



UNIVERSITY OF
LIVERPOOL

Ronan Keegan CCP₄ Group

MR Search Models and Assessing the Solution

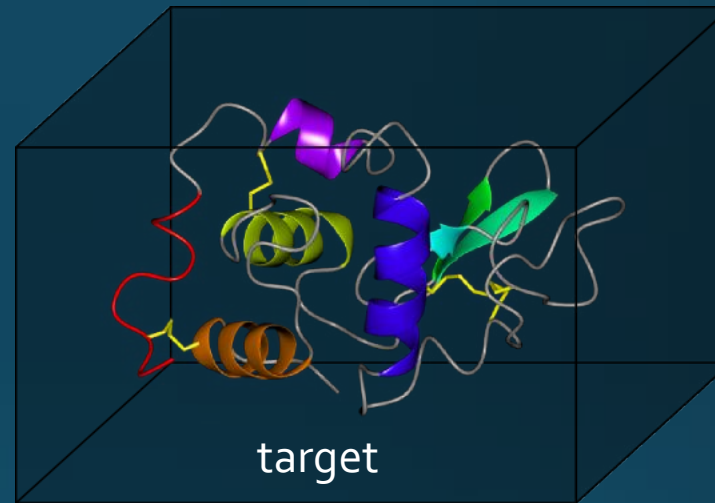
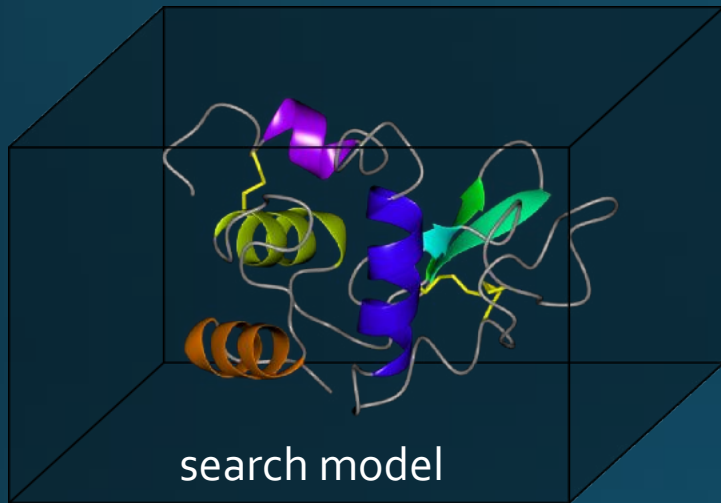
Macromolecular Crystallography School 2018

IFSC, Sao Carlos 14-26.11.2018

Overview

- Introduction
- Step-by-step guide to performing Molecular Replacement
- Automatic Molecular Replacement in CCP4
- What to do if Molecular Replacement doesn't work
- Acknowledgements

Introduction: *Basics of Molecular Replacement*

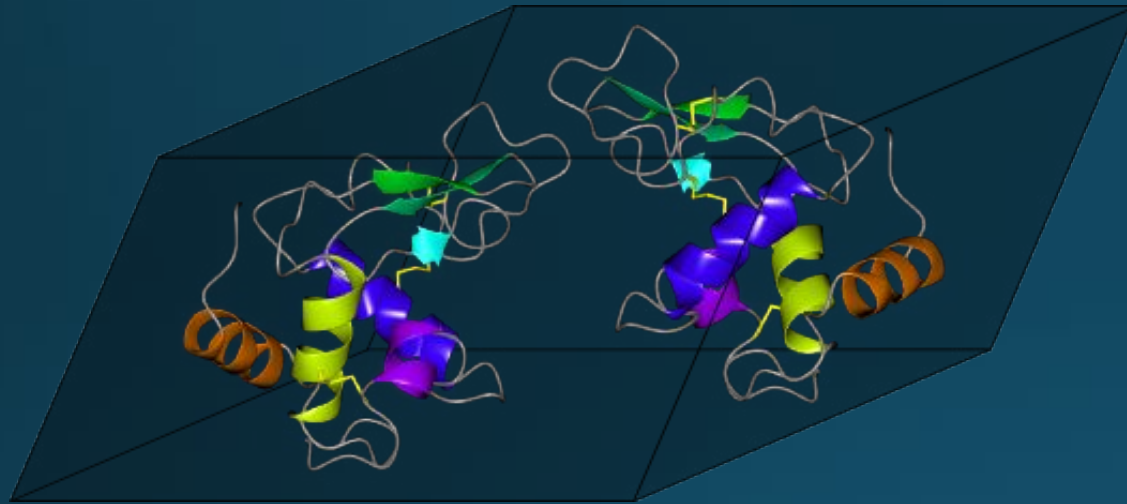


h	k	l	F	ϕ
0	0	1	12.6	123
0	0	2	2.1	12
0	0	3	69.9	287

h	k	l	F	ϕ
0	0	1	10.4	113
0	0	2	3.5	18
0	0	3	57.2	265

(slide from Gabor Bunkoczi)

Introduction: *Basics of Molecular Replacement*



Search model crystal form



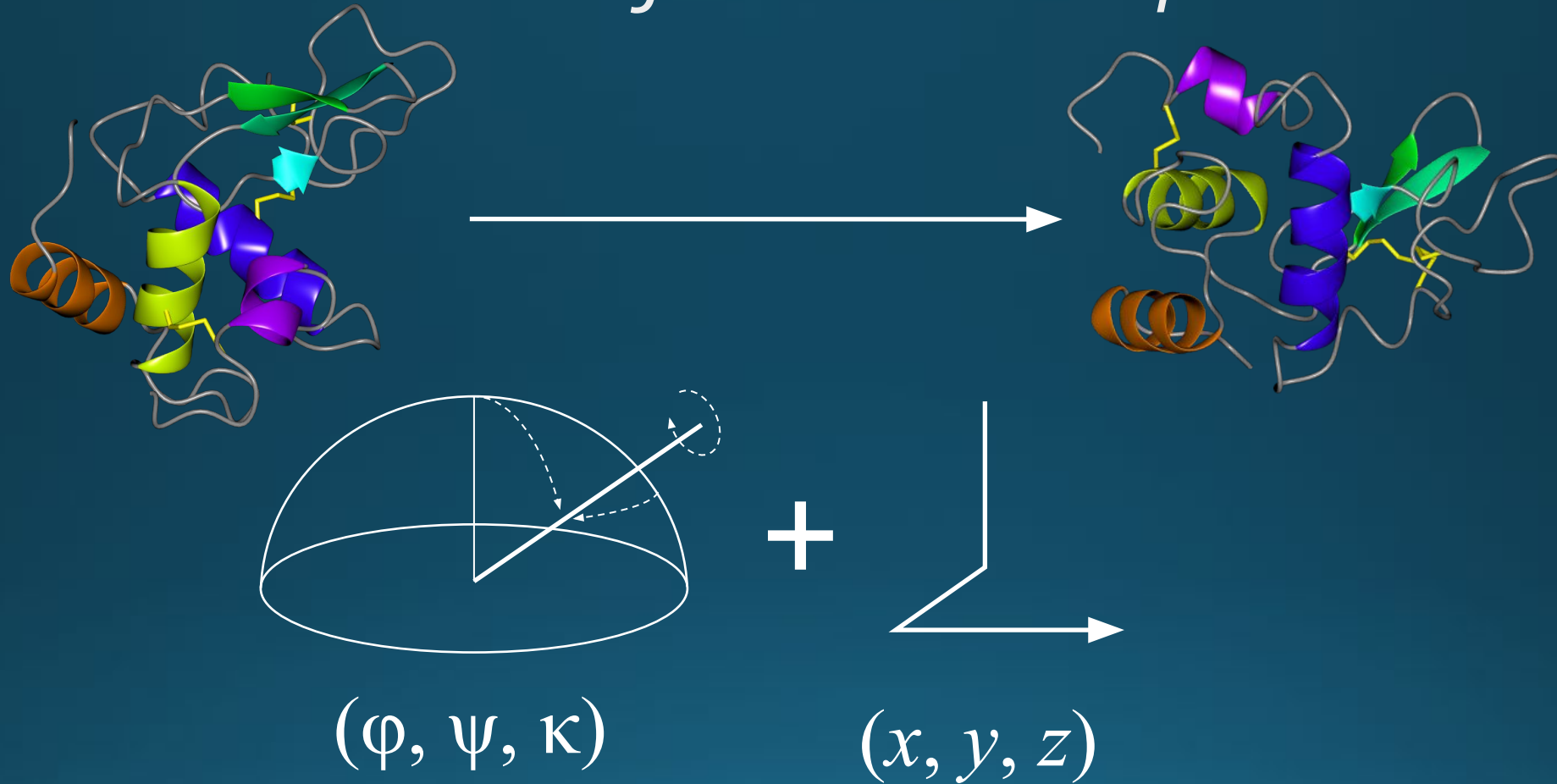
Target crystal form



$$\mathcal{R}\{\varphi, \psi, \kappa, x, y, z\}$$

(slide from Gabor Bunkoczi)

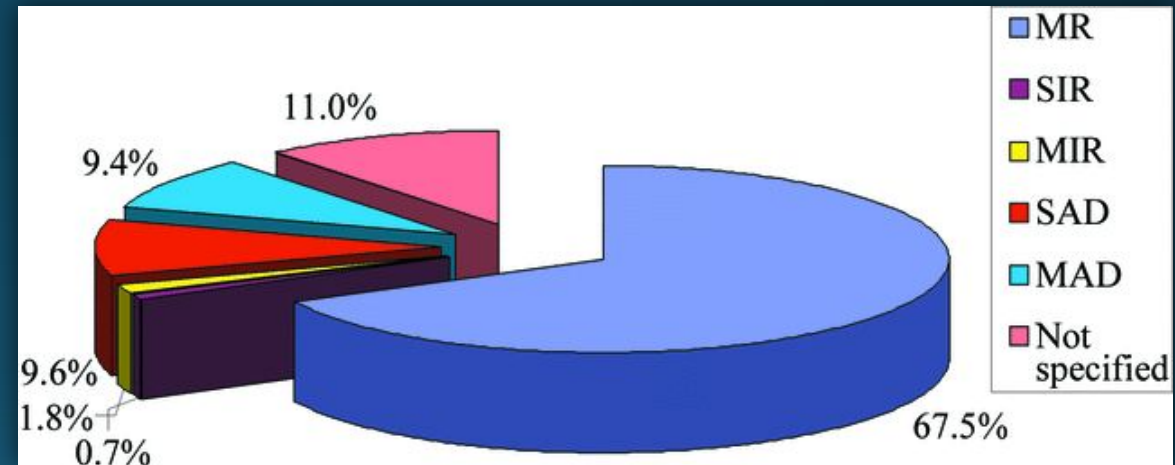
Introduction: *Basics of Molecular Replacement*



(slide from Gabor Bunkoczi)

Introduction: *Molecular Replacement and the PDB*

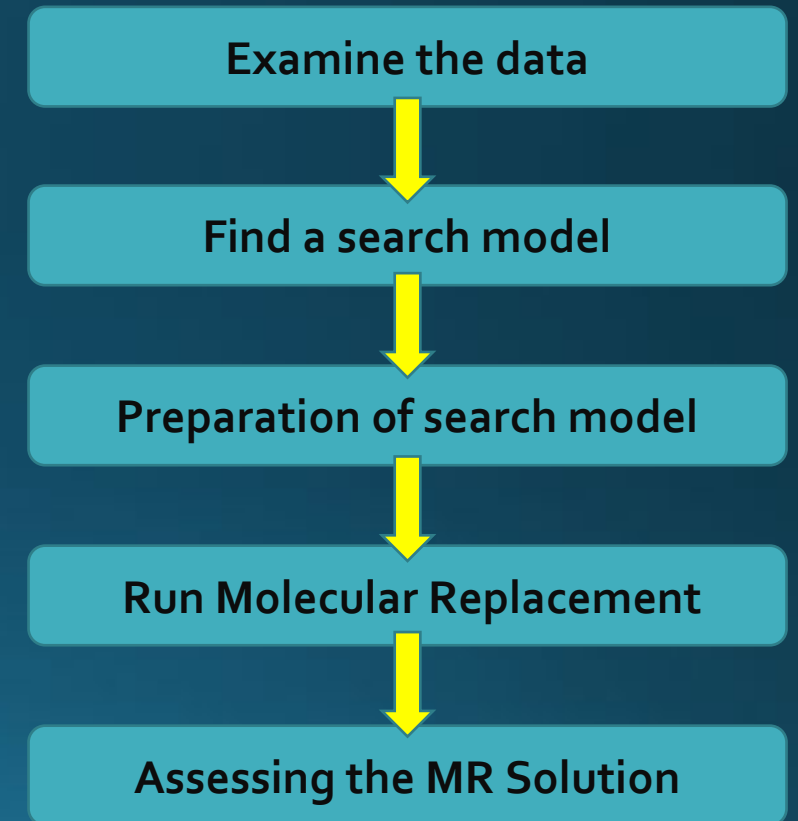
- Today, the vast majority of entries in the PDB have been phased using MR
- Increasing success rate of MR has been due to:
 - Improving methods in software e.g. maximum likelihood in Phaser
 - Increasing availability of suitable search models



(image from Fei Long)

Introduction: *Step-by-step Molecular Replacement*

- Despite this, successfully performing MR depends on attention to detail both before and after running the Molecular Replacement program



Step-by-Step MR: *Examining the data*

- Data issues can have an impact on how well MR will work

Examine the data

Step-by-Step MR: *Examining the data*

- Data issues can have an impact on how well MR will work
- Things to think about:
 - How many copies in the asymmetric unit?
 - What's the resolution of the data?
 - Self-rotation function – signs of NCS?

Examine the data

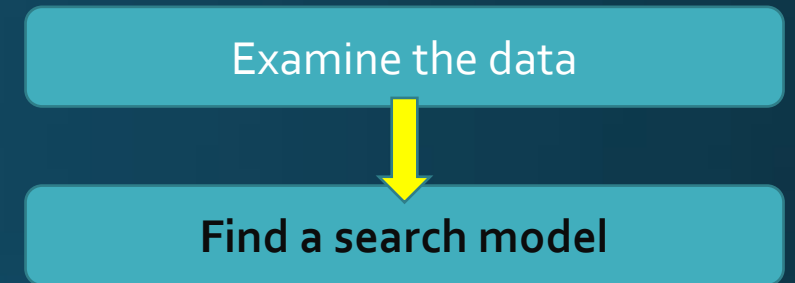
Step-by-Step MR: *Examining the data*

- Data issues can have an impact on how well MR will work
- Things to think about:
 - How many copies in the asymmetric unit?
 - What's the resolution of the data?
 - Self-rotation function – signs of NCS?
- Potential problems:
 - Pseudo translation?
 - Twinned data?
 - Space group assignment correct?

Examine the data

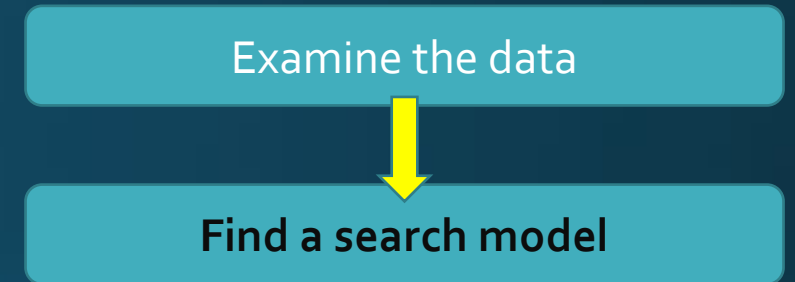
Step-by-Step MR: *Finding a search model*

- A well chosen and optimized search model has two key advantages



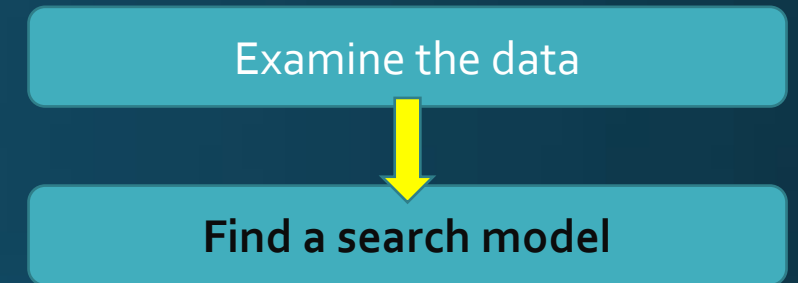
Step-by-Step MR: *Finding a search model*

- A well chosen and optimized search model has two key advantages
 1. Makes it easier for the MR program to find the correct position for that search model and hence sensible starting phases



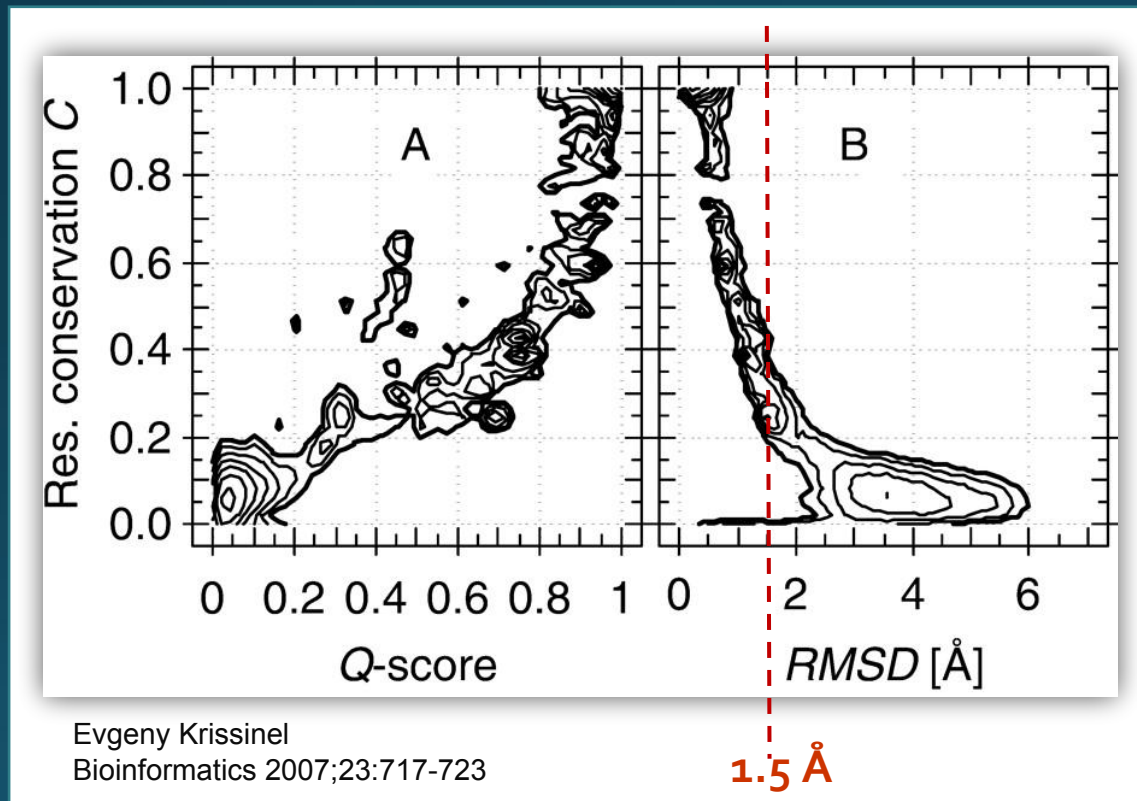
Step-by-Step MR: *Finding a search model*

- A well chosen and optimized search model has two key advantages
 1. Makes it easier for the MR program to find the correct position for that search model and hence sensible starting phases
 2. Facilitates quicker and easier model building and refinement post MR



Step-by-Step MR: *Finding a search model*

- How to find a search model?
 - Amino acid sequence similarity often correlates well with structural similarity



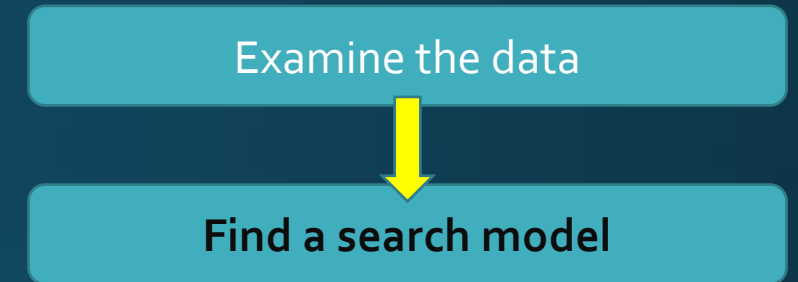
Examine the data



Find a search model

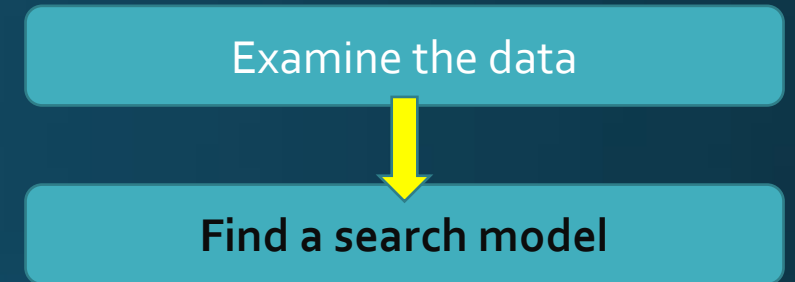
Step-by-Step MR: *Finding a search model*

- Considerations
 - Target Size:
 - The bigger the structure the more likely there will be sizeable conformational changes
 - Data Resolution :
 - Lower than 2.5 Angstroms – search model should represent large fraction of target molecule
 - Between 1.0 and 2.5 Angstroms – search model can be small fragment of target – DM and model building can improve phases
 - Better than 1.0 Angstrom – search model can be single atom (McCoy et al. 2017, PNAS)
 - Homologue Sequence identity < 30%
 - Suitable search models may exist – sequence identity calculation is difficult



