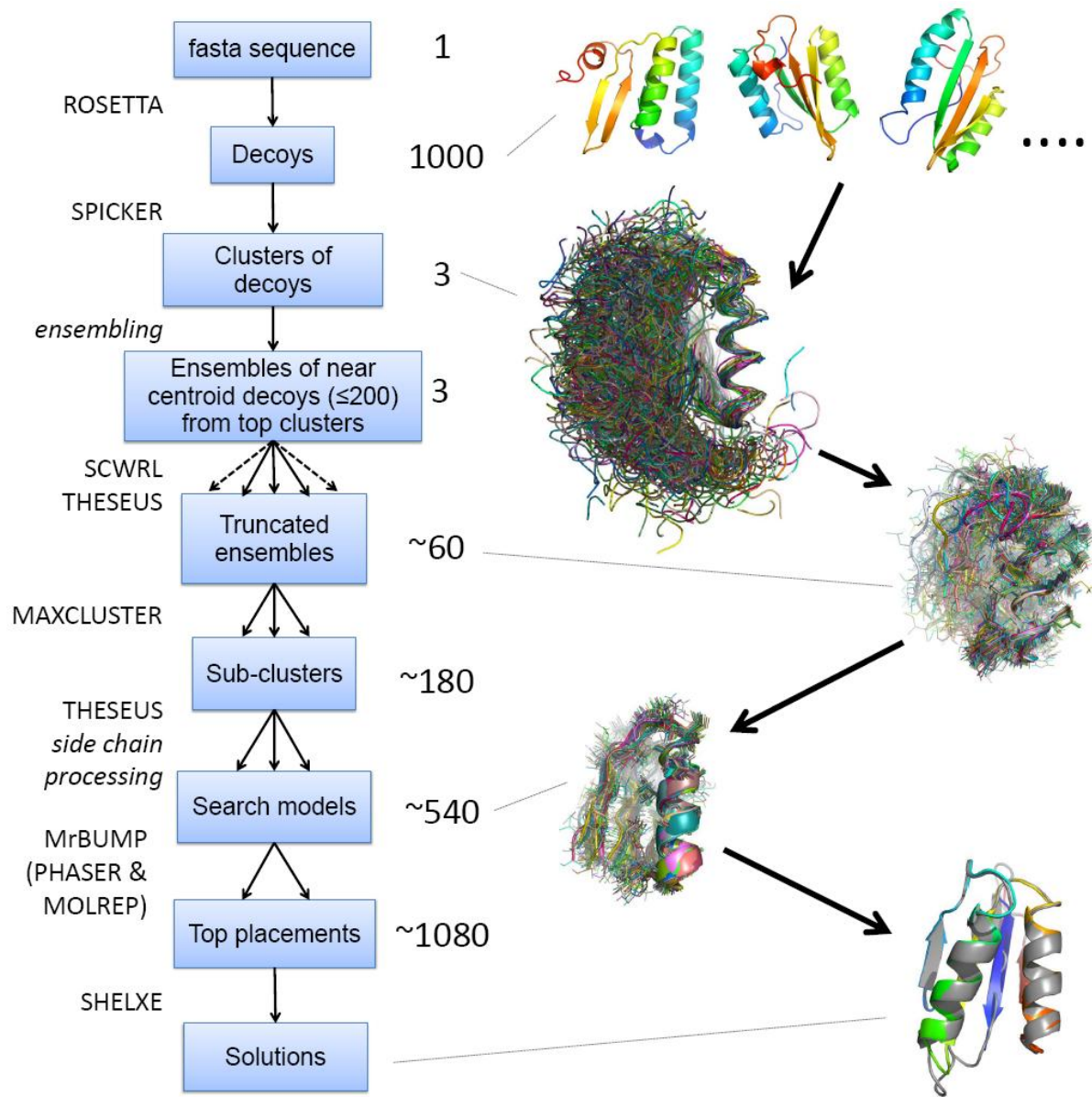


AMPLE

Modest computing power is needed:

1. 1000's of “Decoys” assembled of fragments from PDB structures
 2. Decoys are clustered and centroid representatives of largest cluster are considered candidate fold predictions
 3. Side chains added to selected decoys
-
4. Ensembles generated from decoy models are used for Molecular Replacement





Assessing Solutions

Results of structure rebuilding
as a criterion

- SHELXE C^α tracing
 - partial CC of >25%
 - average fragment length of 10 or more
- Further rebuilding including side chains
 - ARP/wARP
 - Buccaneer.

