Ronan Keegan CCP4 Group

MR model selection, preparation and assessing the solution

CCP4/DLS Workshop, Diamond, 2019

Overview

- Introduction to Molecular Replacement (MR)
- Practical guide to performing MR
- Automatic Molecular Replacement in CCP4
- What to do if Molecular Replacement doesn't work
- Acknowledgements

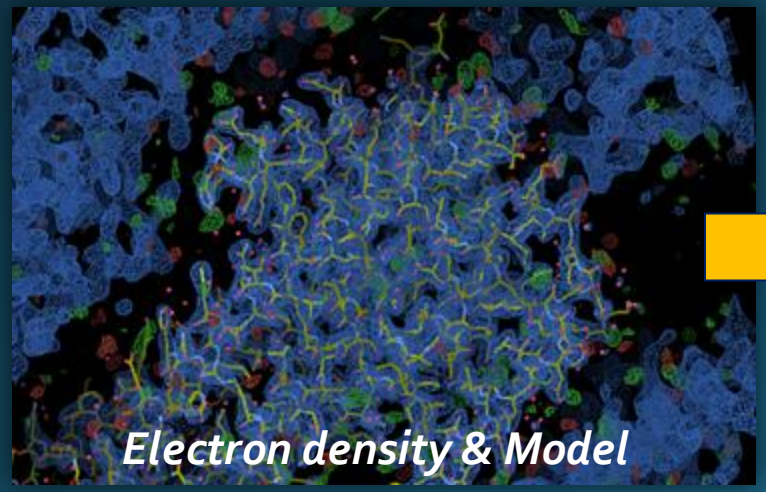
Introduction: *Basics of Molecular Replacement*

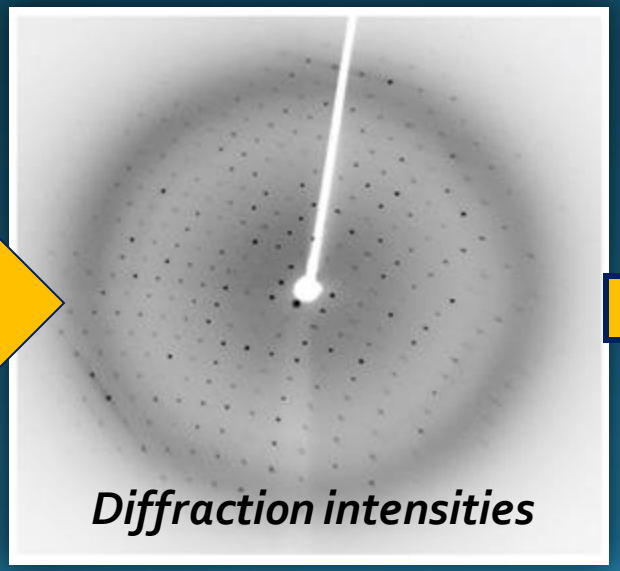
$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |F_o(hkl)| \cos[-2\pi(hx + ky + lz) + \varphi_{hkl}]$$

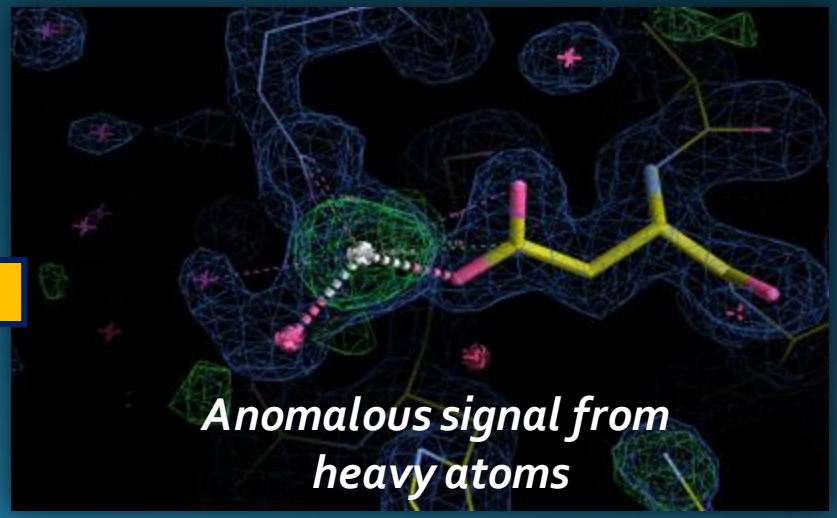
Electron density at (x,y,z)