Step-by-Step MR: *Finding a search model*

- The field of Bioinformatics has given us a wealth of sophisticated and sensitive sequence-based search tools for finding potential homologues from which we can derive search models for MR



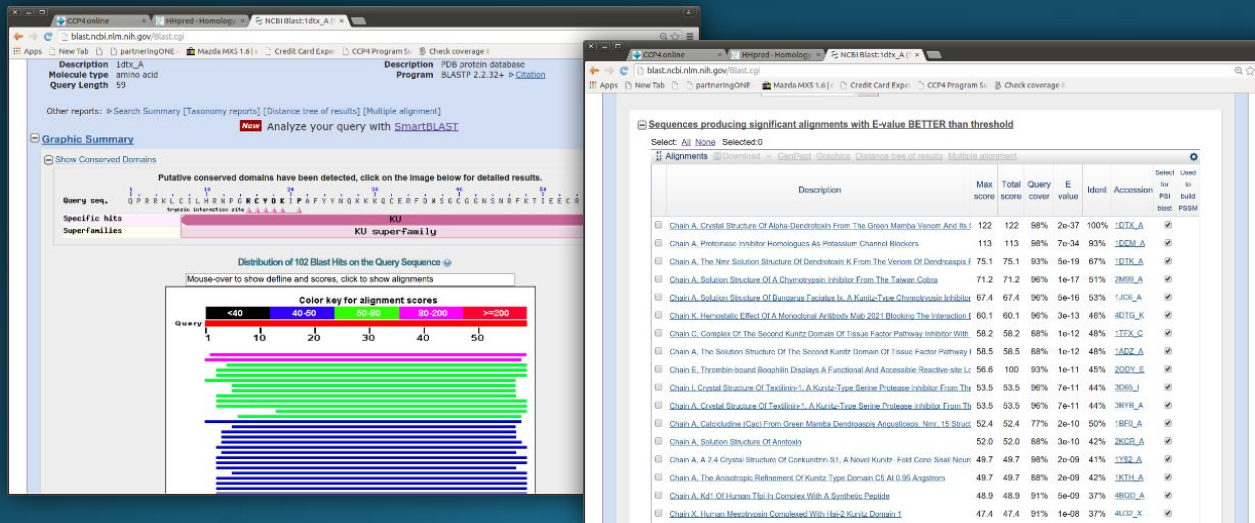
Step-by-Step MR: *Finding a search model*

- PSI-Blast
 - Profile-based searching
 - Online server, fast
 - Works well at finding suitable homologues down to sequence identities of 30%

Examine the data



Find a search model



Step-by-Step MR: *Finding a search model*

- HHpred
 - Hidden Markov Model approach
 - Suitable for more difficult cases where no obvious homologue is available

Examine the data

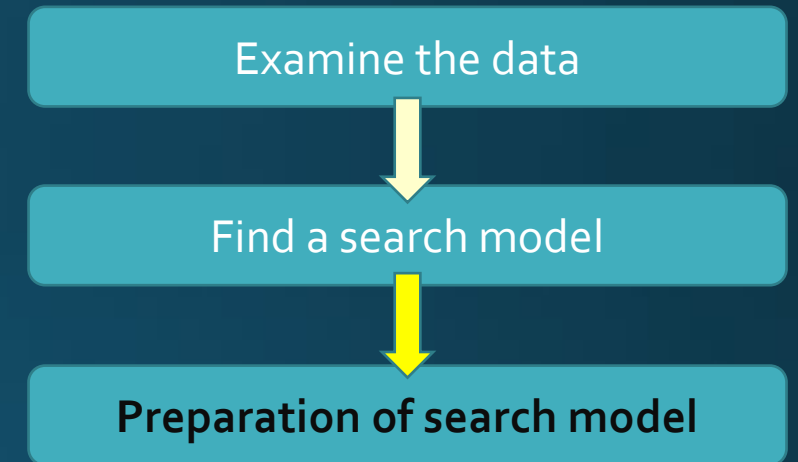


Find a search model

The image shows two screenshots of the HHpred web interface. The left screenshot displays the 'HHpred - Results' page for job ID 7132694, showing a sequence alignment between the query and a template. The right screenshot displays the 'Results' page for job ID 7132694, showing a list of search results with sequence alignments and scores.

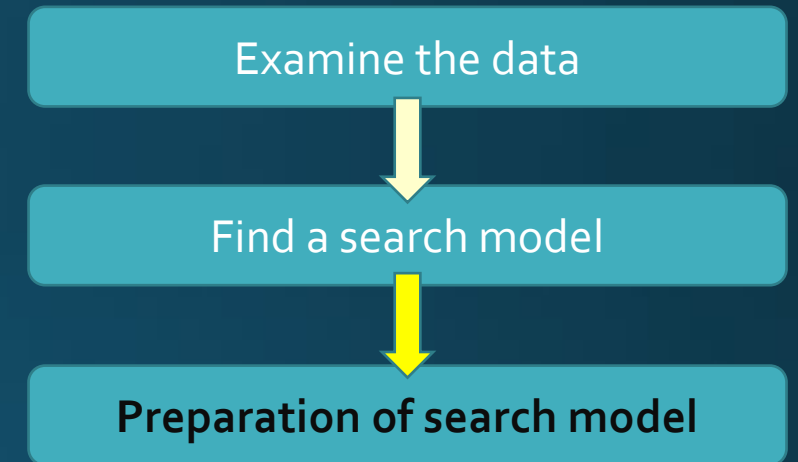
Step-by-Step MR: *Preparation of search model*

- Once a suitable homologue or set of homologues have been found they need to be prepared for use as search models



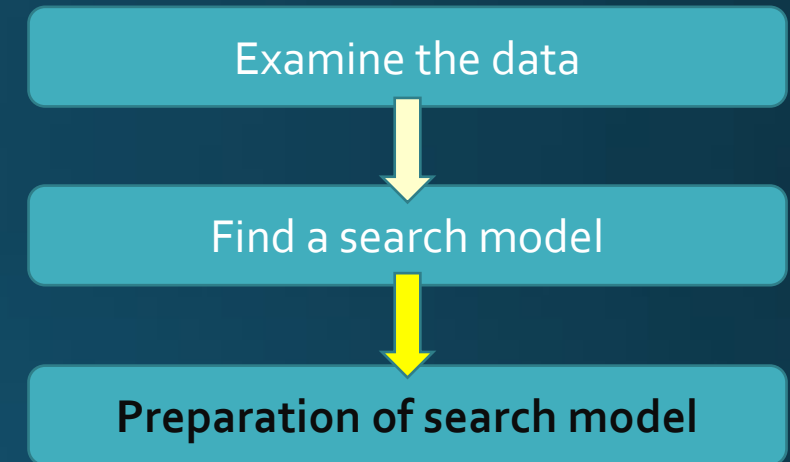
Step-by-Step MR: *Preparation of search model*

- Once a suitable homologue or set of homologues have been found they need to be prepared for use as search models
- The closer we can make the homologue to the target in terms of structural similarity the more likely it is to be successfully positioned in Molecular Replacement



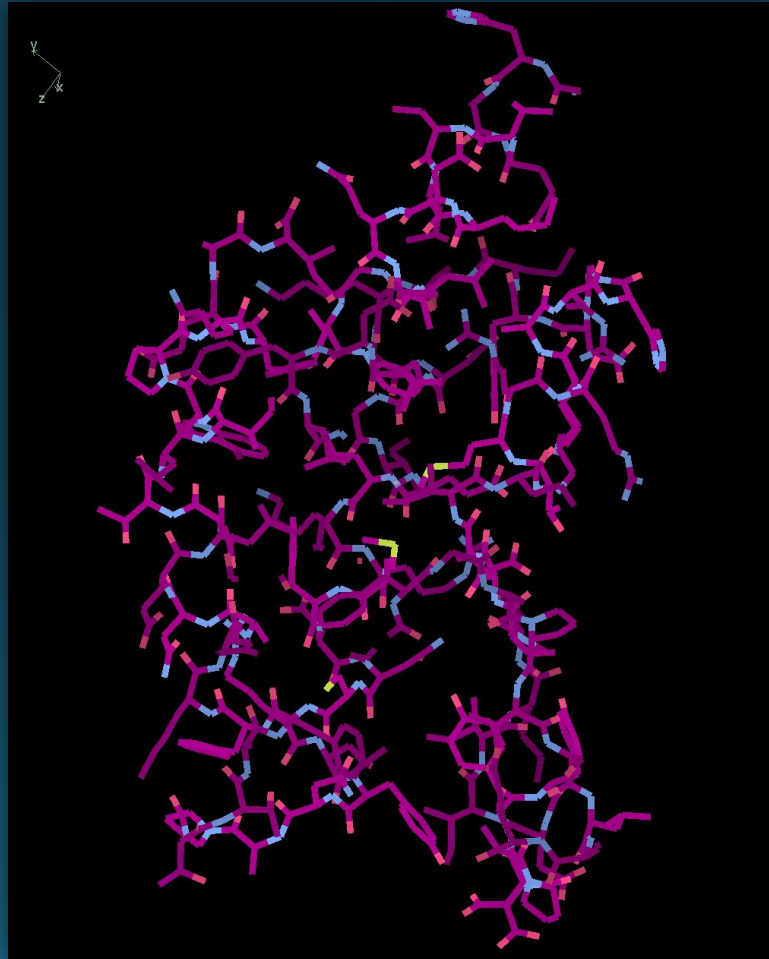
Step-by-Step MR: *Preparation of search model*

- Once a suitable homologue or set of homologues have been found they need to be prepared for use as search models
- The closer we can make the homologue to the target in terms of structural similarity the more likely it is to be successfully positioned in Molecular Replacement
- The sequence alignment generated in the search step can be used as a guide for the pruning of the homologue



Step-by-Step MR: *Preparation of search model*

- Example target:
 - 5TPX - Bromodomain from Plasmodium Faciparum Gcn5



Examine the data



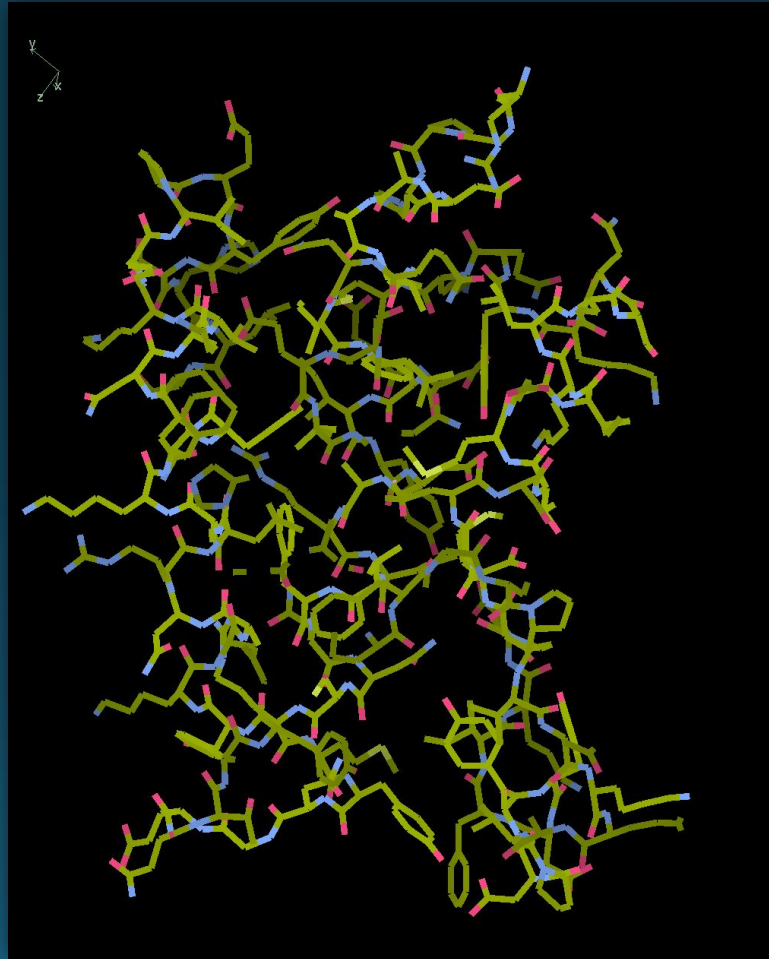
Find a search model



Preparation of search model

Step-by-Step MR: *Preparation of search model*

- Example target:
 - 5TPX - Bromodomain from Plasmodium Faciparum Gcn5
- Homologue:
 - 1E6I – chain A
 - 45% sequence identity over 82% coverage of target



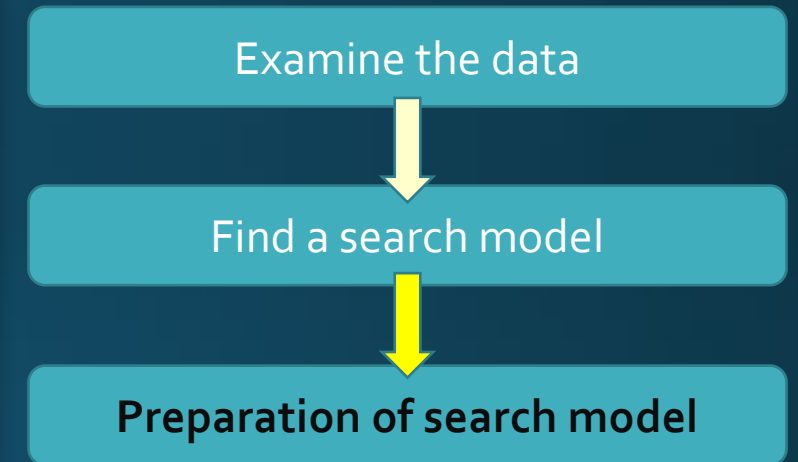
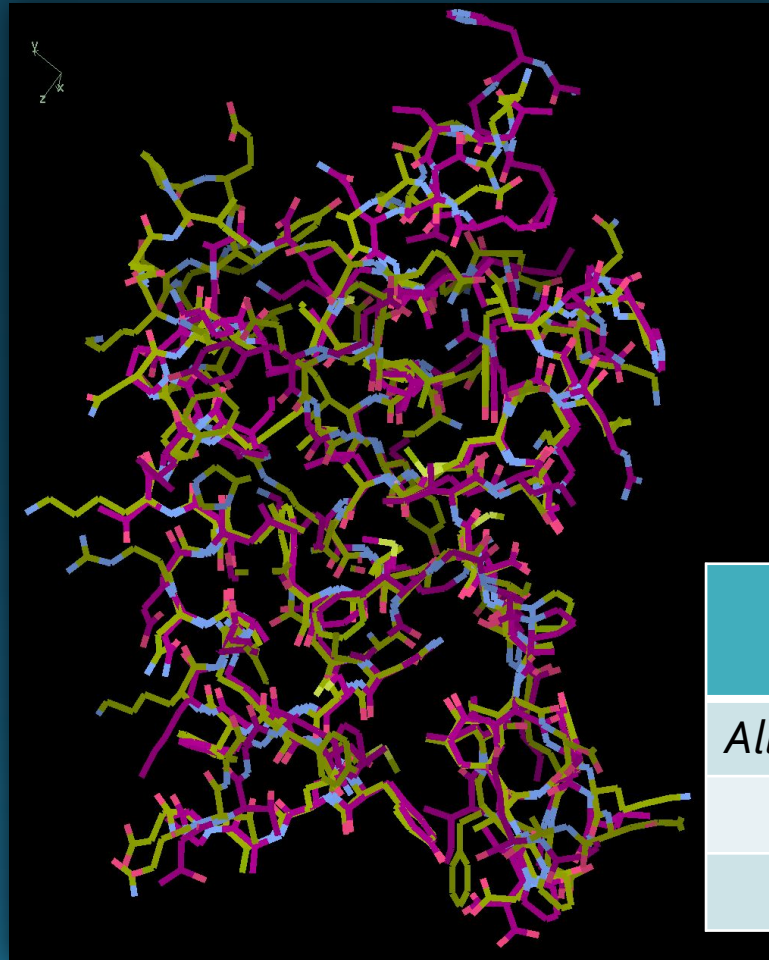
Examine the data

Find a search model

Preparation of search model

Step-by-Step MR: *Preparation of search model*

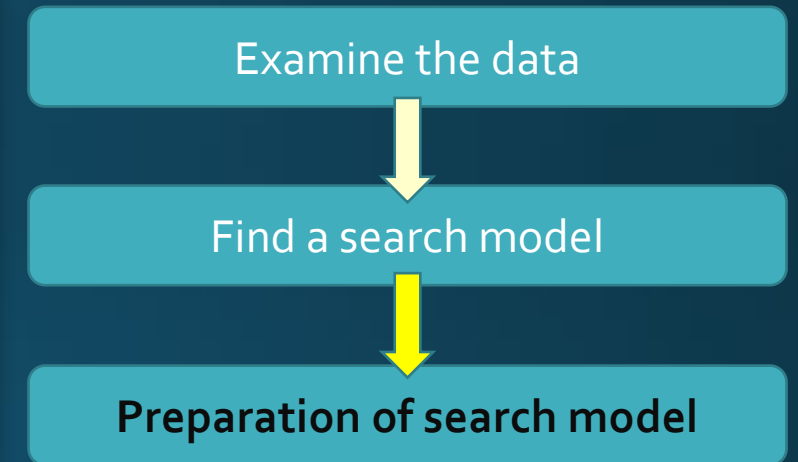
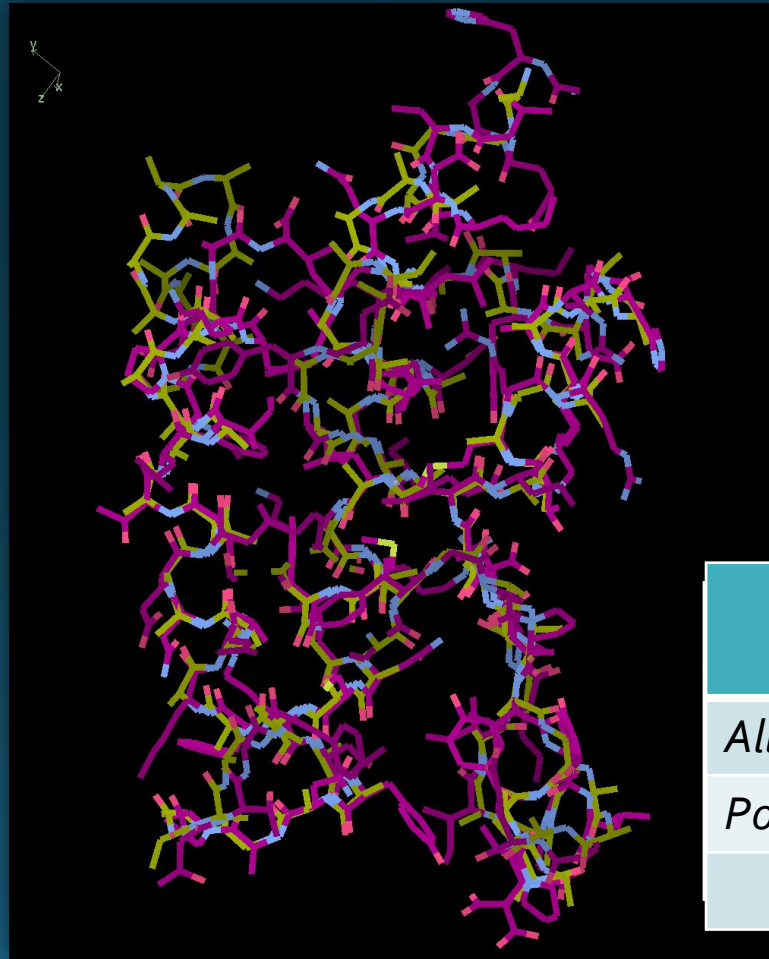
- Search Model preparation and MR:
 - All atoms?



	<i>Molrep</i> TF/sigma	<i>Refmac</i> Rfactor	<i>Refmac</i> Rfree
<i>All atoms</i>	6.82	0.418	0.481

Step-by-Step MR: *Preparation of search model*

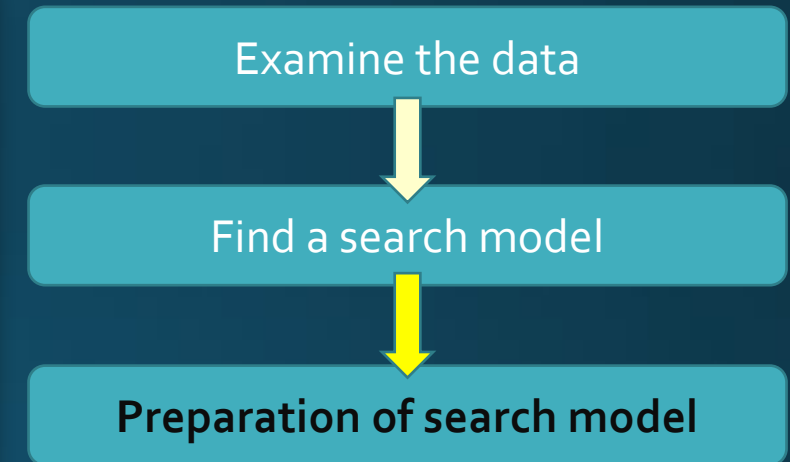
- Search Model preparation and MR:
 - All atoms?
 - Polyalanine?