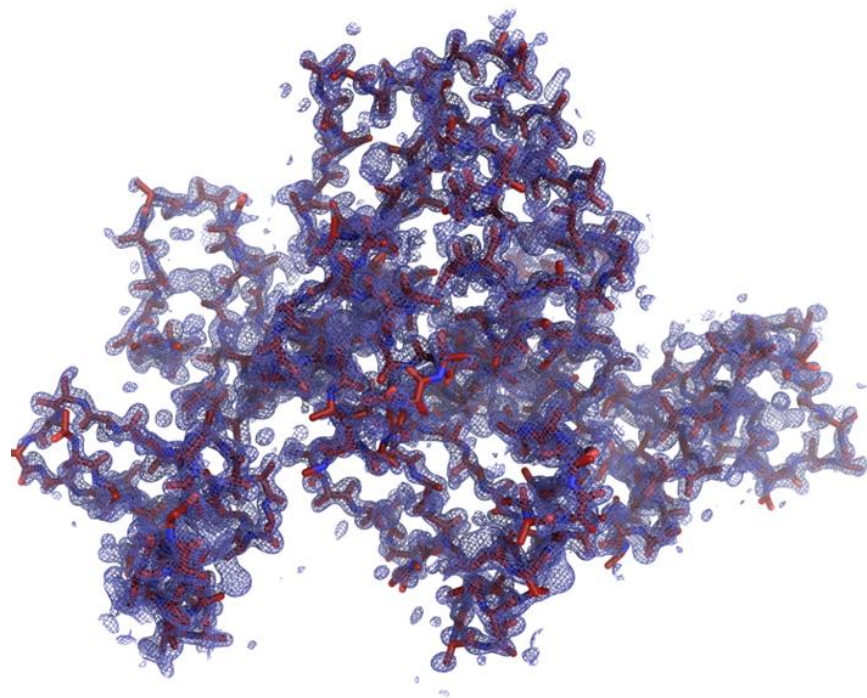
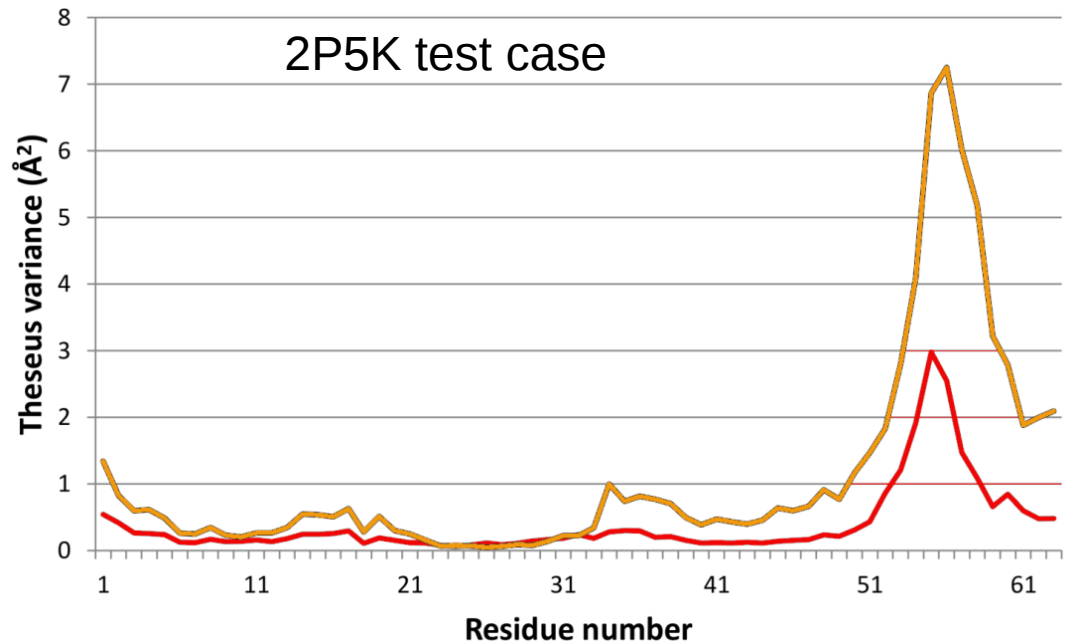


Figure courtesy of Andrea Thorn

Variance and Truncation

- Strong correlation between
 - deviation from the true structure
 - variability within decoy cluster



- Recipe for model generation:
truncation at different ***Theseus*** variance levels