Amplitudes

Phases







CCP4

MrBUMP

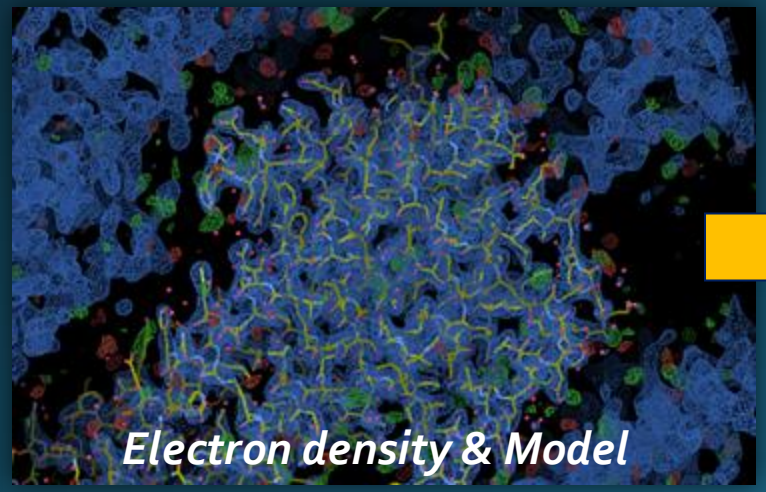
Introduction: *Basics of Molecular Replacement*

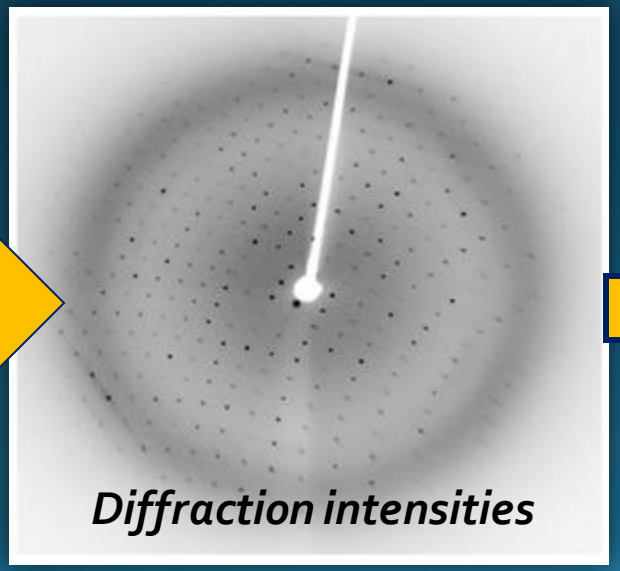
$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |F_o(hkl)| \cos[-2\pi(hx + ky + lz) + \varphi_{hkl}]$$