	<i>Molrep</i> TF/sigma	<i>Refmac</i> Rfactor	<i>Refmac</i> Rfree
<i>All atoms</i>	6.82	0.418	0.481
<i>Poly</i>	4.98	0.46	0.482

Step-by-Step MR: *Preparation of search model*

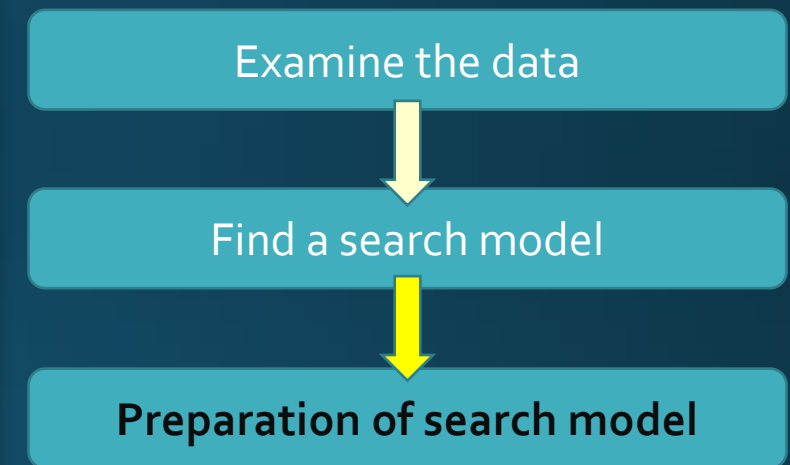
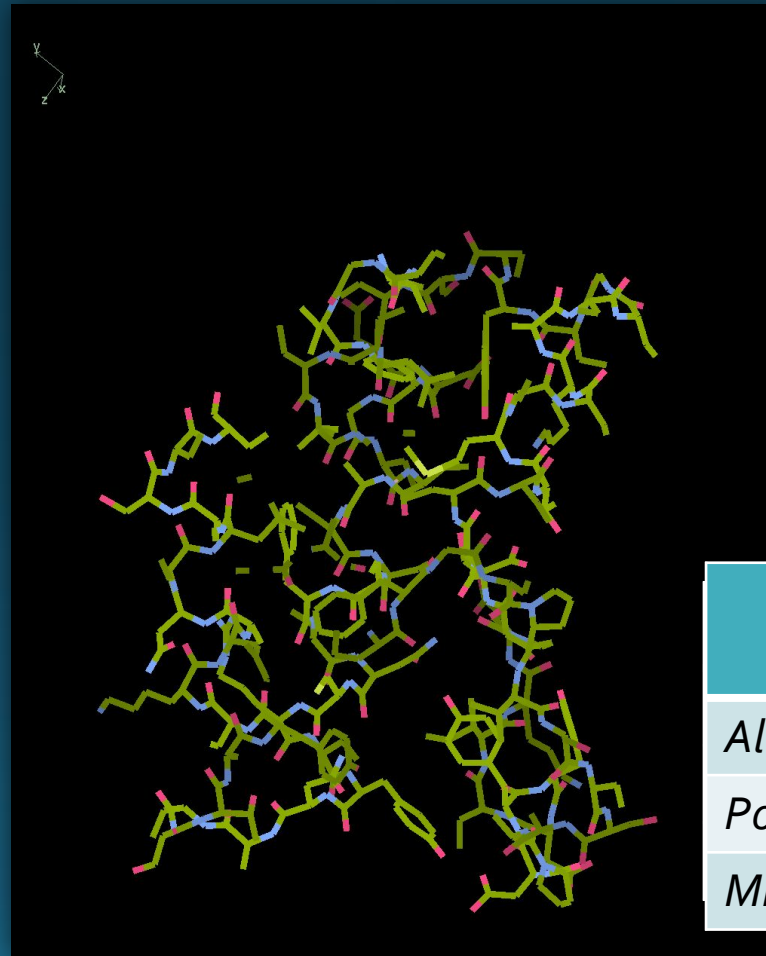
- Search Model preparation and MR:
 - All atoms?
 - Polyalanine?
 - Mixed model based on sequence alignment
 - Keep common residues
 - Truncate to common atoms on aligned residues



	<i>Molrep</i> TF/sigma	<i>Refmac</i> Rfactor	<i>Refmac</i> Rfree
<i>All atoms</i>	6.82	0.418	0.481
<i>Poly</i>	4.98	0.46	0.482
<i>Mixed</i>	10.27	0.379	0.418

Step-by-Step MR: *Preparation of search model*

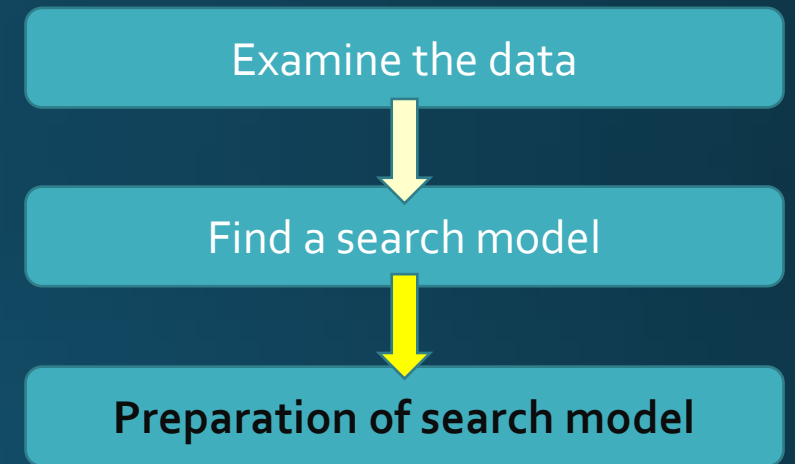
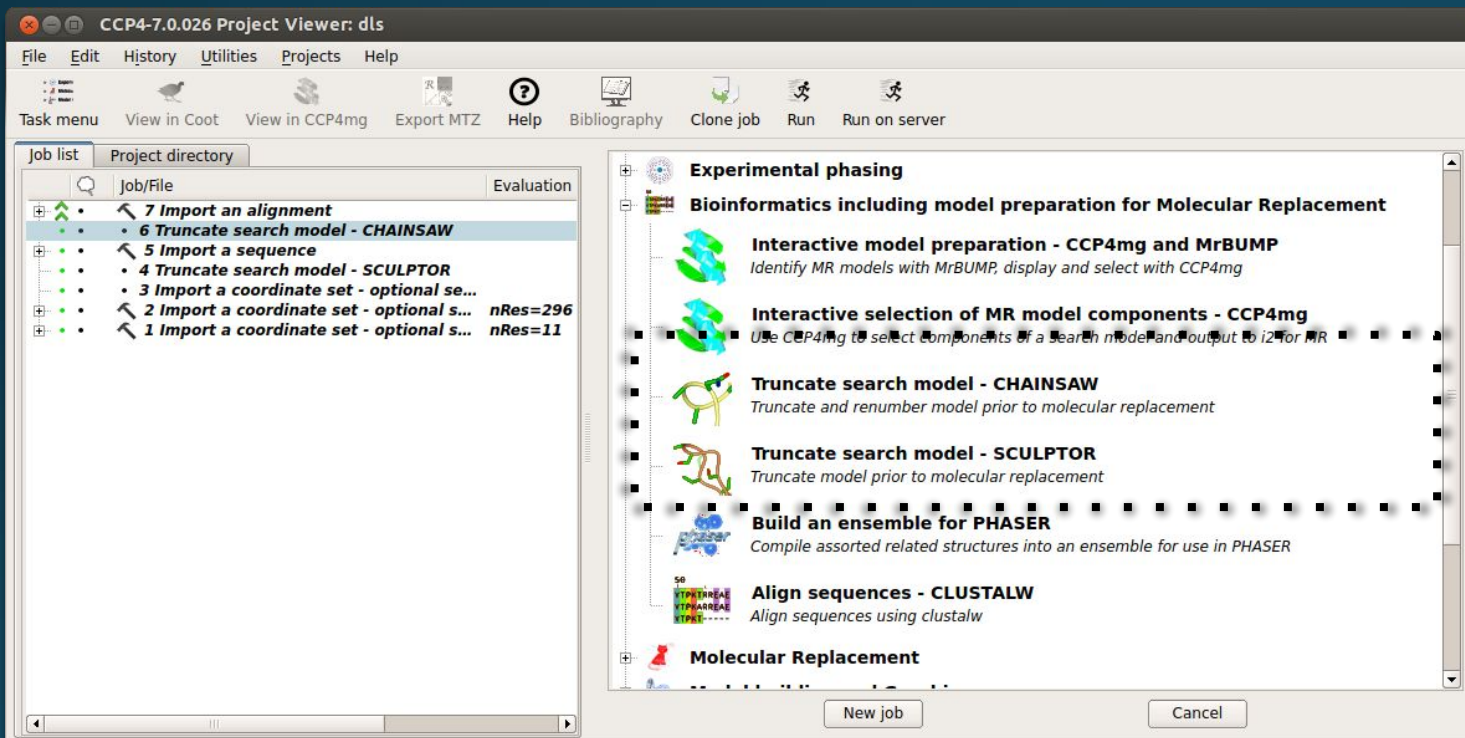
- Search Model preparation and MR:
 - All atoms?
 - Polyalanine?
 - Mixed model based on sequence alignment
 - Keep common residues
 - Truncate to common atoms on aligned residues



	<i>Molrep</i> TF/sigma	<i>Refmac</i> Rfactor	<i>Refmac</i> Rfree
<i>All atoms</i>	6.82	0.418	0.481
<i>Poly</i>	4.98	0.46	0.482
<i>Mixed</i>	10.27	0.379	0.418

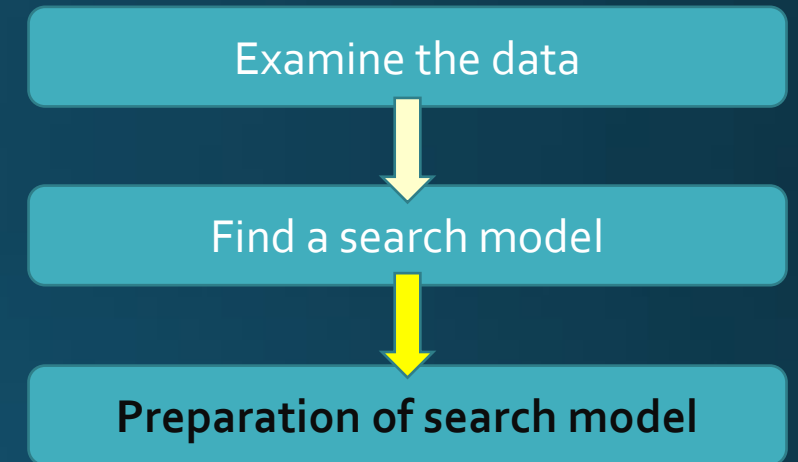
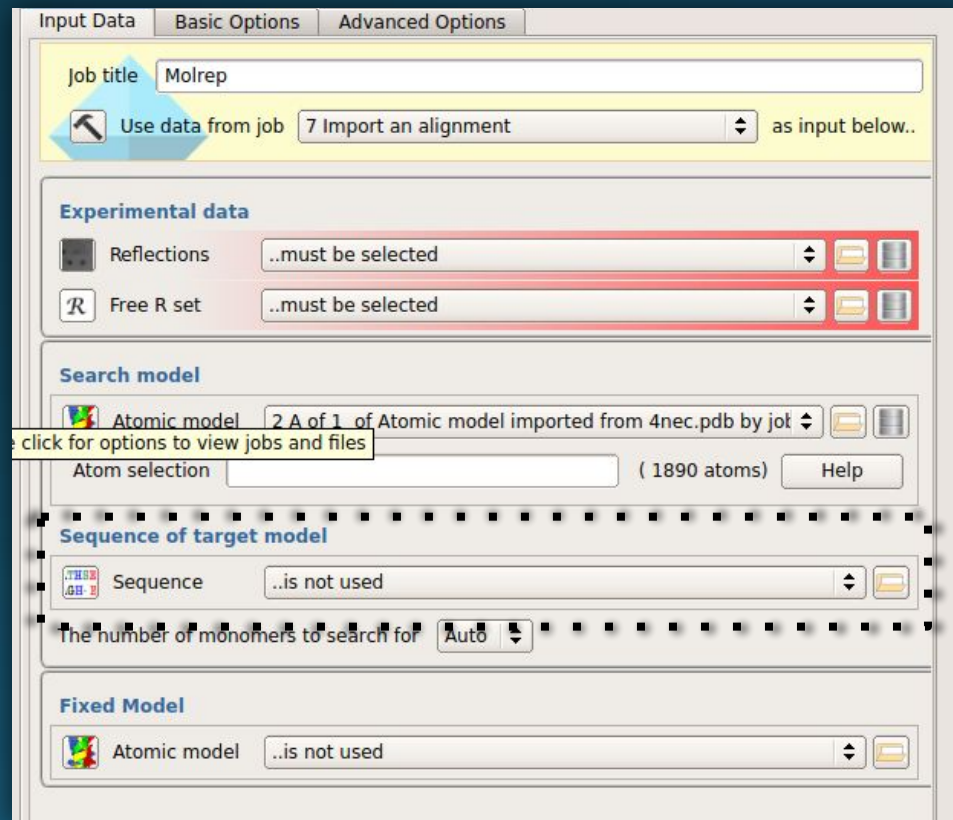
Step-by-Step MR: *Preparation of search model*

- Mixed-model generation in CCP4
- Chainsaw & Sculptor



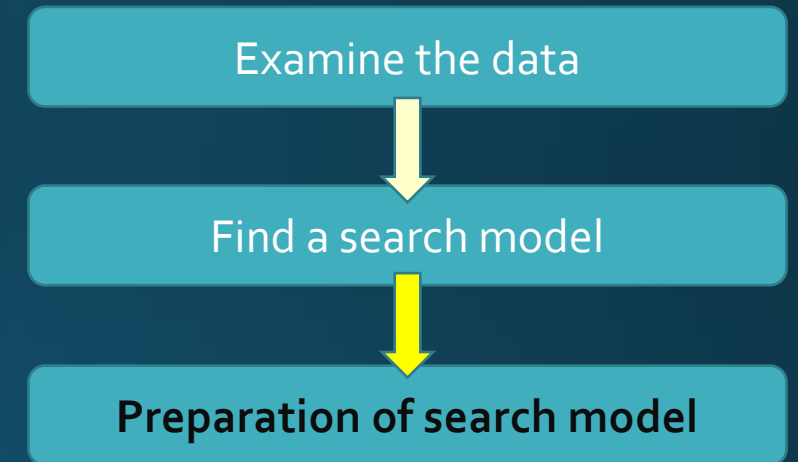
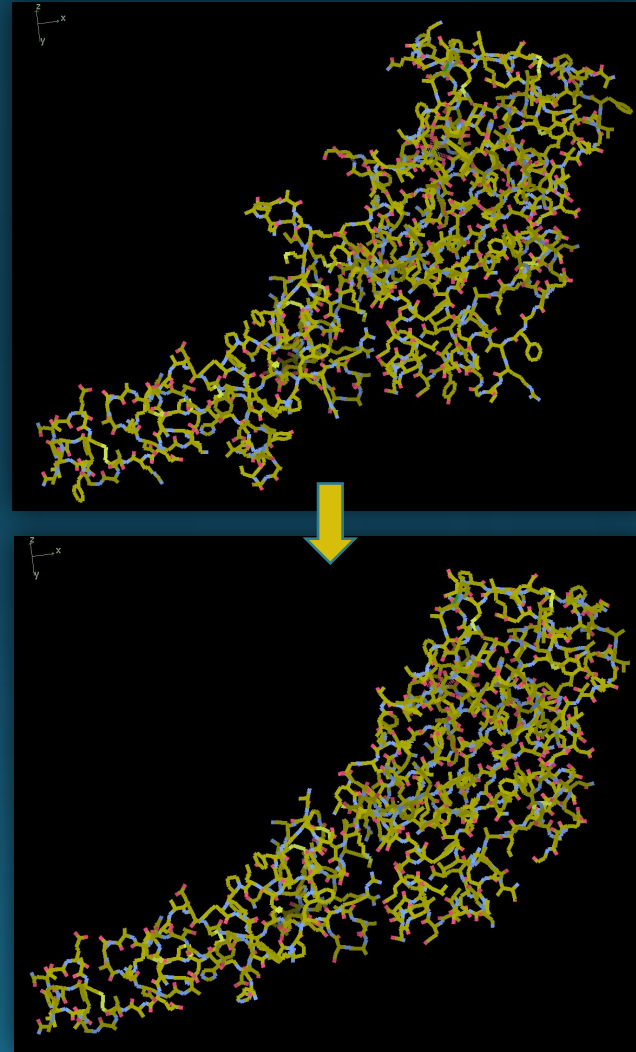
Step-by-Step MR: *Preparation of search model*

- Mixed-model generation in CCP4
 - Molrep



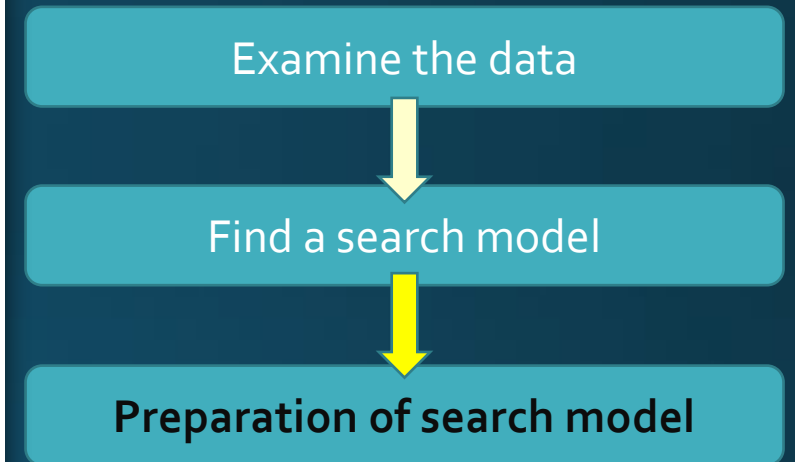
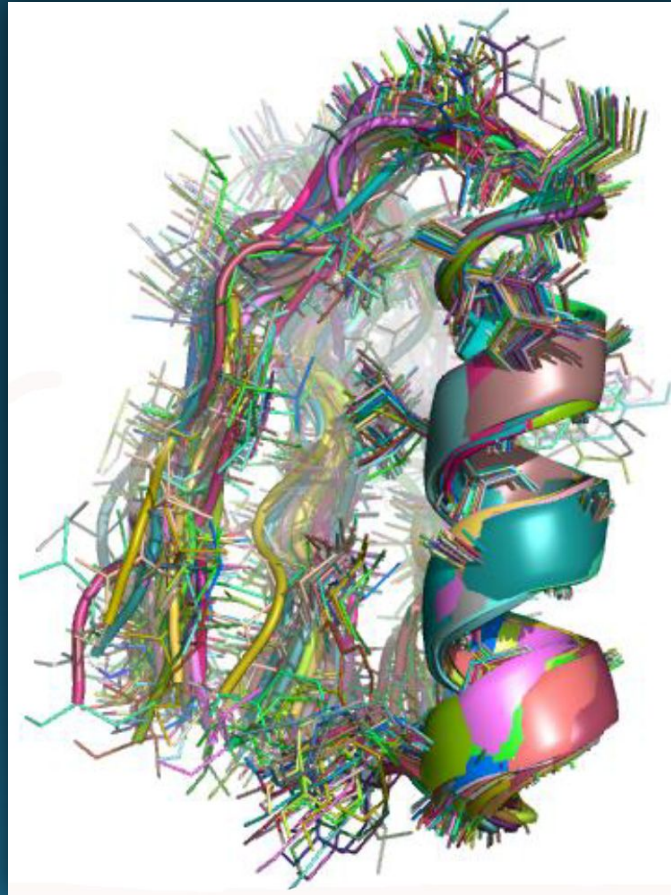
Step-by-Step MR: *Preparation of search model*

- Manual pruning in Coot
- Remove anything that is likely to be flexible:
 - External loops
 - Longer side chains on the surface of the molecule (e.g. LYS)
 - Domains if there is evidence of “hinge motion”



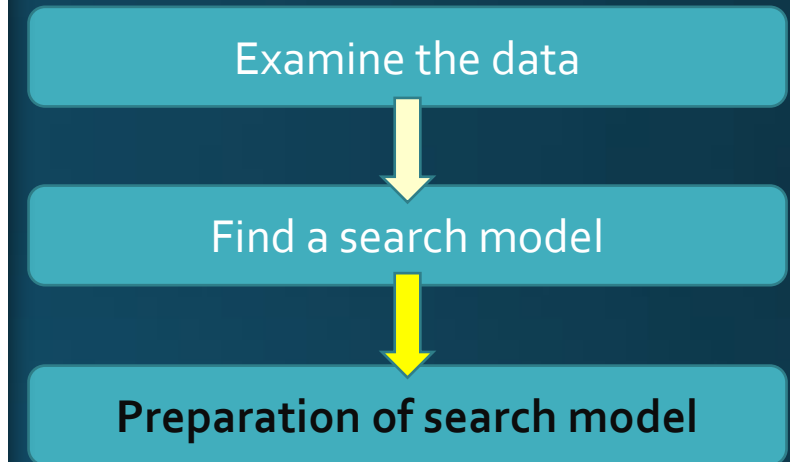
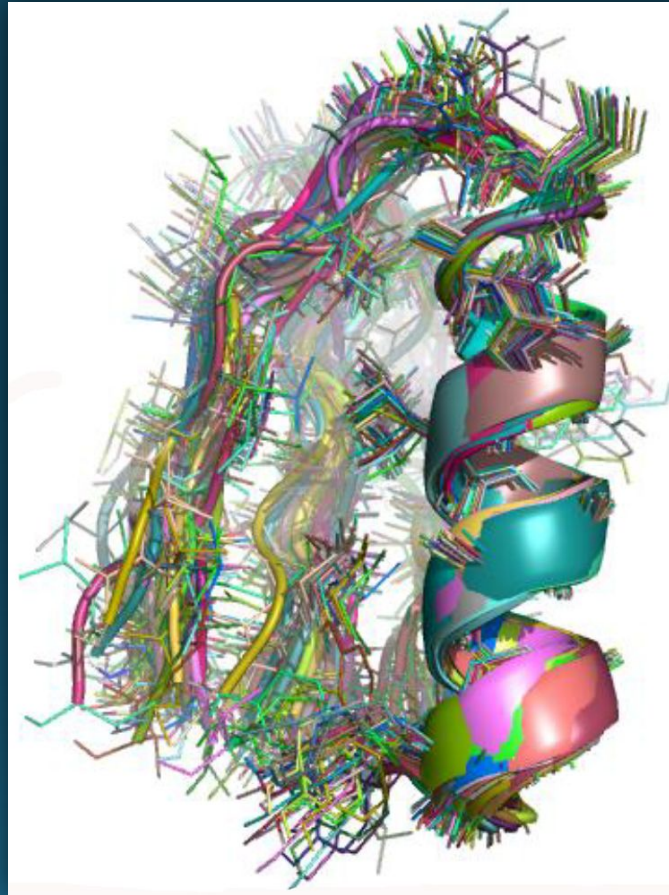
Step-by-Step MR: *Preparation of search model*