Synergies

<i>Ab initio</i> modelling	Molecular Replacement
Produces clusters of similar model structures	Works effectively with superposed ensembles
Variance within the cluster correlates with similarity to the target structure	Works most effectively if unreliable parts of model are truncated



Ab initio tests

Test set of 295 small proteins from the PDB (40-120 residues)

- Resolution of 2.2 Å or better
- Single molecule in the asymmetric unit
- Fragment generation step: **homologues** were **excluded**
- 1000 decoys generated for each case using Rosetta

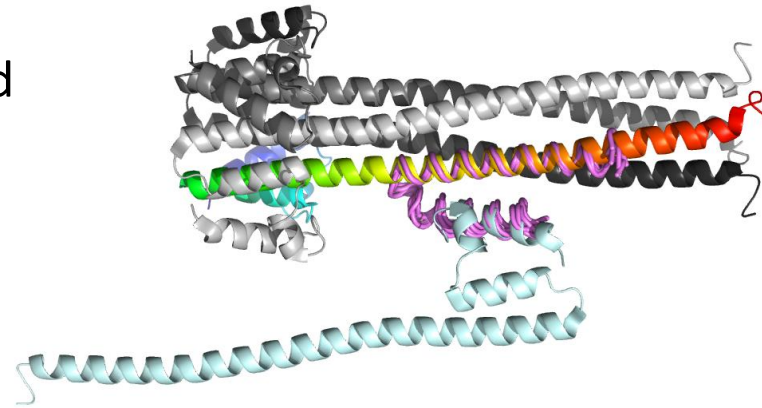
Fold	all- α	all- β	mixed α - β	all cases
Success rate	80%	2%	37%	43%



All α -helical proteins are possible *ab-initio* targets

Coiled-coil proteins

- Difficult to solve in MR even with good homologues
- Initial tests: 80% success rate
- Some novel structures have also been solved



Transmembrane Proteins

- Very difficult to work with/crystallise
- 18 test cases of 1.45 - 2.5Å resolution
- 7 clear and 5 possible successes
- 223 residue structure (3GD8) could be largest ever solved *ab initio*.

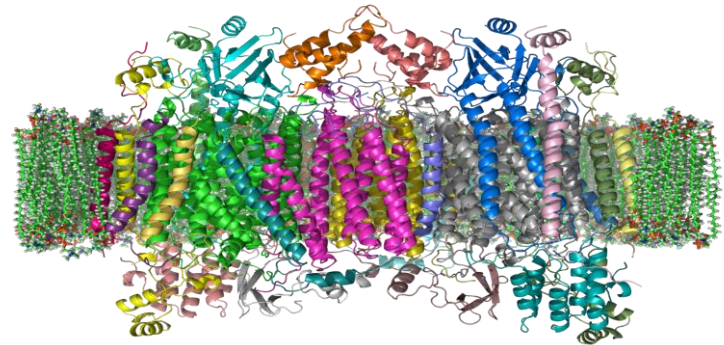


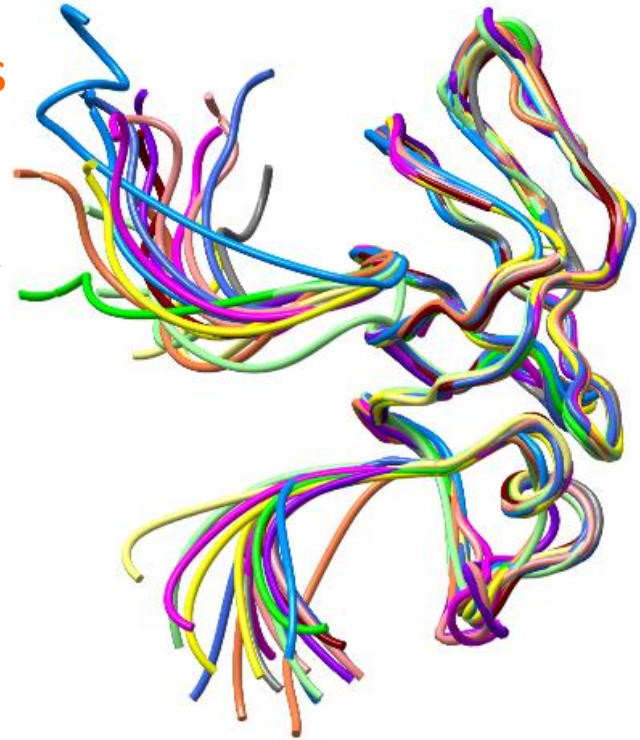
Image:

http://en.wikiversity.org/wiki/File:Cytochrome_C_Oxidase_1OCC_in_Membrane_2.png

General targets

AMPLE can use Rosetta to “re-model” a **particular set of templates**

- Remodelling related NMR structures
- Exploiting distant homologues



AMPLE usage

- Complete run: 2 CPU days
- A parallelised version on a cluster: 1 hour

Distributed as a part of CCP4 suite

- 6.3.0: beta version
- 6.4.0 (upcoming and installed here): improved and more robust version
- Currently required non-CCP4 packages:
 - Rosetta, SHELXE, Theseus, SPICKER, Maxcluster