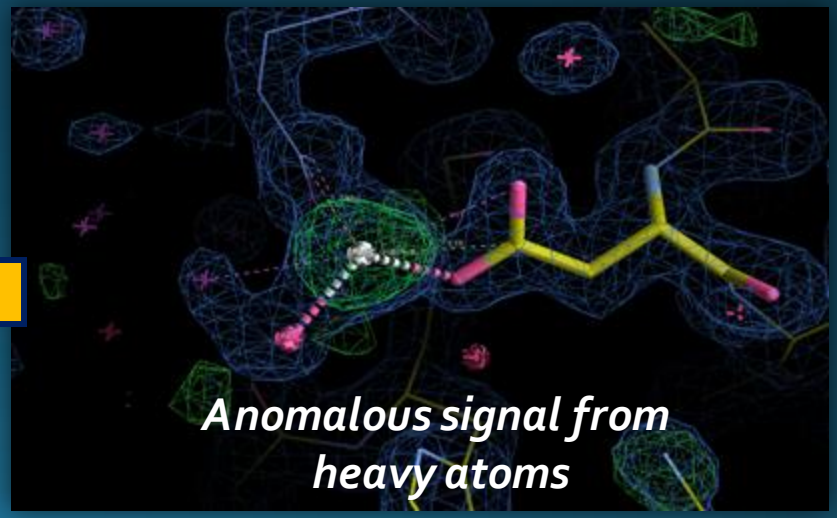
Electron density at (x,y,z)

Amplitudes

Phases







CCP4

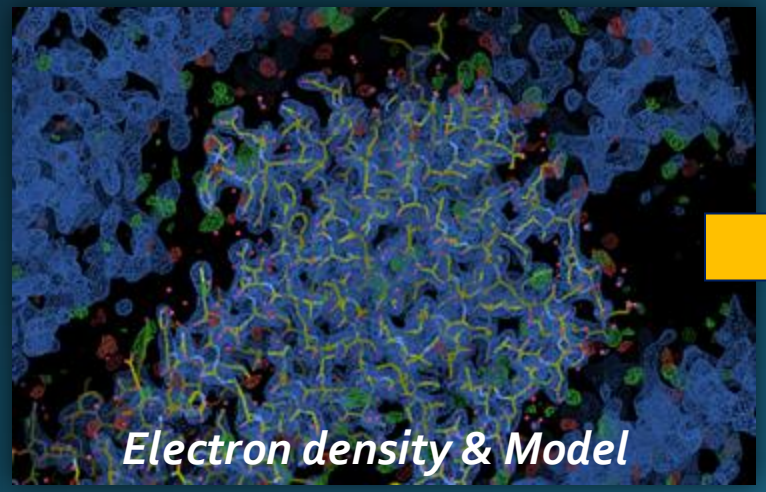
MrBUMP

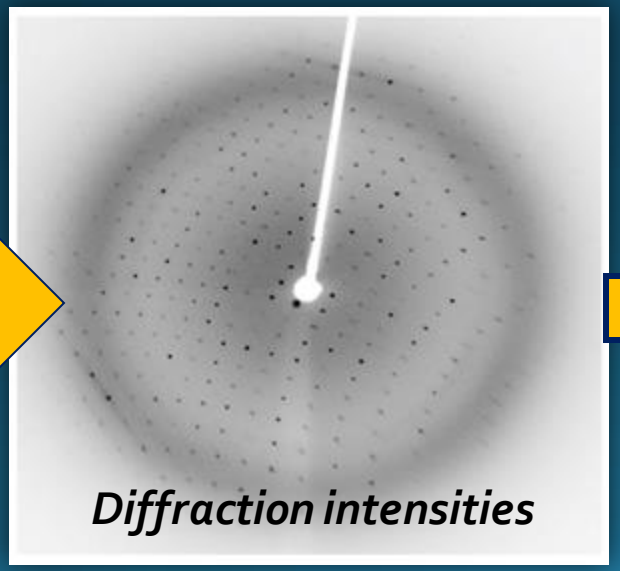
Introduction: *Basics of Molecular Replacement*