- Ensembles: alignment of search models



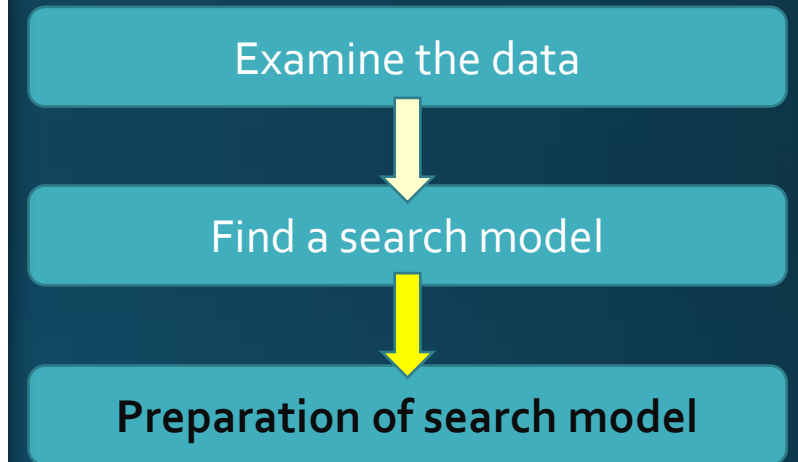
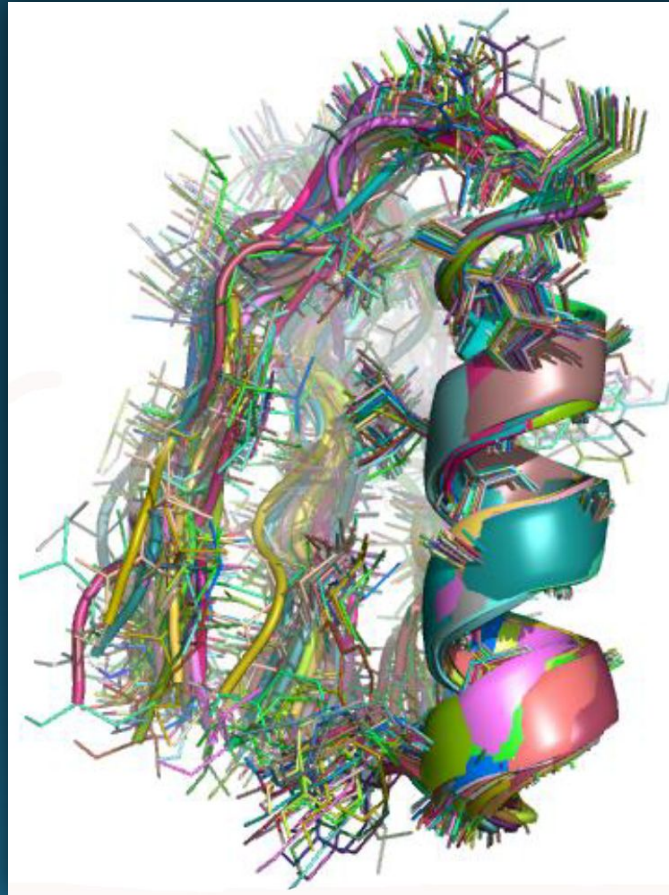
Step-by-Step MR: *Preparation of search model*

- Ensembles: alignment of search models
 1. Variance across members can guide experimental data weighting in Phaser



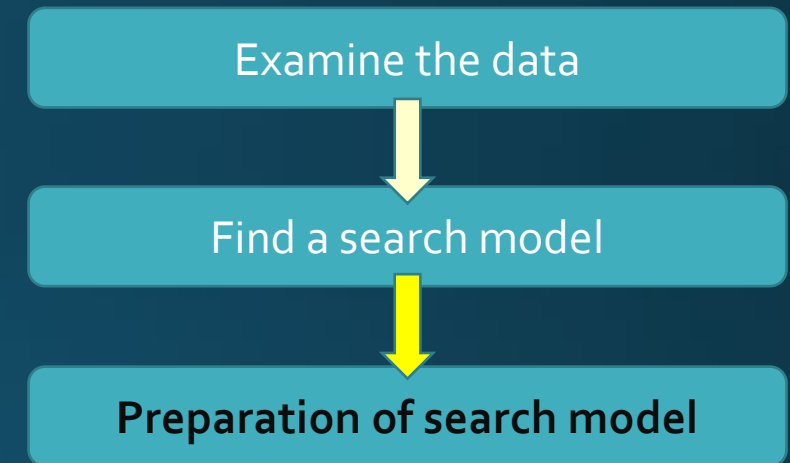
Step-by-Step MR: *Preparation of search model*

- Ensembles: alignment of search models
 1. Variance across members can guide experimental data weighting in Phaser
 2. Truncation based on alignment variance can identify conserved regions or core



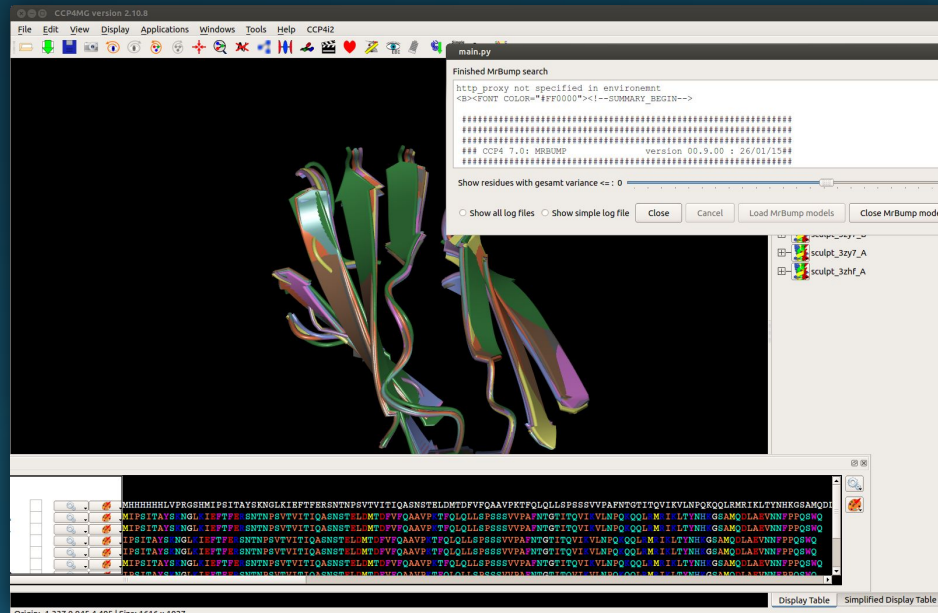
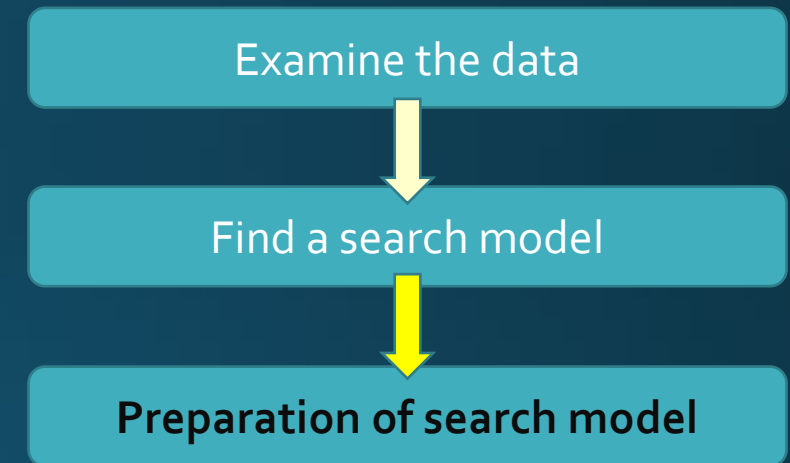
Step-by-Step MR: *Preparation of search model*

- Ensembles: Generating ensembles in CCP4
 - Ensembler:
 - Align models and truncate based on a variance threshold
 - General alignment tools:
 - Gesamt
 - Superpose
 - Can also be run through CCP4mg and Coot graphical interfaces



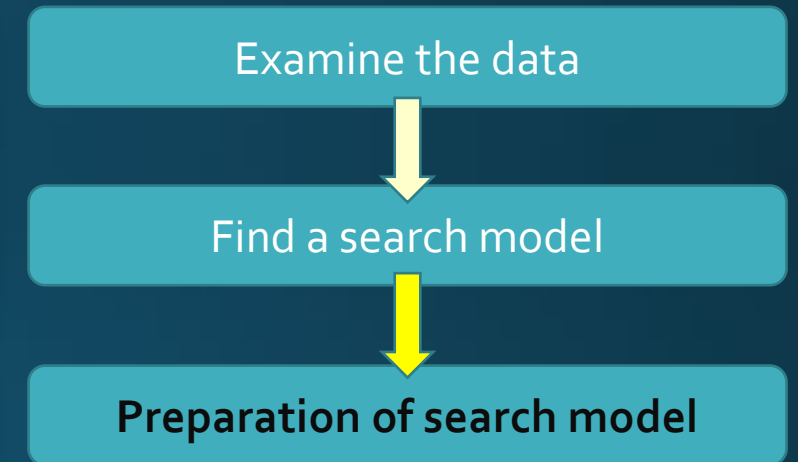
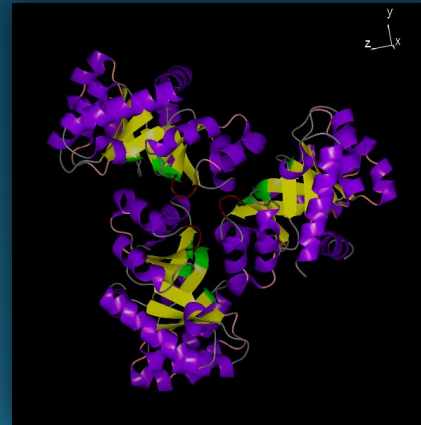
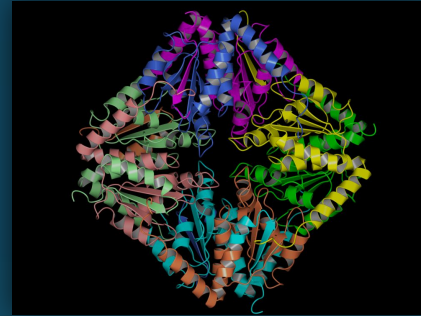
Step-by-Step MR: *Preparation of search model*

- Ensembles: CCP4mg & MrBUMP
 - Finds homologues using Phmmer
 - Runs Chainsaw to prune and create search models
 - Aligns models using Gesamt
 - Truncation tool to adjust variance threshold



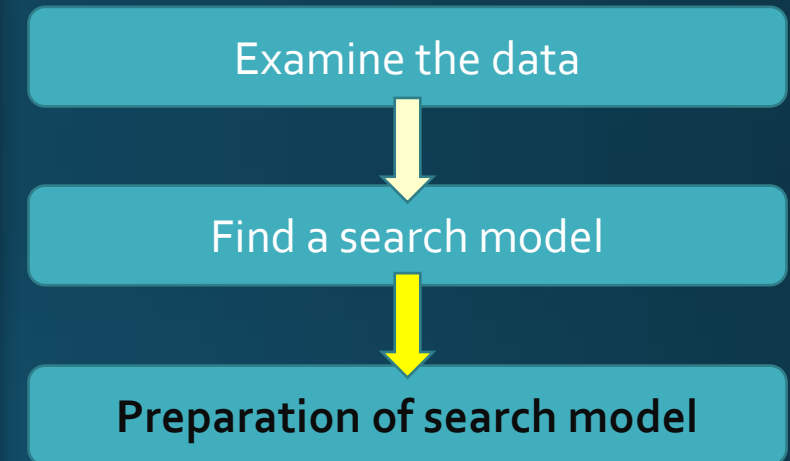
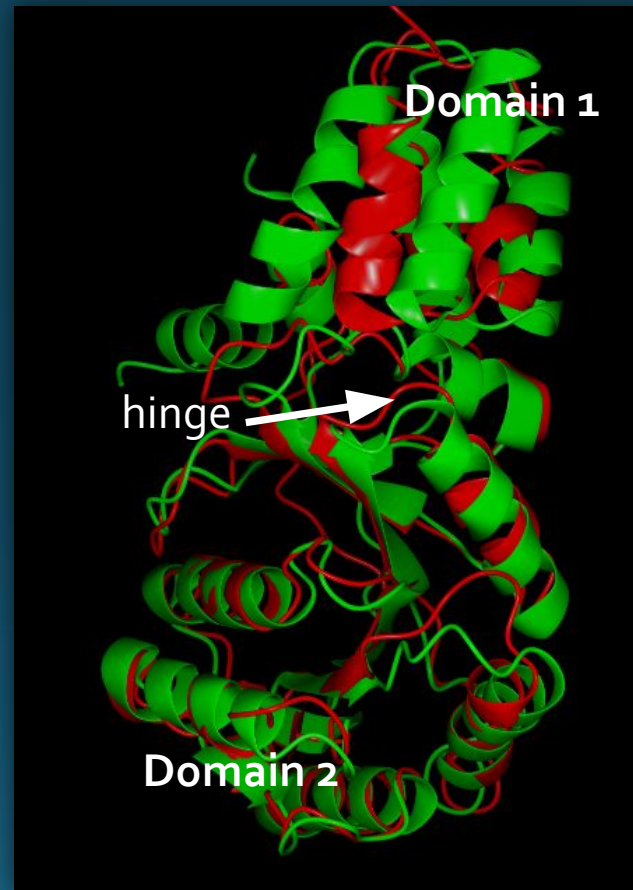
Step-by-Step MR: *Preparation of search model*

- Multimers as search models
 - A single chain search model can be too small if the target has crystallised in multimeric form
 - The signal for the correct position is too weak against the background noise of incorrect positions
 - Particularly a problem at lower resolutions and crystals with high symmetry



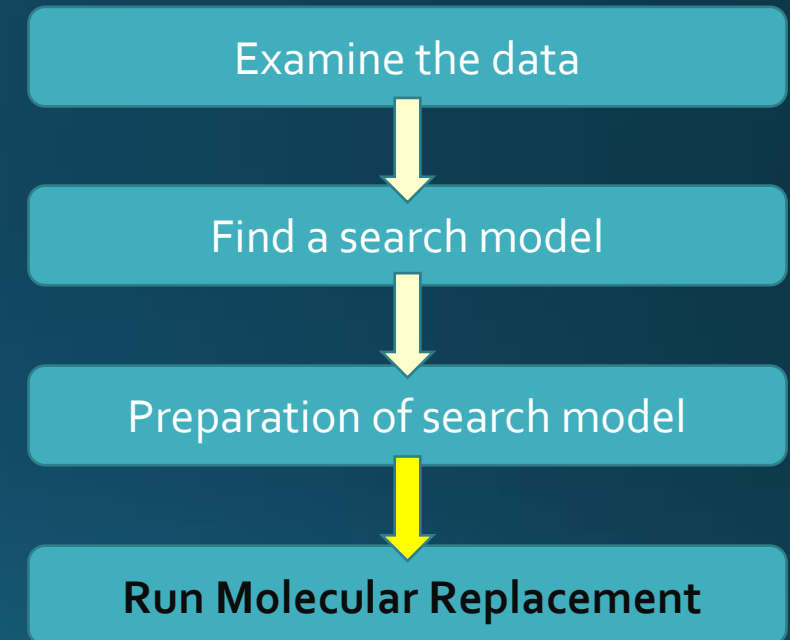
Step-by-Step MR: *Preparation of search model*

- Domains as search models
 - A hinge motion may alter the relative orientation of domains within a chain
 - Domain models should be isolated from the parent search model and used separately in MR



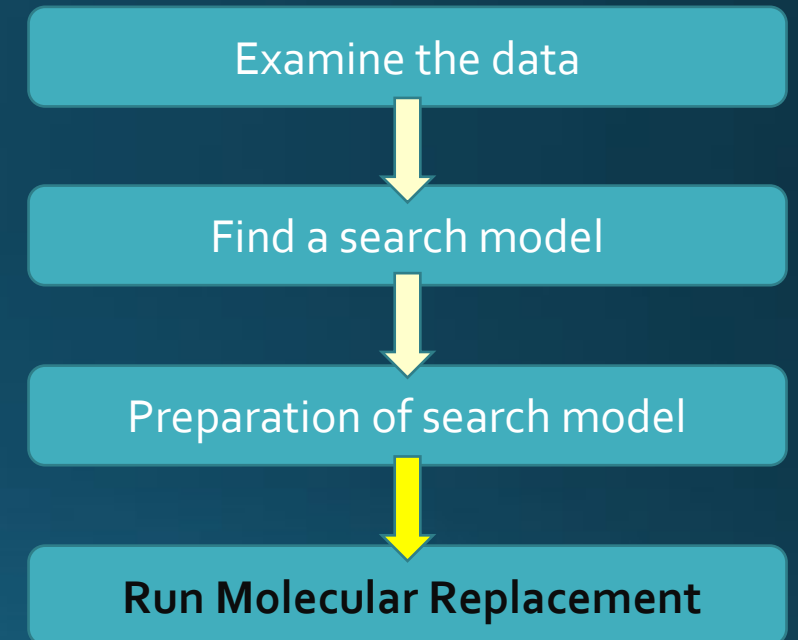
Step-by-Step MR: *Run Molecular Replacement*

- CCP4 has several programs for doing Molecular Replacement
 - Amore
 - Manual steps but very fast
 - Molrep
 - Automated MR
 - Several useful features e.g. searching a map
 - Phaser
 - Maximum likelihood approach
 - Accounts for potential model errors
 - Best for difficult cases and for correctly positioning fragment search models



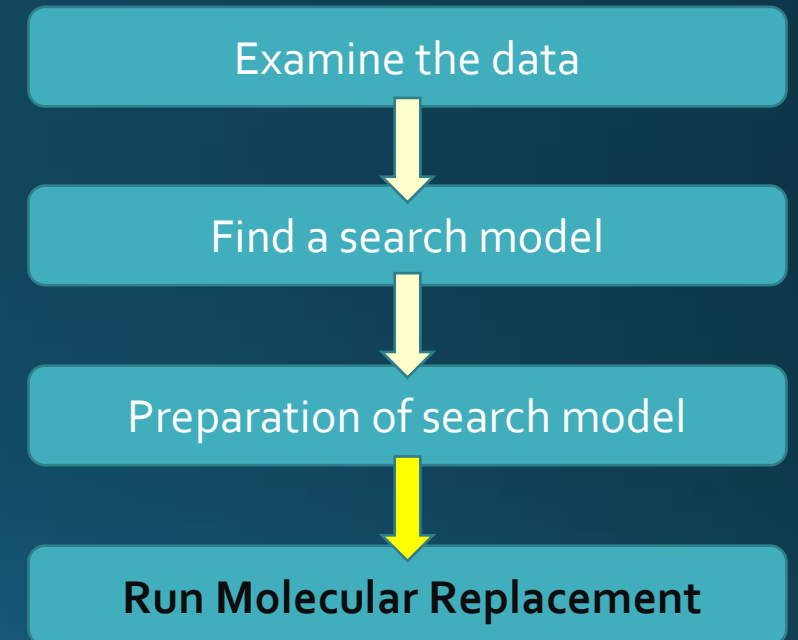
Step-by-Step MR: *Run Molecular Replacement*

- Important points on using Phaser



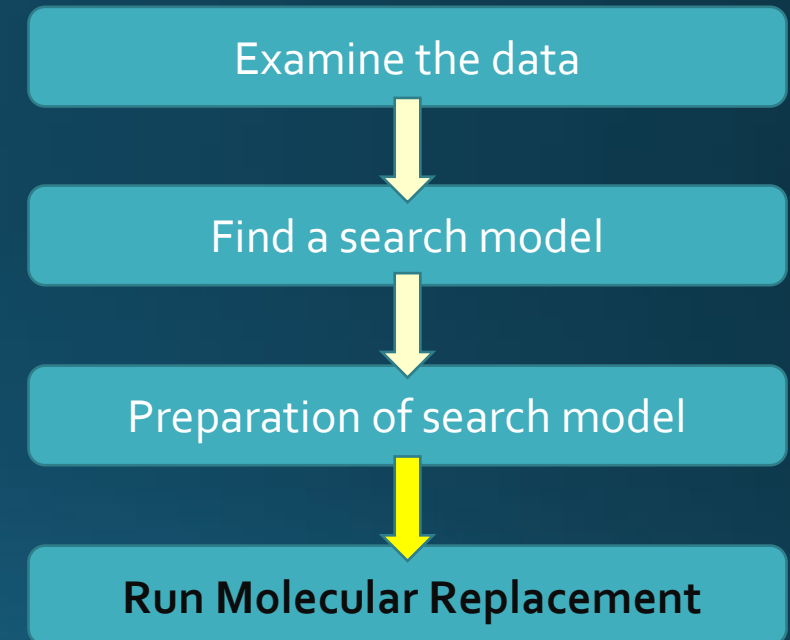
Step-by-Step MR: *Run Molecular Replacement*