CCP4 Program Suite 6.3.0 CCP4Interface 2.2.0 running on ccp4i1 Project: aps5-raja 100% 320KB 319.7KB/s 00:00

Change Project Help

Molecular Replacement

- Analysis
- Model Generation
- Phaser MR
- Run Molrep - auto MR
- Run MrBUMP
- Run Balbes
- Run AMPLE (beta)
- AMoRe Suite
- Utilities

Job ID	Date	Status	Description
15	11 Jul 12	FINISHED	arp_warp_classic automated model building starting
14	11 Jul 12	FINISHED	buccaneer_pipeline [No title given]
13	11 Jul 12	FINISHED	refmac5 refine jac soln
12	10 Jul 12	FINISHED	buccaneer_pipeline [No title given]
11	10 Jul 12	FINISHED	arp_warp_classic automated model building starting
10	10 Jul 12	FINISHED	refmac5 refine jac soln
9	10 Jul 12	FINISHED	refmac5 refine jac soln
8	10 Jul 12	FINISHED	refmac5 refine jac soln
7	10 Jul 12	FINISHED	refmac5 refine jac soln
6	10 Jul 12	FINISHED	arp_warp_classic automated model building starting
5	10 Jul 12	FINISHED	buccaneer_pipeline [No title given]
4	10 Jul 12	FINISHED	refmac5 refine jac soln
3	10 Jul 12	FINISHED	refmac5 refine jac soln
2	10 Jul 12	FINISHED	refmac5 refine jac soln
1	10 Jul 12	FINISHED	refmac5 refine jac soln

Directories&ProjectDir

View Any File

View Files from Job

Search/Sort Database..

Graphical View of Project

Delete/Archive Files..

AMPLE (beta) - Ab initio modelling for Molecular Replacement Initial parameters from /home/rola/

Job title 2cmp ab initio model MR Ample User Guide

Input Files

SEQ in ample 2cmp.fasta Browse View

MTZ in ample 2cmp.mtz Browse View

F FP Sigma SIGFP

Free-R FREE

Number of models to generate: 500

Number of Processors 1

Fragment Files

To generate fragment files for your sequence you need to use the Robetta Fragment library server

Click here to go to online Robetta Server (registration required)

3mers (aaXXXX_03_05.200_v1_3) /home/rmk65/CCP4-Projects/ample Browse

9mers (aaXXXX_09_05.200_v1_3) /home/rmk65/CCP4-Projects/ample Browse

Rosetta Installation

Rosetta Installation Directory /home/rmk65/opt/src/rosetta3.4 Browse

Run Save or Restore Close

robetta.bakerlab.org

ROBETTA BETA

Full-chain Protein Structure Prediction Server

Model 1 Target - T0513

2.66 Å over 62 residues

0.84 Å over 39 residues

REGISTRATION

[Register / Update] [Login]

DOCUMENTATION

[Docs / FAQs]

SERVICES

Domain Parsing & 3-D Modeling

[Queue] [Submit]

Interface Atomic Scanning

[Queue] [Submit]

Fragment Libraries

[Queue] [Submit]

DNA Interface Flexible Scanning

[Queue] [Submit]

RELATED SITES

RosettaBackrub Server

RosettaAntibody Server

RosettaDesign Server

RosettaDock Server

RosettaCambridge

Rosetta

Rosetta/Amber

Rosetta/Amber/Amber

Documentation available from
<http://ccp4wiki.org>

Acknowledgements

AmoRe

Jorge Navaza

IBS Grenoble

Molrep

Alexey Vagin

University of York

Phaser

Randy Read

University of Cambridge

Airlie McCoy

University of Cambridge

Gabor Bunkozci

University of Cambridge

Acknowledgements

Balbes

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

MrBUMP

Ronan Keegan	STFC RAL
Martyn Winn	STFC DL

Ample

Ronan Keegan	STFC RAL
Jaclyn Bibby	University of Liverpool
Daniel Rigden	University of Liverpool
Olga Mayans	University of Liverpool