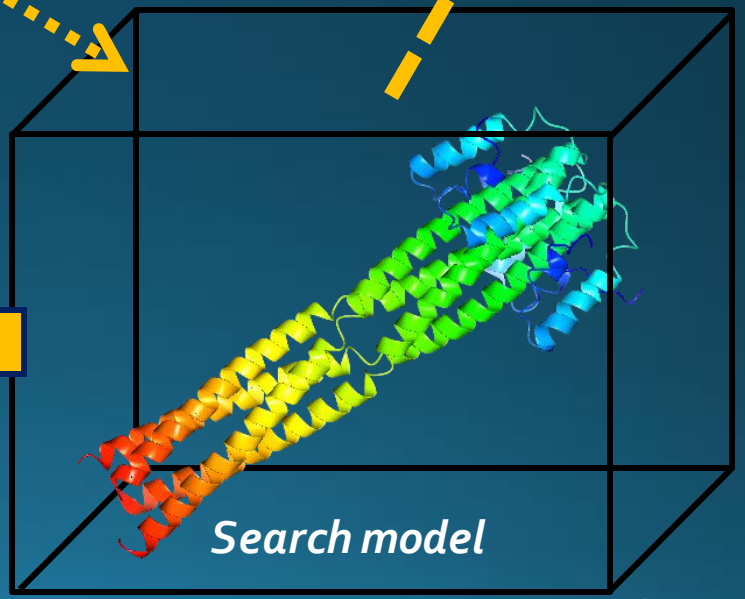
$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |F_o(hkl)| \cos[-2\pi(hx + ky + lz) + \varphi_{hkl}]$$

Electron density at (x,y,z)