- Important points on using Phaser
 - Phaser accounts for errors in:



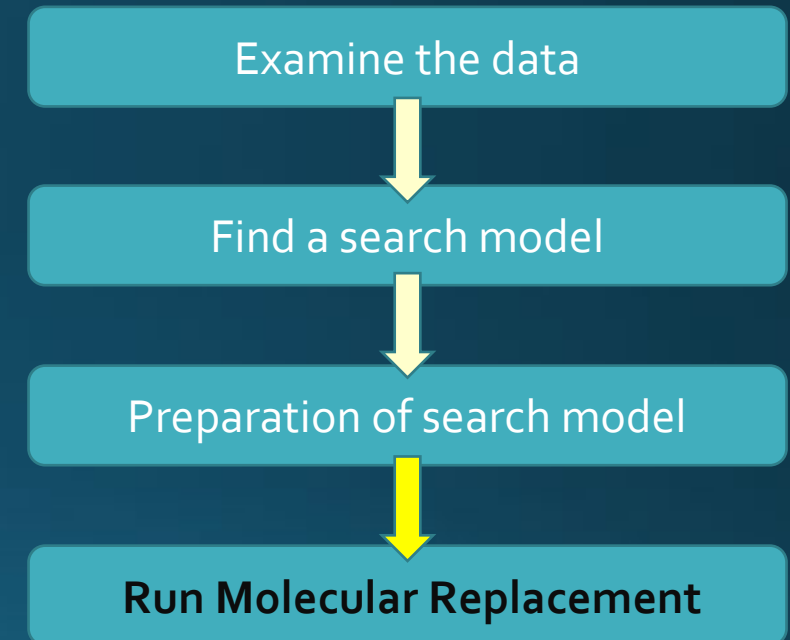
Step-by-Step MR: *Run Molecular Replacement*

- Important points on using Phaser
 - Phaser accounts for errors in:
 1. Model
 - Provide accurate details of AU composition
 - Use RMS value rather than sequence identity and try different values if first attempt doesn't work



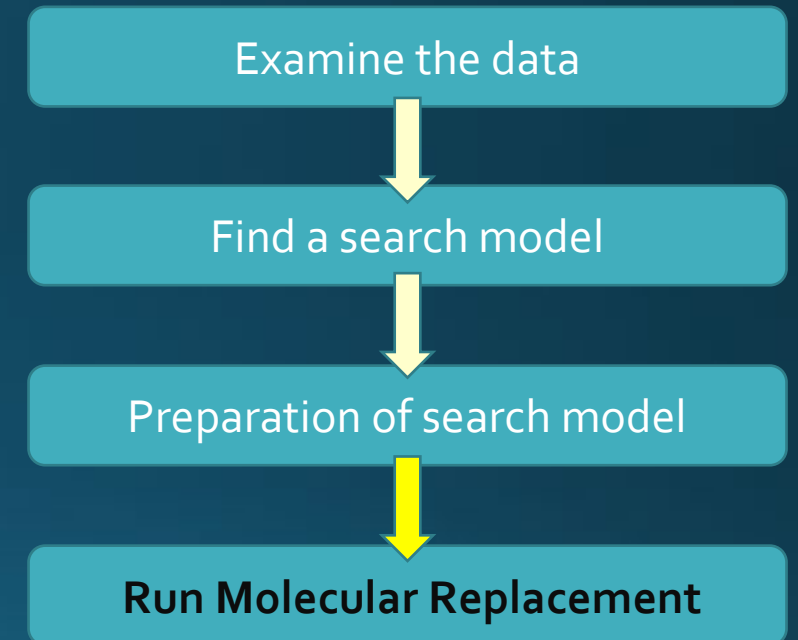
Step-by-Step MR: *Run Molecular Replacement*

- Important points on using Phaser
 - Phaser accounts for errors in:
 1. Model
 - Provide accurate details of AU composition
 - Use RMS value rather than sequence identity and try different values if first attempt doesn't work
 2. Data
 - Provide intensities – internally works out amplitudes accounting for experimental errors

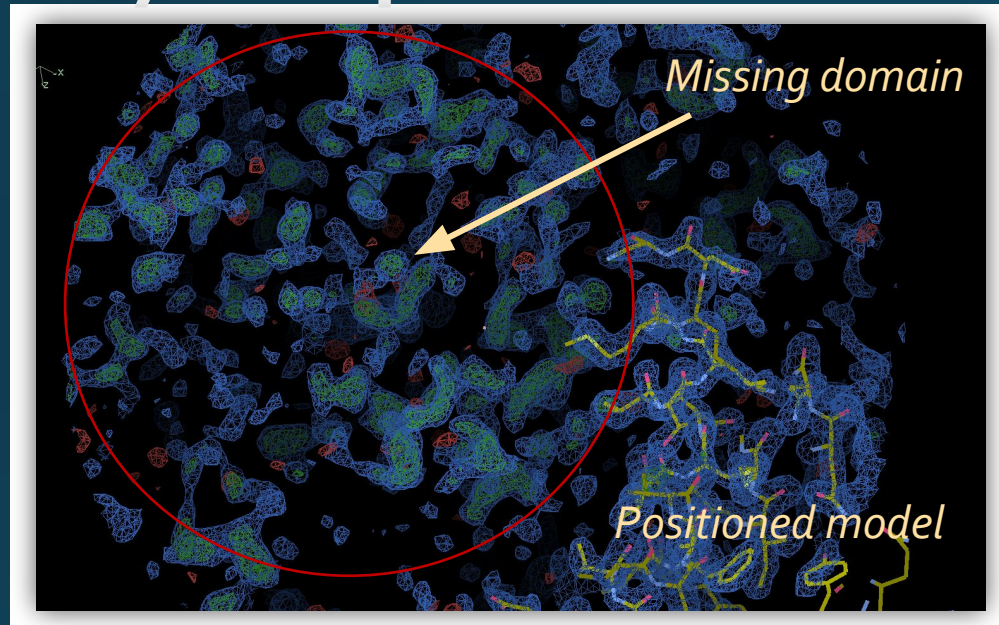


Step-by-Step MR: *Run Molecular Replacement*

- Important points on using Phaser
 - Phaser accounts for errors in:
 1. Model
 - Provide accurate details of AU composition
 - Use RMS value rather than sequence identity and try different values if first attempt doesn't work
 2. Data
 - Provide intensities – internally works out amplitudes accounting for experimental errors
 - Phaser performs clever decision making for automation
 - Provide minimal details and let Phaser make it's own decisions e.g. search order, search all possible space groups
 - If it doesn't work take step-by-step approach – 1 copy at a time

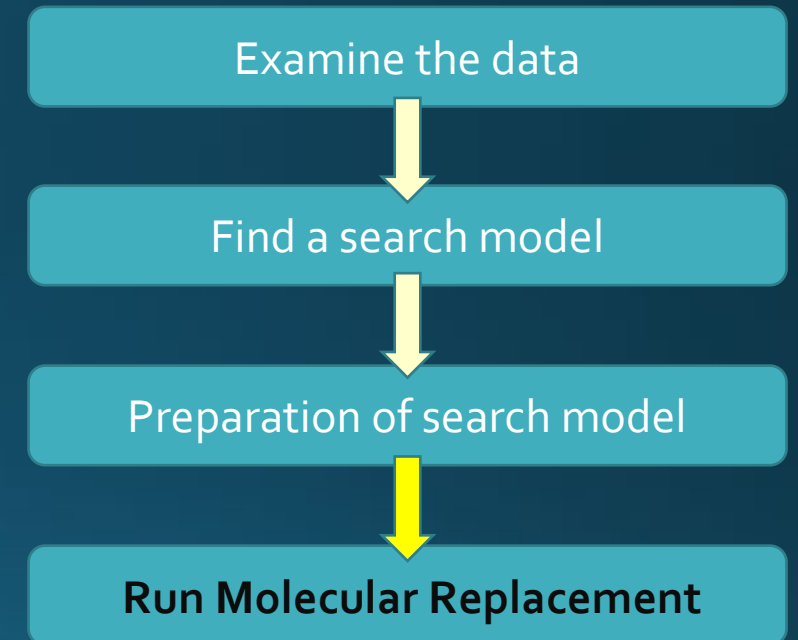


Step-by-Step MR: *Run Molecular Replacement*



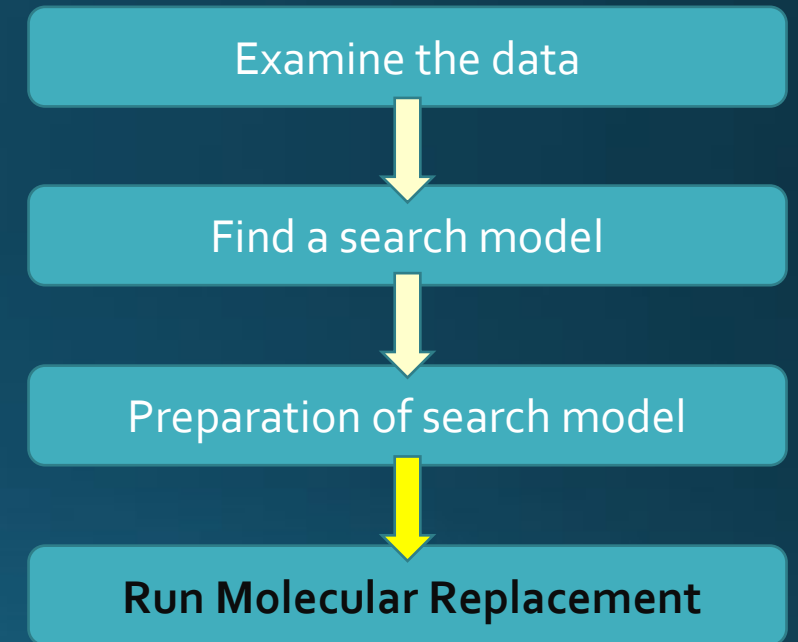
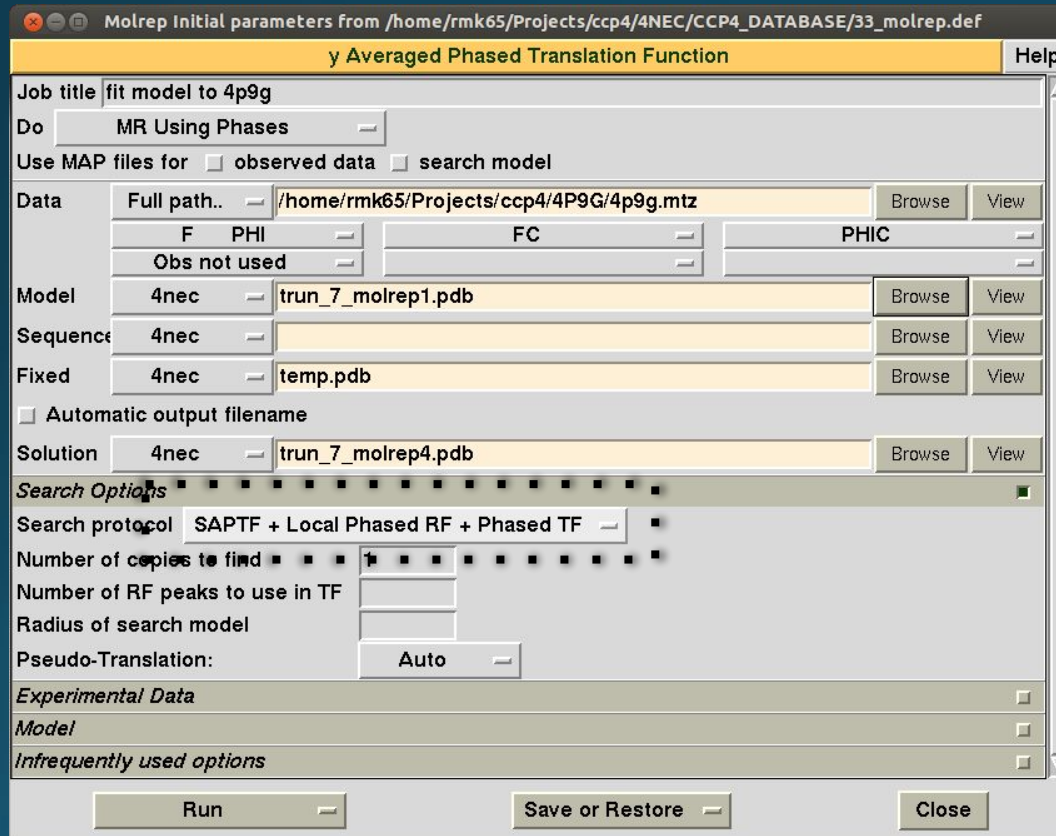
- Phased translation search

- Used when searching for several copies or dealing with a complex
- Available through MOLREP (3 protocols) and Phaser
- Often more successful than standard MR search approach particularly when looking for small domains



Step-by-Step MR: *Run Molecular Replacement*

- Phased translation search
 - From MOLREP (CCP4i):
 - 3 Protocols for phased search - Try all!



Step-by-Step MR: *Run Molecular Replacement*

- Phased translation search
 - From Phaser (CCP4i):

Maximum Likelihood Molecular Replacement Initial parameters from /home/rmk65/Projects/ample/concord

Space group read from mtz file 'P 21 21 21' ; Run Phaser with mtz space group and enantiomorph

Define ensembles (models)

Ensemble # 1

Ensemble Name ensemble1 Define ensemble via pdb file(s) Use HETATM atoms

PDB #1 Full path.. /home/rmk65/Projects/ample/concord/concord_13_merge Browse View

Similarity of PDB #1 to the target structure sequence identity 0.2

Edit list Add superimposed PDB file to the ensemble

Ensemble # 2

Ensemble Name ensemble2 Define ensemble via pdb file(s) Use HETATM atoms

PDB #1 concord 3c71_A.pdb Browse View

Similarity of PDB #1 to the target structure rms difference 0.2

Edit list Add superimposed PDB file to the ensemble

Edit list Add ensemble

Define composition of the asymmetric unit

Total scattering determined by components in asymmetric unit

Component 1 protein sequence file Number in asymmetric unit 2

SEQ file Full path.. /home/rmk65/Projects/ample/concord/data/3DCY/3dcy.fasta Browse View

Edit list Define another component

Search parameters

Perform search using ensemble2 Number of copies to search for 1

Edit list Add another search

Additional Search parameters

Allow search with alternative ensembles (models) for a single component of ASU off

Define any known partial structure Fix input coordinates Ensemble: ensemble1

Rotation function search target Fast

Cluster rotation peaks before passing to translation function on Turn OFF for Brute Search

Translation function search target Phased

User parameters

Output control

Expert parameters

Developer parameters

Run Save or Restore Close

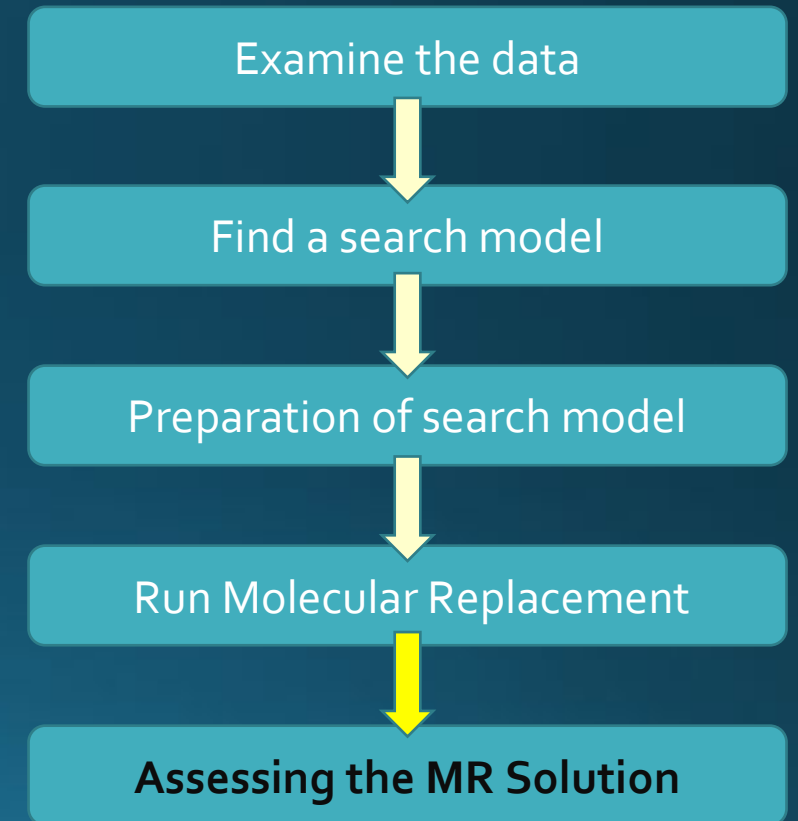
Examine the data

Find a search model

Preparation of search model

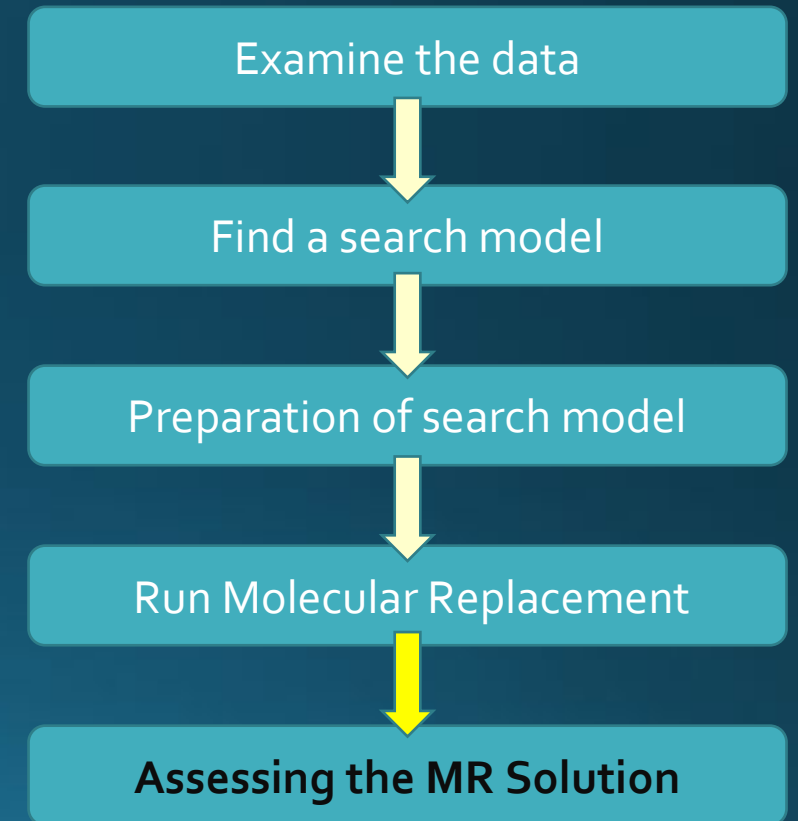
Run Molecular Replacement

Step-by-Step MR: *Assessing the MR solution*



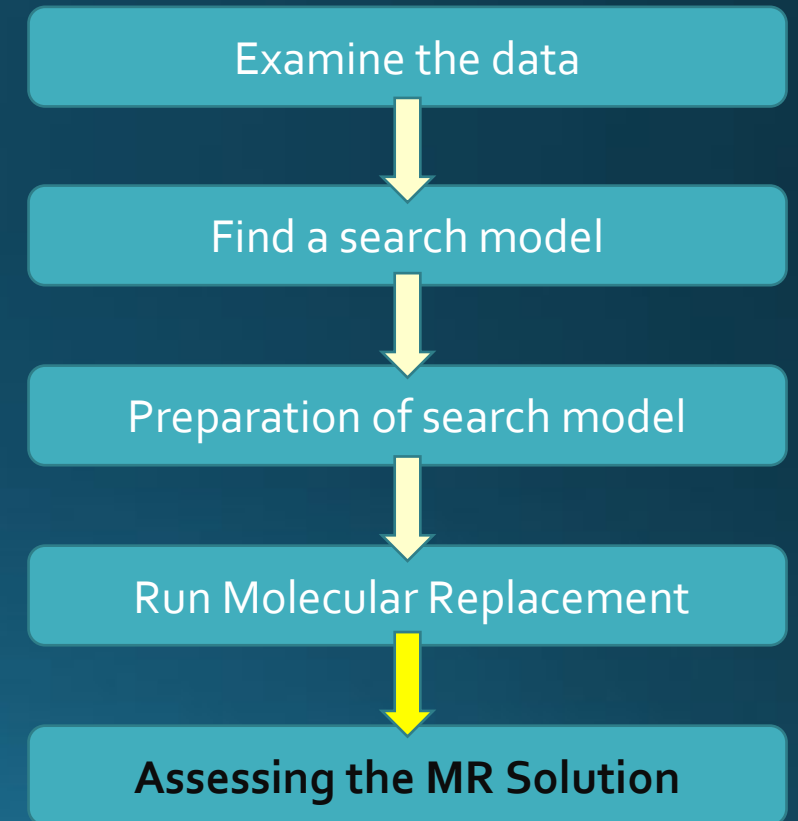
Step-by-Step MR: *Assessing the MR solution*

- How to know if the search model is correctly positioned in the target unit cell