Amplitudes



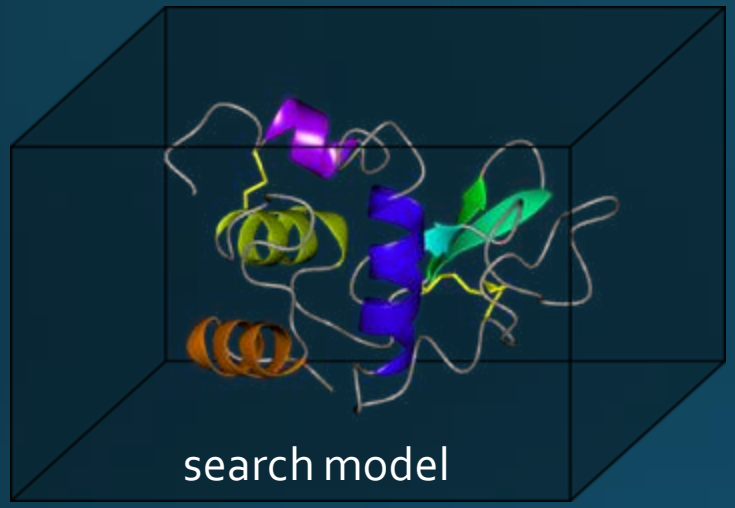




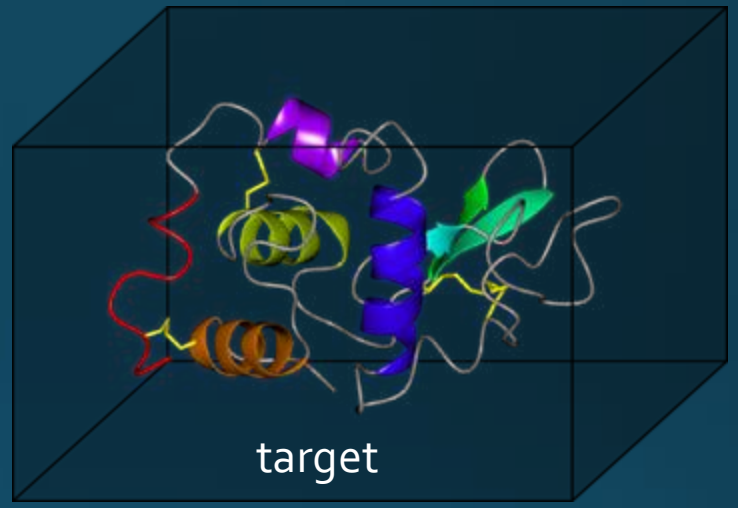
CCP4

MrBUMP

Introduction: *Basics of Molecular Replacement*



search model



target

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

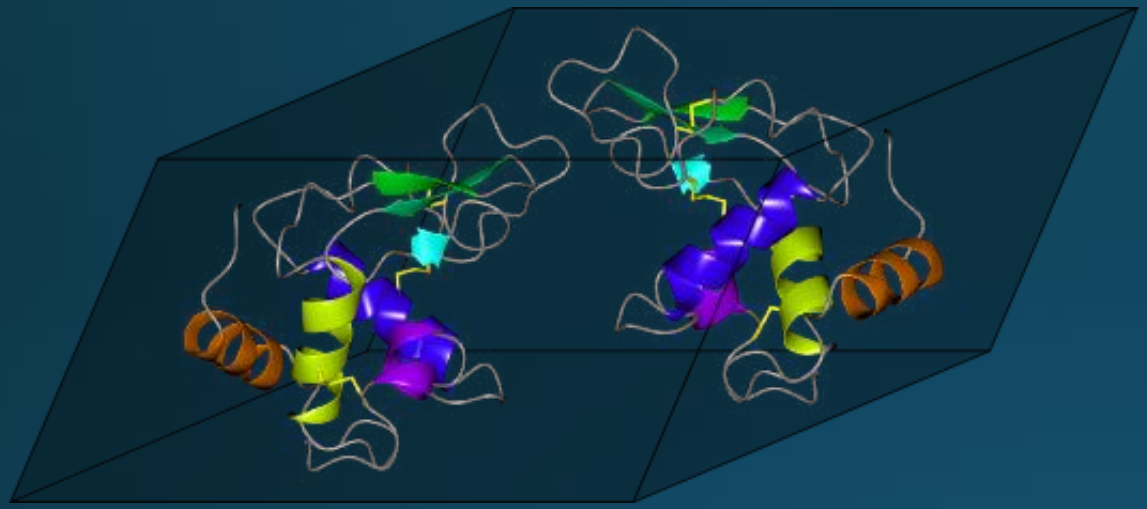
h	k	l	F	ϕ
0	0	1	10.4	113
0	0	2	3.5	18
0	0	3	57.2	265

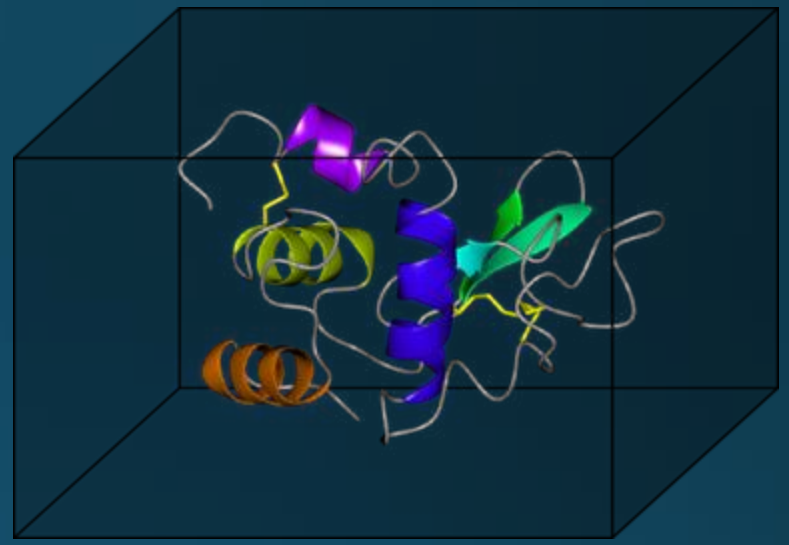
(slide from Gabor Bunkoczi)