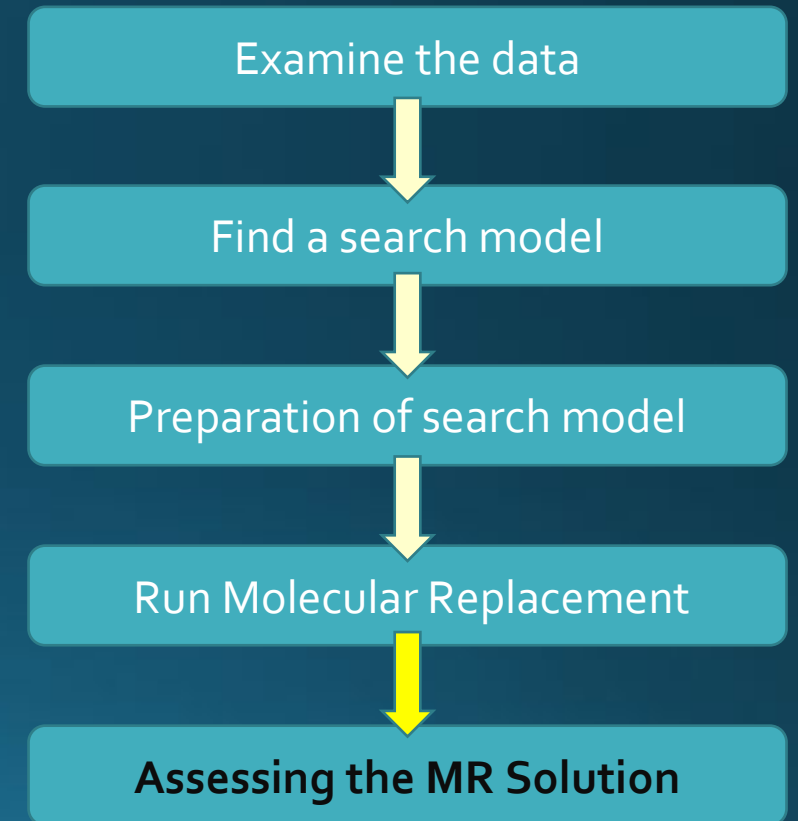
Step-by-Step MR: *Assessing the MR solution*

- How to know if the search model is correctly positioned in the target unit cell
- In difficult cases the position may be correct but getting from the MR solution to a complete model may not be straight forward



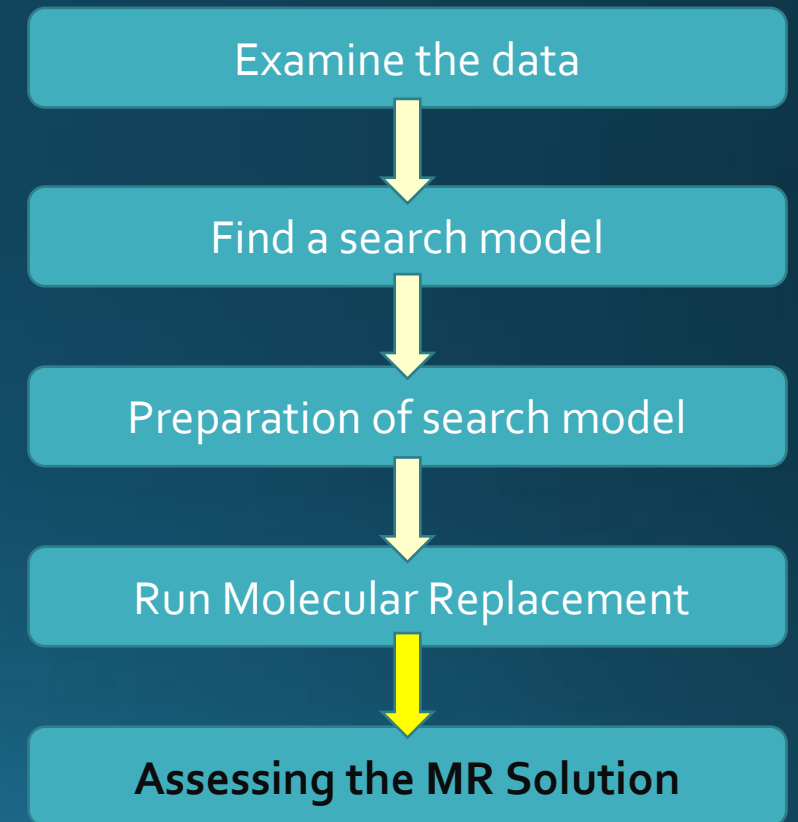
Step-by-Step MR: *Assessing the MR solution*

- Rough guide to MR program scoring



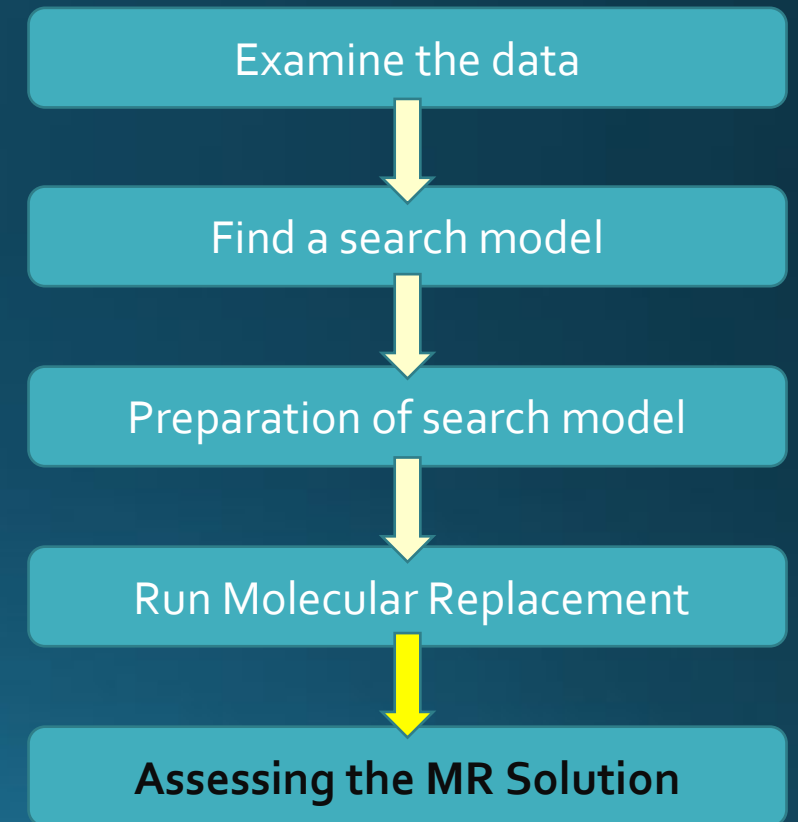
Step-by-Step MR: *Assessing the MR solution*

- Rough guide to MR program scoring
 - Phaser scores
 - LLG scores – has it increased by 60 or more after the placement of a new molecule?
(resolution and space group dependent)
 - TFZ – greater than 8?
 - Few or single solution almost always indicative of success



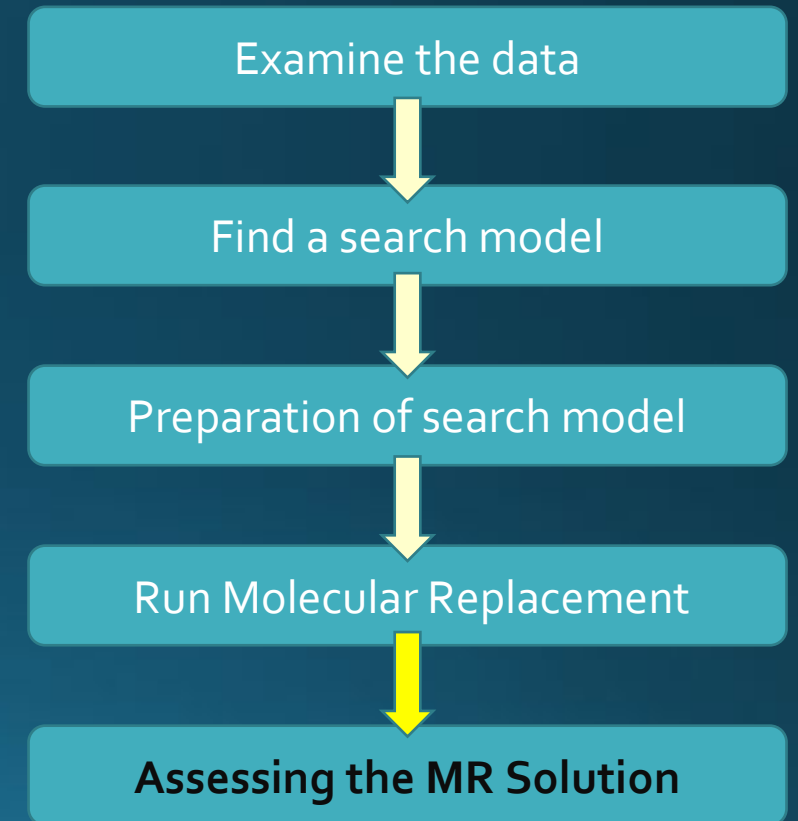
Step-by-Step MR: *Assessing the MR solution*

- Rough guide to MR program scoring
 - Phaser scores
 - LLG scores – has it increased by 60 or more after the placement of a new molecule?
(resolution and space group dependent)
 - TFZ – greater than 8?
 - Few or single solution almost always indicative of success
 - Molrep scores
 - RFZ – rotation search score greater than 5 – is there a clear peak?
 - TFZ – translation search score – is there a clear peak?



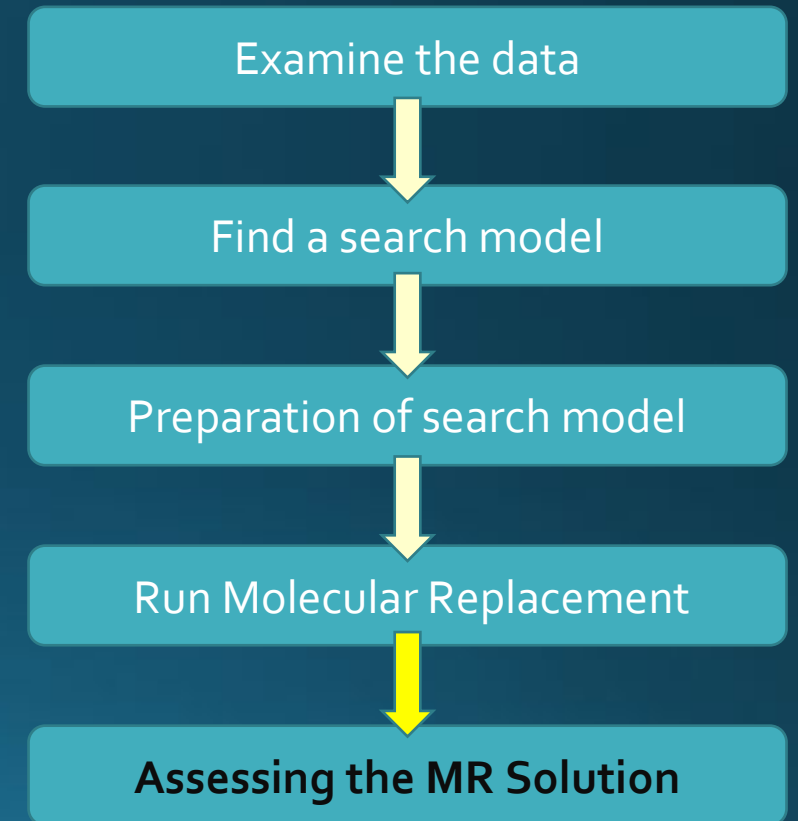
Step-by-Step MR: *Assessing the MR solution*

- Refinement



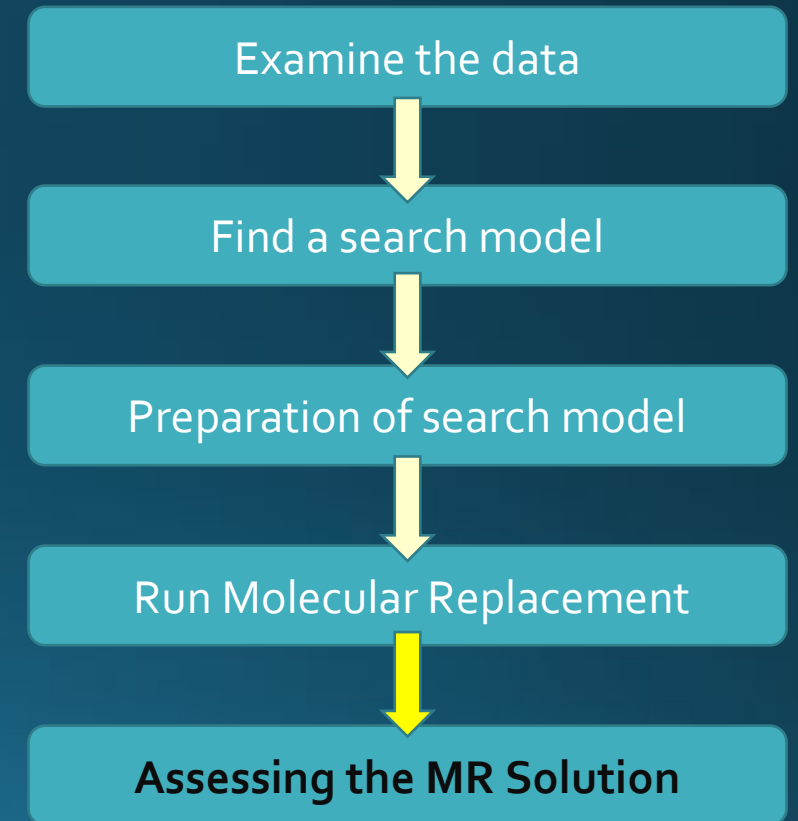
Step-by-Step MR: *Assessing the MR solution*

- Refinement
 - Look at Rfactor/Rfree
 - are they falling? Is Rfree below 0.5?



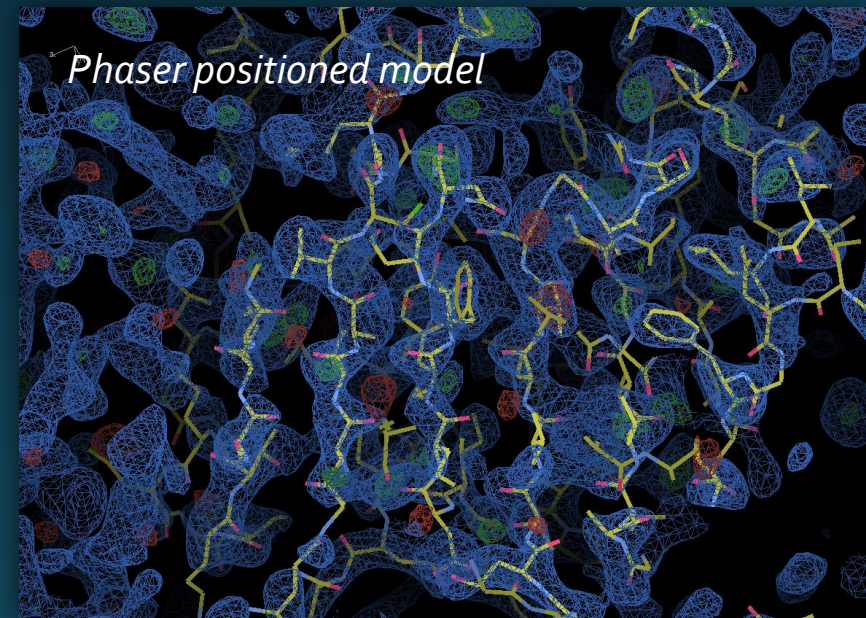
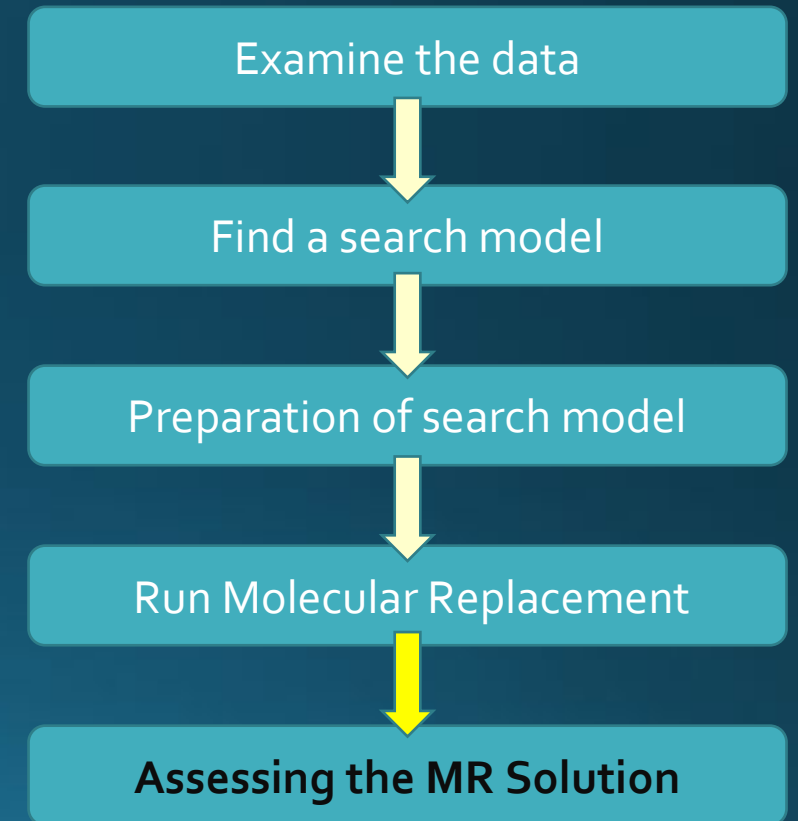
Step-by-Step MR: *Assessing the MR solution*

- Refinement
 - Look at Rfactor/Rfree
 - are they falling? Is Rfree below 0.5?
 - Use 200 cycles of jelly body refinement option in Refmac post MR



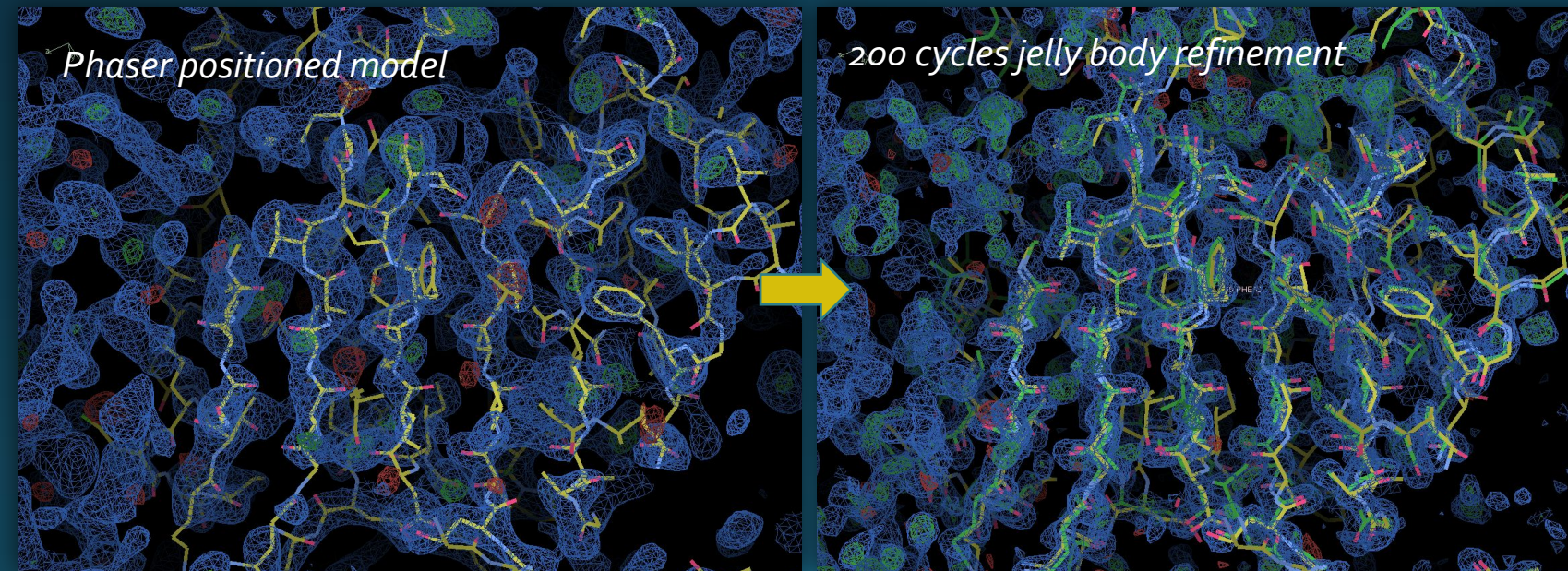
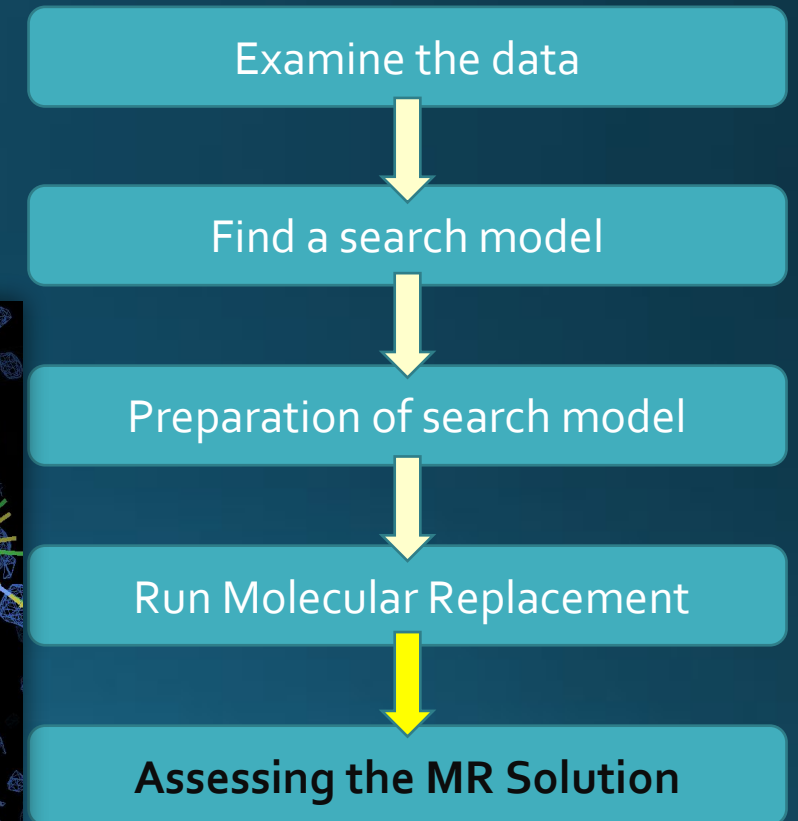
Step-by-Step MR: *Assessing the MR solution*

- Refinement
 - Look at Rfactor/Rfree
 - are they falling? Is Rfree below 0.5?
 - Use 200 cycles of jelly body refinement option in Refmac post MR



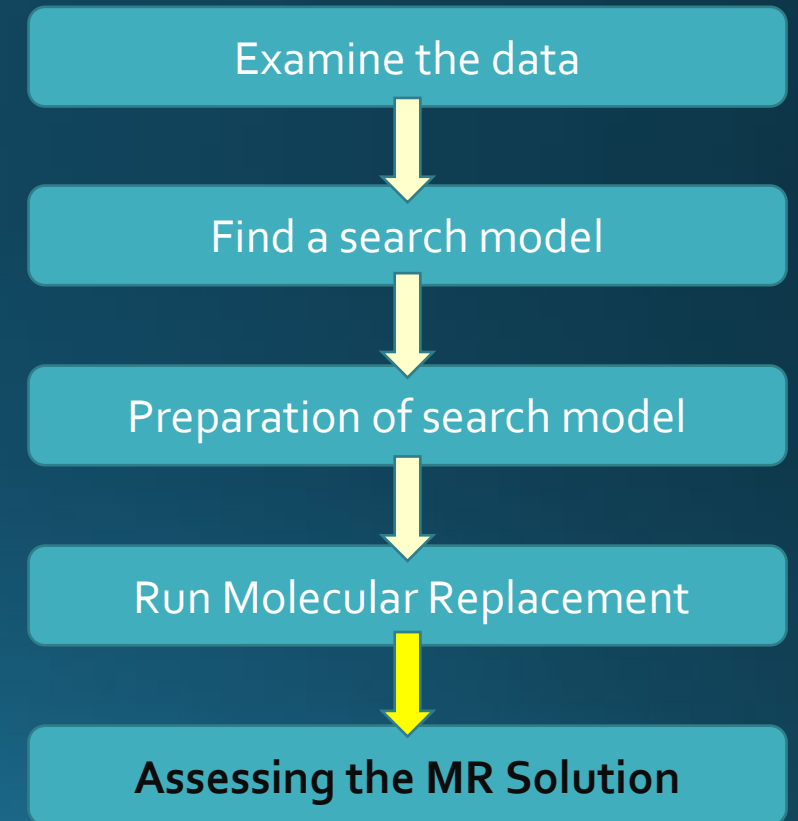
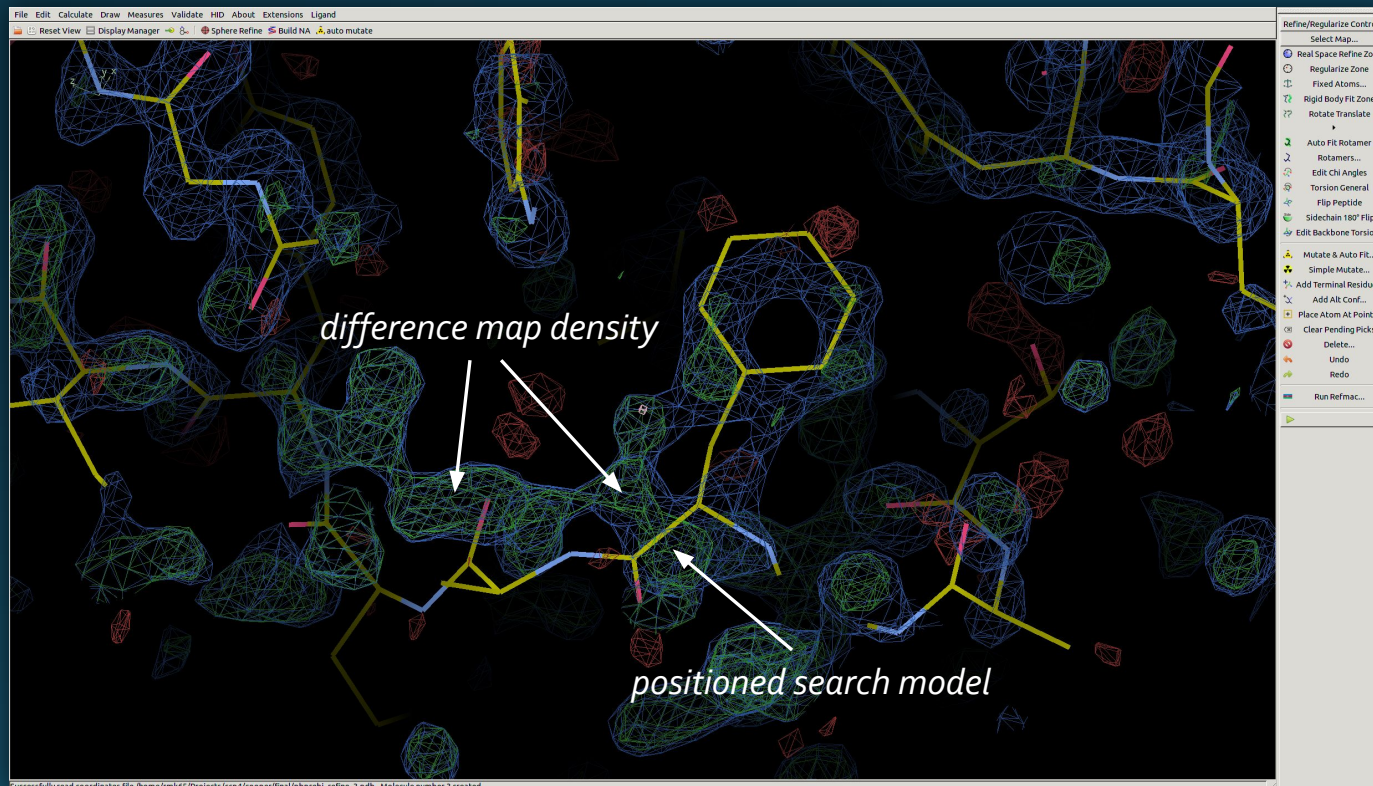
Step-by-Step MR: Assessing the MR solution

- Refinement
 - Look at Rfactor/Rfree
 - are they falling? Is Rfree below 0.5?
 - Use 200 cycles of jelly body refinement option in Refmac post MR



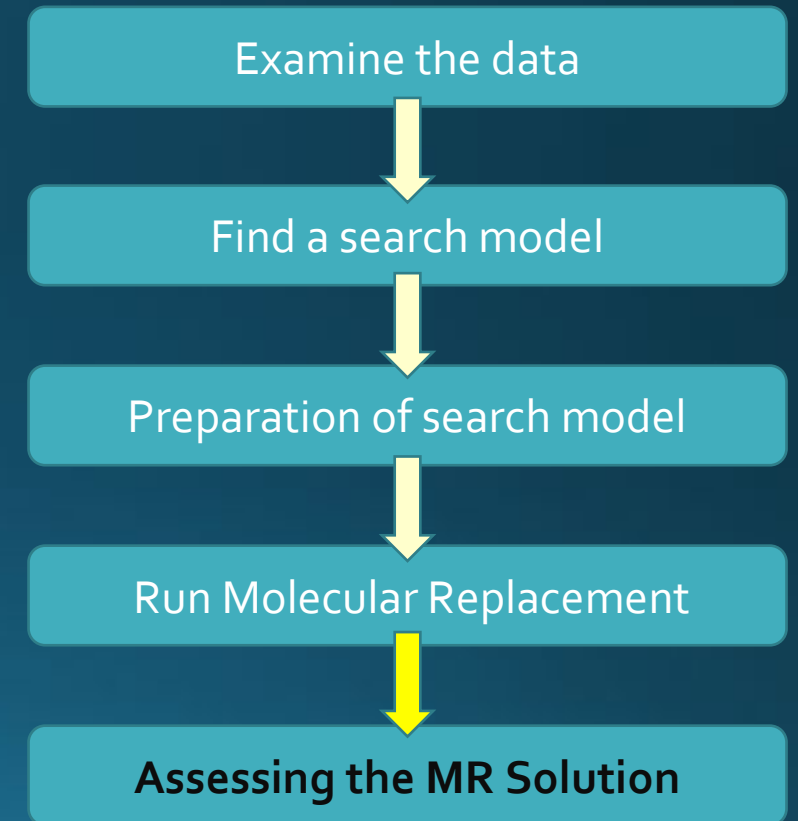
Step-by-Step MR: Assessing the MR solution

- Examine solution by eye
 - Use Coot to examine positioned models & maps



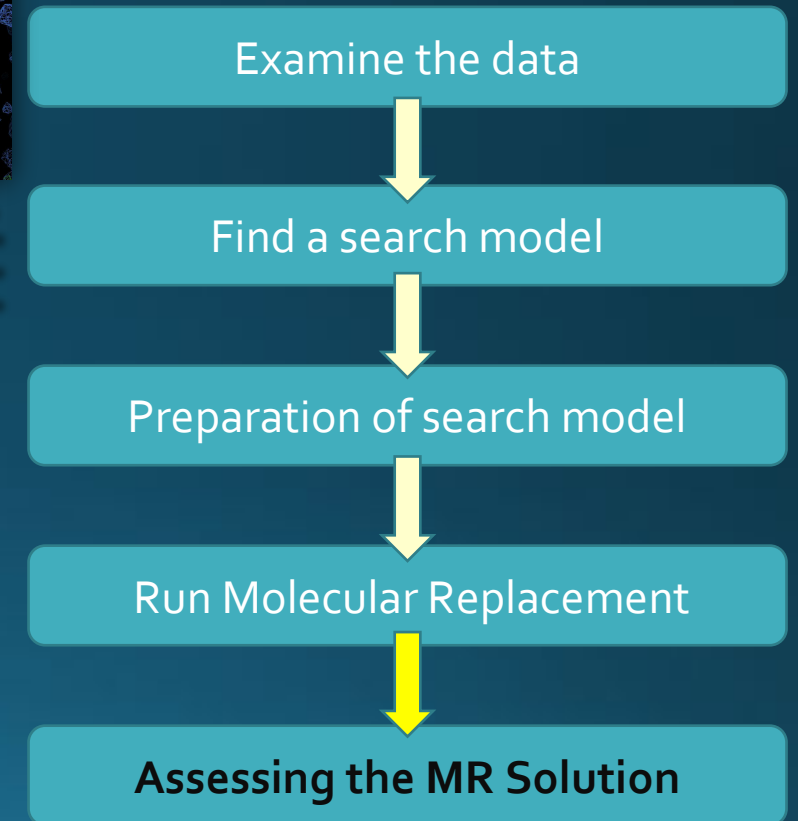
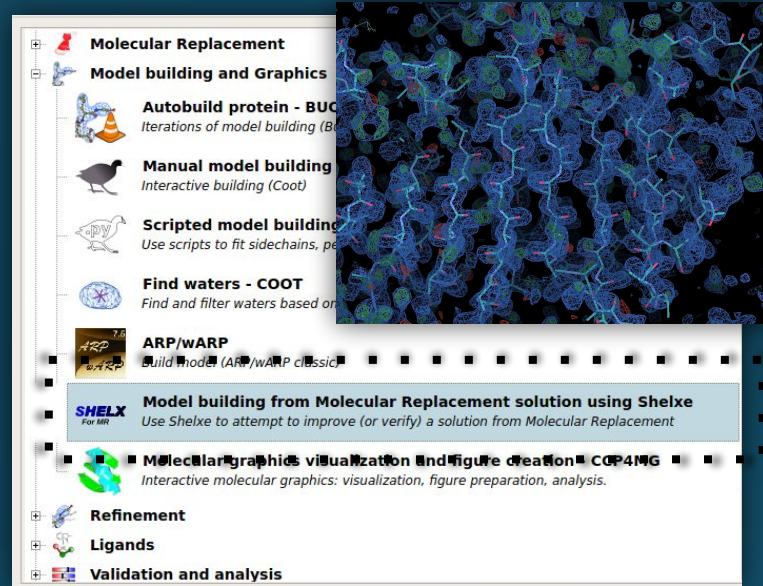
Step-by-Step MR: *Assessing the MR solution*

- Phase Improvement & C-alpha tracing



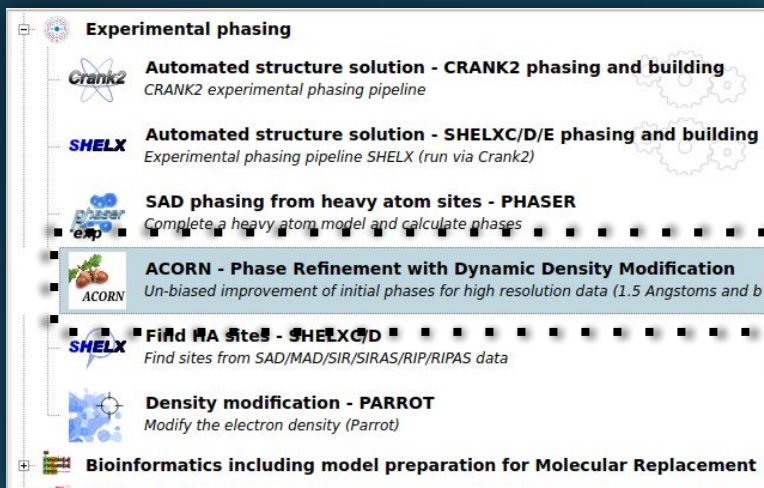
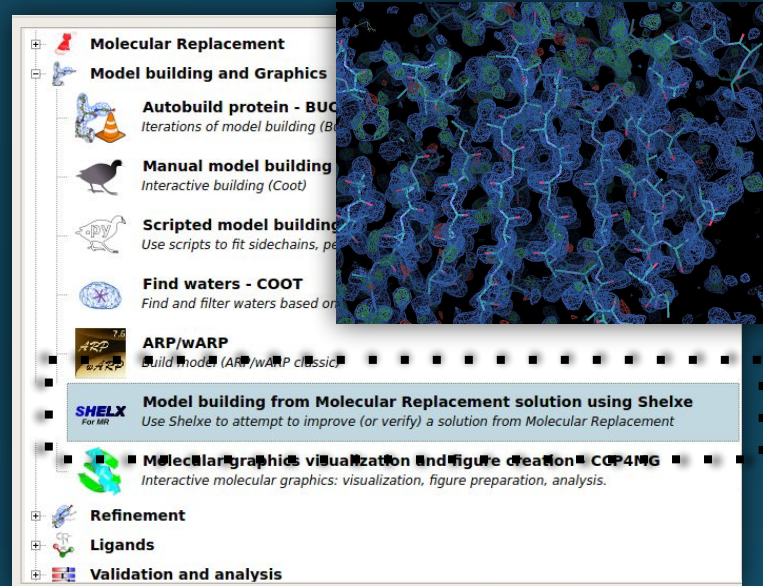
Step-by-Step MR: Assessing the MR solution

- Phase Improvement & C-alpha tracing
- SHELXE for MR
(resolution better than 2.5Å)

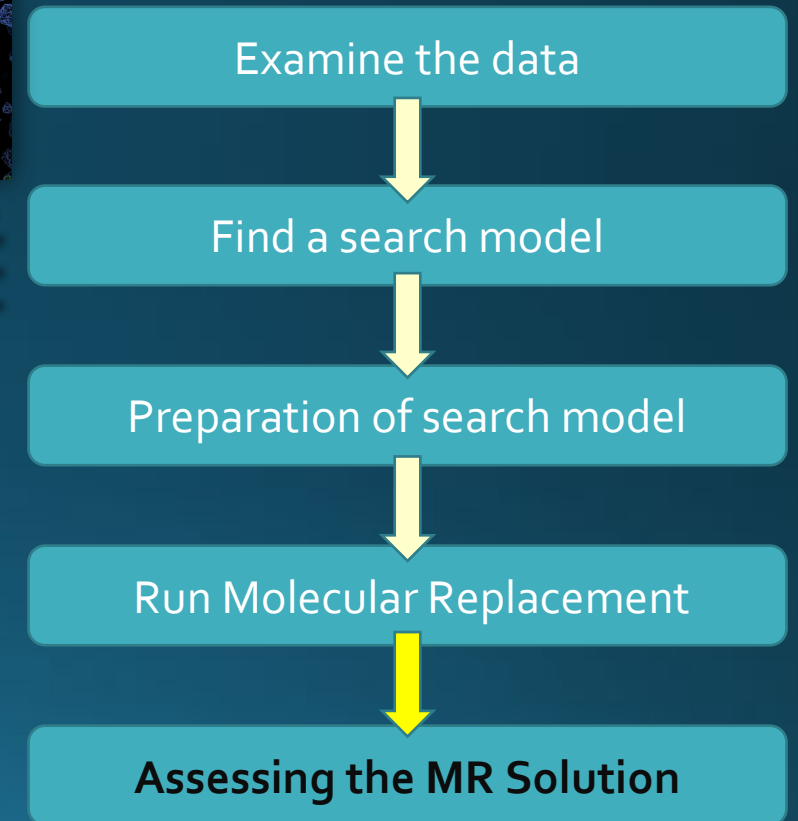


Step-by-Step MR: Assessing the MR solution

- Phase Improvement & C-alpha tracing
- SHELXE for MR (*resolution better than 2.5Å*)

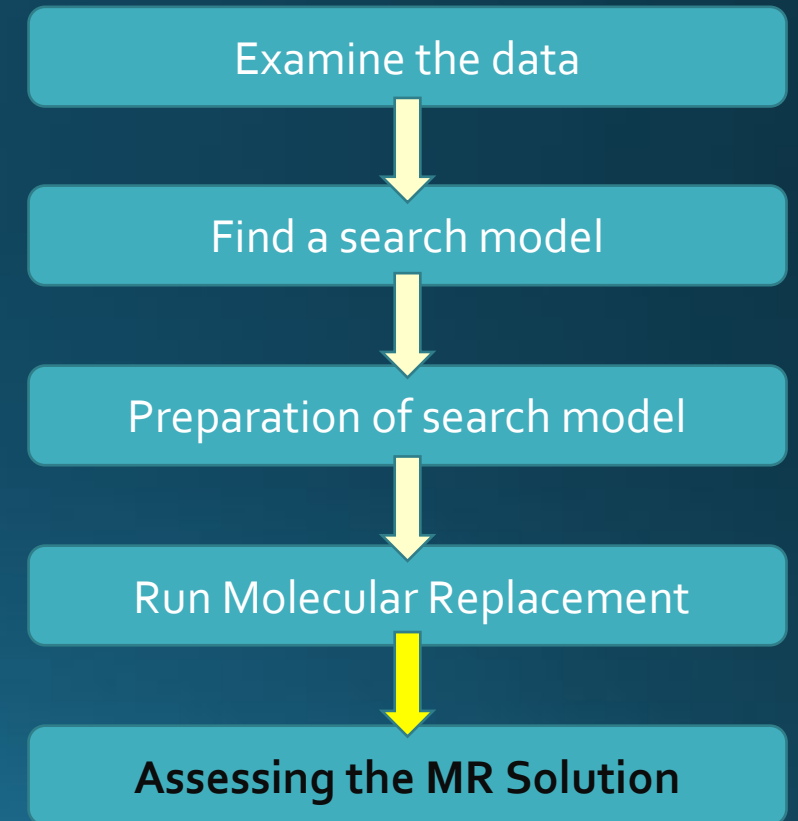
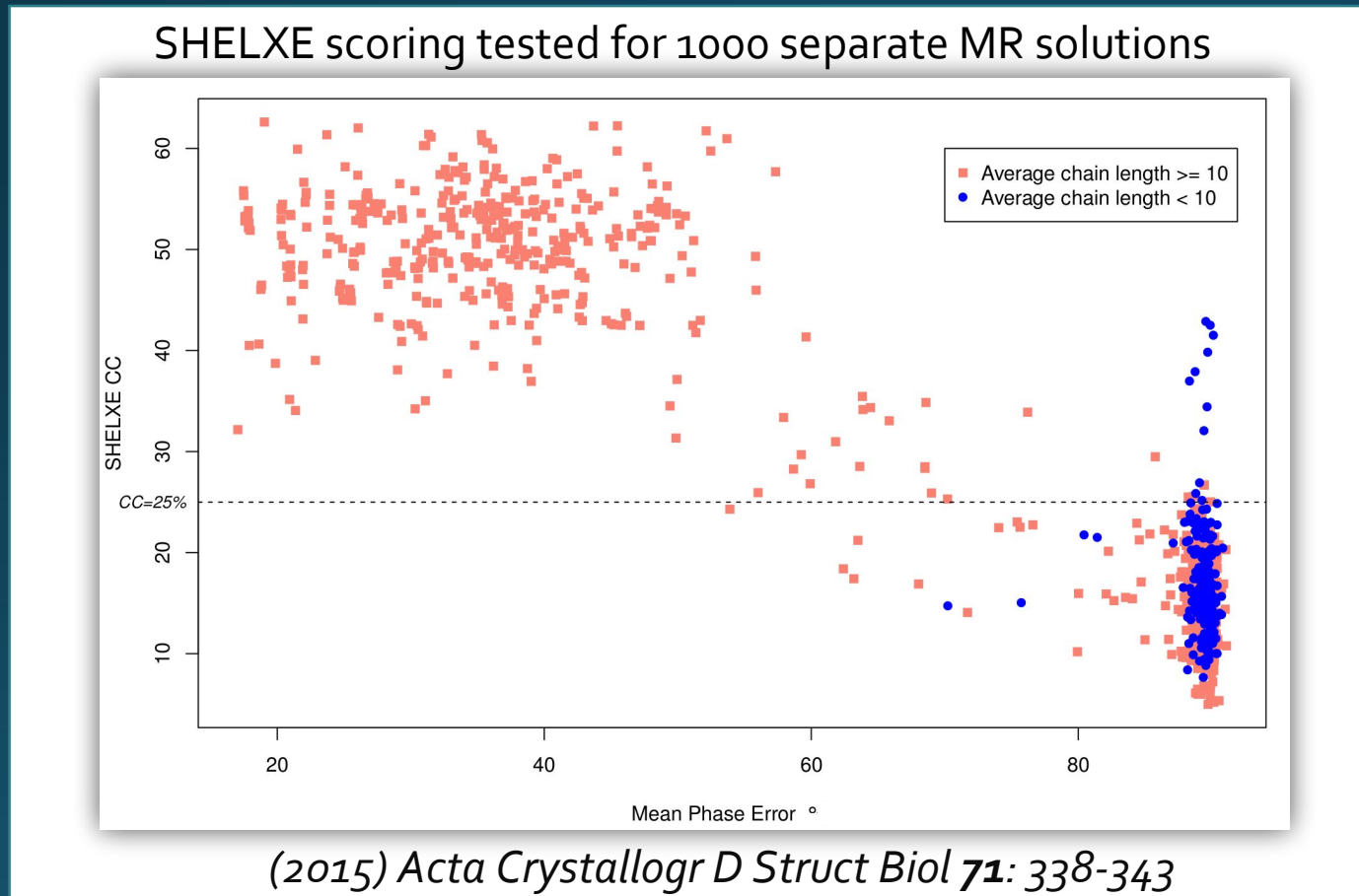


- ACORN (*resolution better than 1.7Å*)



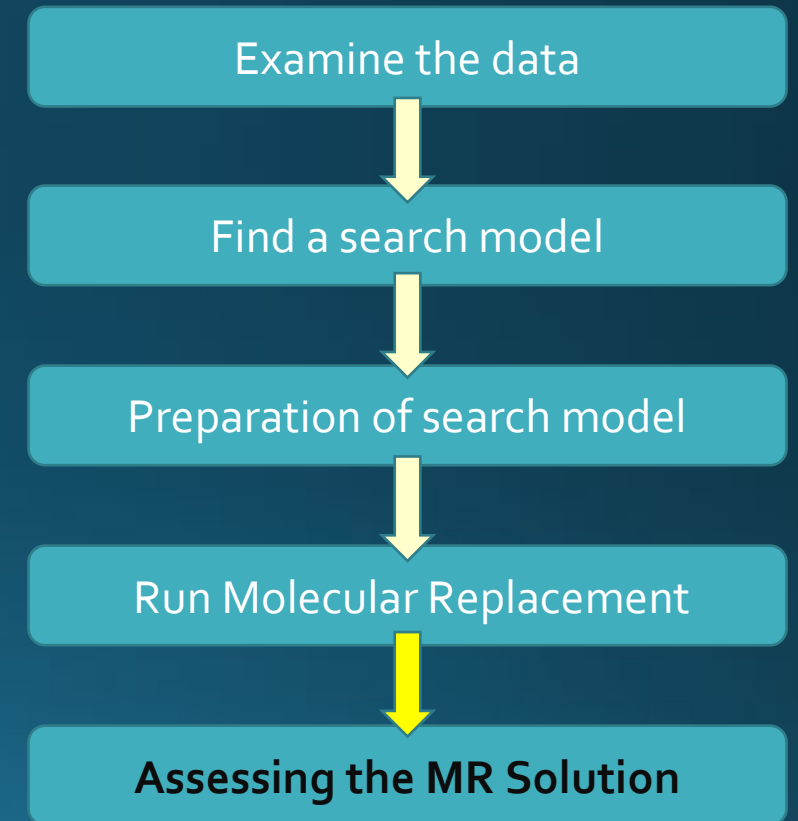
Step-by-Step MR: Assessing the MR solution

- SHELXE scoring: $CC \geq 25\%$
(average c-alpha fragment length ≥ 10)



Step-by-Step MR: *Assessing the MR solution*

- Automatic model building: Buccaneer & ARP/wARP
 - Can be used post-MR for generation of better model and phases for the target
 - Rebuilding parts that may not be present in search model
 - Useful for assessing whether or not your positioned MR model is true – eliminates bias

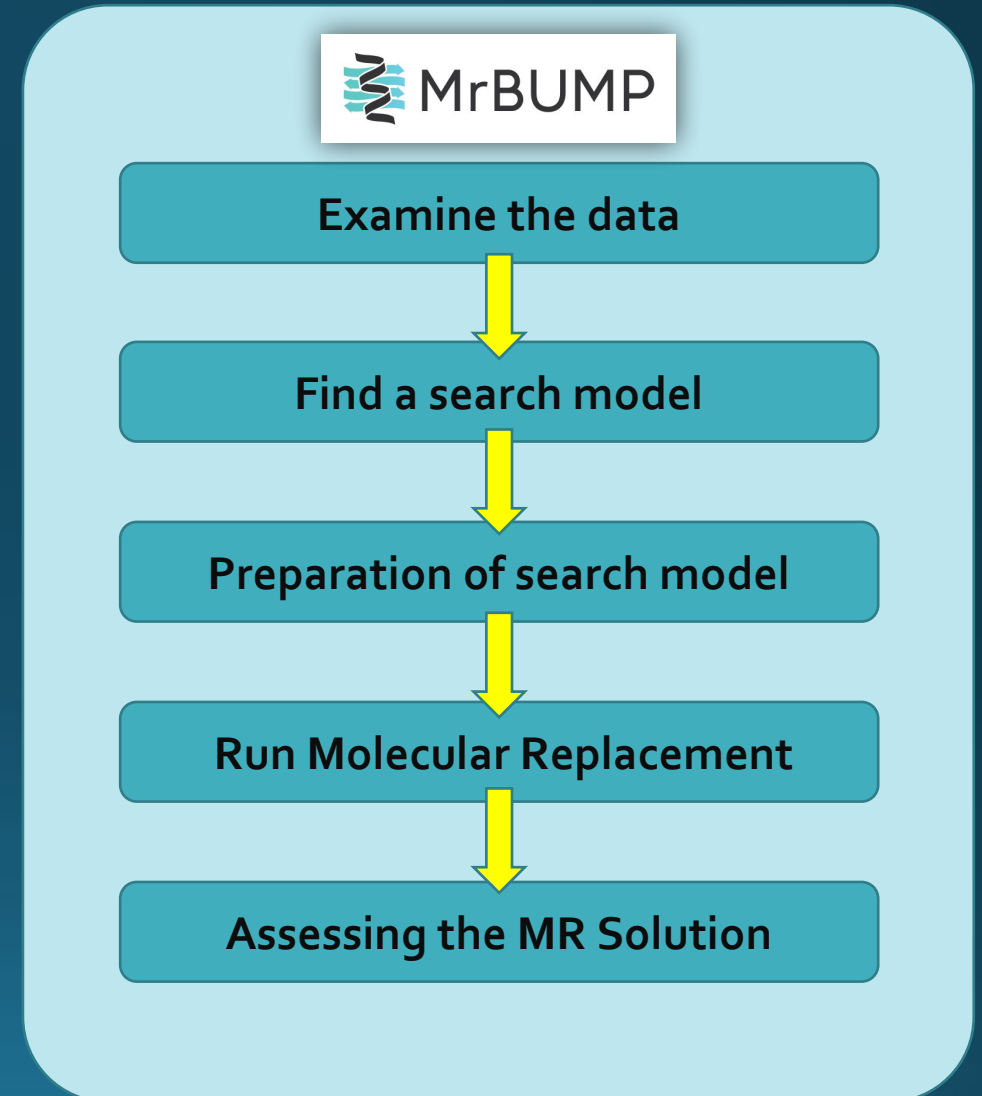


Automated Molecular Replacement in CCP4

- Several automation pipelines for MR in CCP4:
 - *MrBUMP* – model search, preparation, MR and refinement
 - *BALBES* – model search in custom version of PDB database
 - *MoRDa* – similar to BALBES

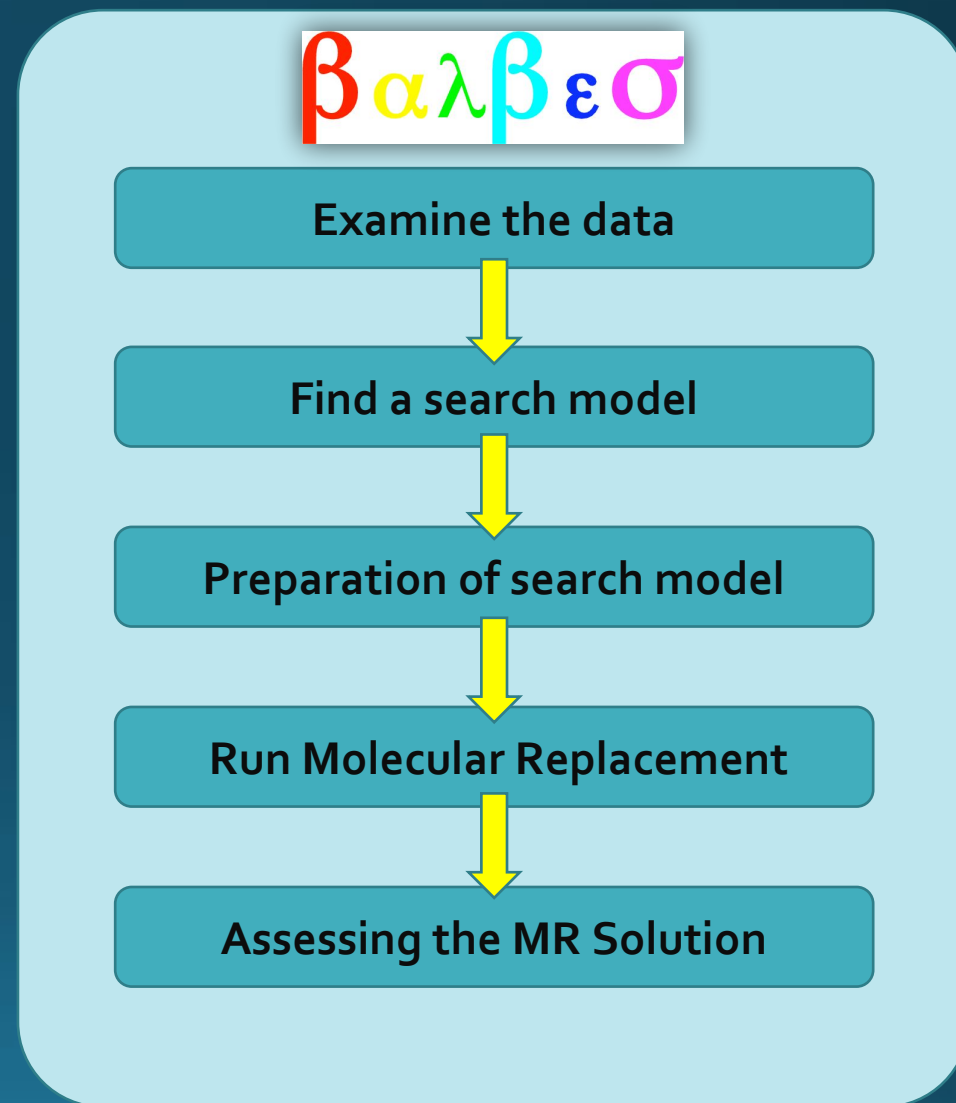
MrBUMP

- Covers all stages from data examination through to initial model building
- Two modes:
 1. Model generation only
 2. Model generation and Molecular Replacement through to model building
- Can be run from:
 - CCP4i2, CCP4i & CCP4 Online
 - CCP4mg – model generation step with graphical view



BALBES/MoRDa

- Covers all stages from data examination through to refinement
- Custom database:
 1. Non-redundant version of PDB
 2. Basic entry is a domain with definitions for constructing full chains, multimers and ensembles
- Other features:
 - Search all spacegroups
 - Uses Molrep for MR and Refmac for refinement
 - Available in CCP4i & on CCP4 Online



CCP4 Online: MrBUMP, BALBES, MoRDa & more

The screenshot displays the CCP4 Online web interface. The main page is titled "New MrBUMP Run" and includes a form for submitting a new run. The form has fields for "Structure Factors" and "Sequence Target", both with "Choose File" buttons and "No file chosen" status. Below these is a text area for pasting a FASTA sequence, with a note: "Instead of entering a Sequence Target file you can paste your FASTA sequence below: (Note that a comment line beginning with a '>' character must precede each sequence)". There is also a checkbox for "I have the SHELX licence (optional)".

To the right, a table shows interface parameters for two monomers (F and E) and their interaction. The table is as follows:

Monomer 1		Monomer 2	
F	E	F	E
Protein		Protein	
X, Y, Z		X, Y, Z	
1_555		1_555	
232	7.3%	221	6.9%
1925	60.5%	1921	60.4%
3184	100.0%	3183	100.0%
66	16.1%	65	15.9%
380	92.9%	385	94.4%
409	100.0%	408	100.0%
2146.6	9.6%	2345.5	10.6%
2253.4	100.0%	22072.6	100.0%
	-331.9		-334.3
	-7.5		-5.0

Below the table is an "Interaction radar" chart with axes labeled IA, DG, BE, PV, HB, SB, and DS. The chart shows the relative values of these parameters for the two monomers.

Below the chart is a list of "Interface parameters":

- IA : Interface area, Å² 2246
- DG : Solvation Energy, kcal/mol -12.46
- BE : Total Binding Energy, kcal/mol -26.65
- PV : Hydrophobic P-value 0.3542
- HB : Number of Hydrogen Bonds 11
- SB : Number of Salt Bridges 25
- DS : Number of Disulphide Bonds 0

At the bottom of the page, there are buttons for "Download PDB", "Download PDBx", and "View in JSMol".

<http://www.ccp4.ac.uk/ccp4online>

What to do if Molecular Replacement doesn't work

What to do if Molecular Replacement doesn't work

- Try different search models or different preparation methods for the homologues

What to do if Molecular Replacement doesn't work

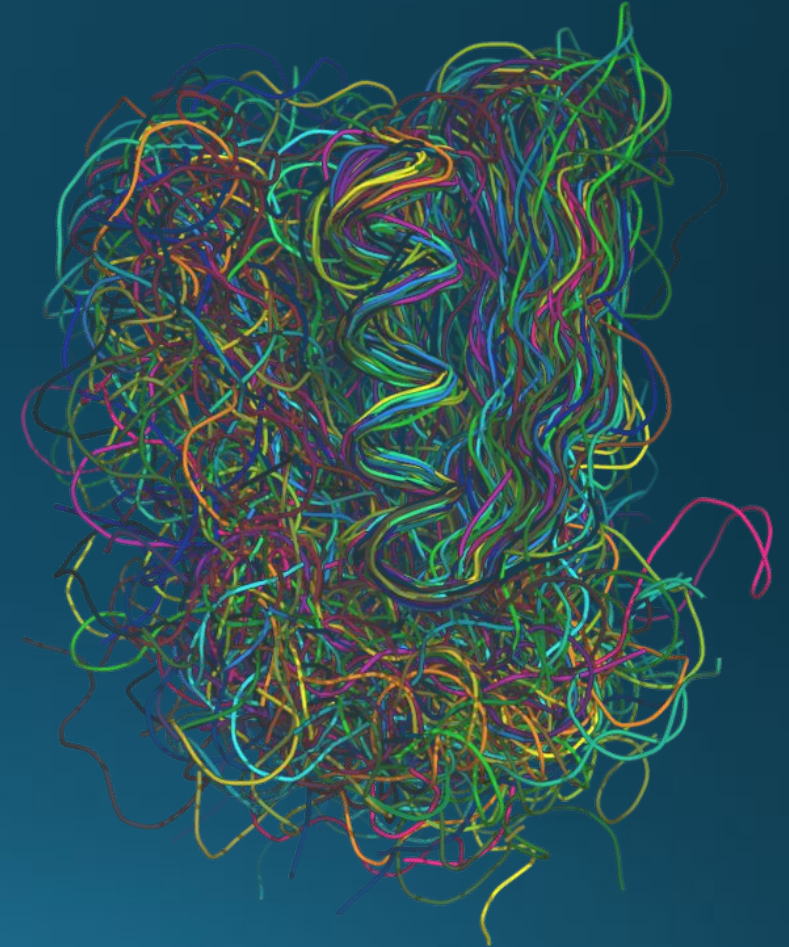
- Try different search models or different preparation methods for the homologues
- Try more distant homologues
 - Sequence similarity does not always imply structural similarity e.g. S100 family of proteins
 - Distant homologues can be structurally similar e.g. globular enzymes

What to do if Molecular Replacement doesn't work

- Try different search models or different preparation methods for the homologues
- Try more distant homologues
 - Sequence similarity does not always imply structural similarity e.g. S100 family of proteins
 - Distant homologues can be structurally similar e.g. globular enzymes
- Experimental Phasing
 - MR-SAD – Phaser, Crank2 or SHELX
 - Sulphur SAD

What to do if Molecular Replacement doesn't work

- **AMPLE**: *Ab initio* modelling to generate models for molecular replacement
- Uses programs like *Rosetta* and *Quark* to generate search models from the target sequence
- Can exploit sequence alignment derived residue contact predictions



What to do if Molecular Replacement doesn't work

• Crystal Contaminants

- Always check that you don't have a contaminant present
- Perform mass spectroscopy on your solution and crystals if possible
- Test your data immediately after data collection using SIMBAD or Contamminer



METHODS AND APPLICATIONS

Protein purification and crystallization artifacts: The tale usually not told

Ewa Niedzialkowska,^{1,2,3} Olga Gasiorowska,^{1,3} Katarzyna B. Handing,^{1,3}
Karolina A. Majorek,^{1,3,4} Przemyslaw J. Porebski,^{1,3} Ivan G. Shabalin,^{1,3,4,5}
Ewelina Zasadzinska,⁶ Marcin Cymborowski,^{1,3} and Wladek Minor^{1,3,4,5*}

¹Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, 1340 Jefferson Park Avenue, Jordan Hall, Room 4223, Charlottesville, Virginia, 22908

²Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Niezapominajek 8, Krakow, 30-239, Poland

³Midwest Center for Structural Genomics (MCSG), Argonne, Illinois, 60439

⁴Center for Structural Genomics of Infectious Diseases (CSGID), Chicago, Illinois, 60611

⁵New York Structural Genomics Research Consortium (NYSGRC), Bronx, New York, 10461

⁶Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, 1340 Jefferson Park Avenue, Jordan Hall, Room 6044, Charlottesville, Virginia, 22908

Received 1 October 2015; Accepted 2 December 2015

DOI: 10.1002/pro.2861

Published online 13 December 2015 proteinscience.org

Abstract: The misidentification of a protein sample, or contamination of a sample with the wrong protein, may be a potential reason for the non-reproducibility of experiments. This problem may occur in the process of heterologous overexpression and purification of recombinant proteins, as well as purification of proteins from natural sources. If the contaminated or misidentified sample is used for crystallization, in many cases the problem may not be detected until structures are deter-

Abbreviations: CSGID, Center for Structural Genomics of Infectious Diseases; GNAT, Gcn5-related N-acetyltransferase; IMAC, immobilized metal affinity chromatography; MAD, multi-wavelength anomalous diffraction; MR, molecular replacement; MCSG, Midwest Center for Structural Genomics; NYSGRC, New York Structural Genomics Research Consortium; PDB, Protein Data Bank; RMSD, root mean square deviation; SAD, single-wavelength anomalous dispersion; SEC, size exclusion chromatography; TEV, tobacco etch virus.

Additional Supporting Information may be found in the online version of this article.

Ewa Niedzialkowska and Olga Gasiorowska have contributed equally to this work.

The authors declare that there is no conflict of interest.

Description of Supporting Information material: Summary of data collection and refinement statistics for the deposited structures and the list of deposits used to identify crystallization artifacts by MR. Filename: Supplementary Materials.

Structural data are available in PDB database under accession numbers 4TNN, 4YYC, and 4ZNZ.

This work focuses on a particular difficulty that may occur as a result of accidental purification or contamination of the sample with a protein different than the protein of interest. Examples where the incorrect protein species was purified and/or crystallized are given.

Acknowledgements

- Daniel Rigden, Jens Thomas, Felix Simkovic, Adam Simpkin, University of Liverpool
- Stuart McNicholas, Keith Wilson, Christian Roth, University of York
- Martyn Winn, Eugene Krissinel, Charles Ballard, Ville Uski, Kyle Stevenson, Marcin Wojdyr, Andrey Lebedev ,CCP4
- Andrea Thorn, Isabel Uson, Tim Gruene & George Sheldrick (SHELX)
- Developers of Rosetta and Quark
- Refmac: Garib Murshudov, LMB-MRC Cambridge
- Phaser: Randy Read, Airlie McCoy & Gabor Bunkozci
- Molrep: Alexei Vagin & Andrey Lebedev
- Thanks to authors of all underlying programs
- Funding:
 - BBSRC, CCP4, University of Liverpool
- Support from the Research Complex at Harwell