CCP4

MrBUMP

Introduction: *Basics of Molecular Replacement*



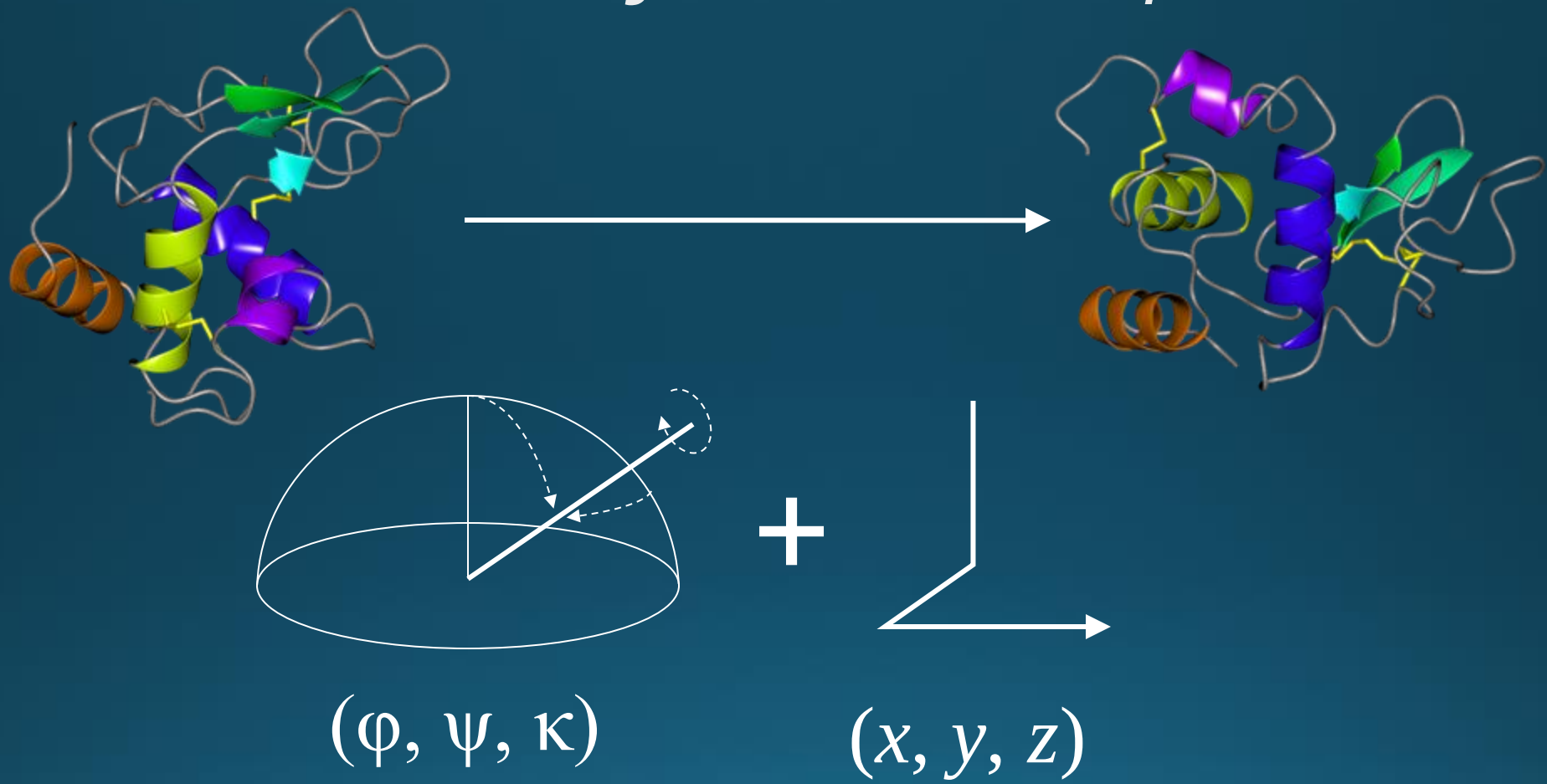


Search model crystal form → Target crystal form

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

(slide from Gabor Bunkoczi)

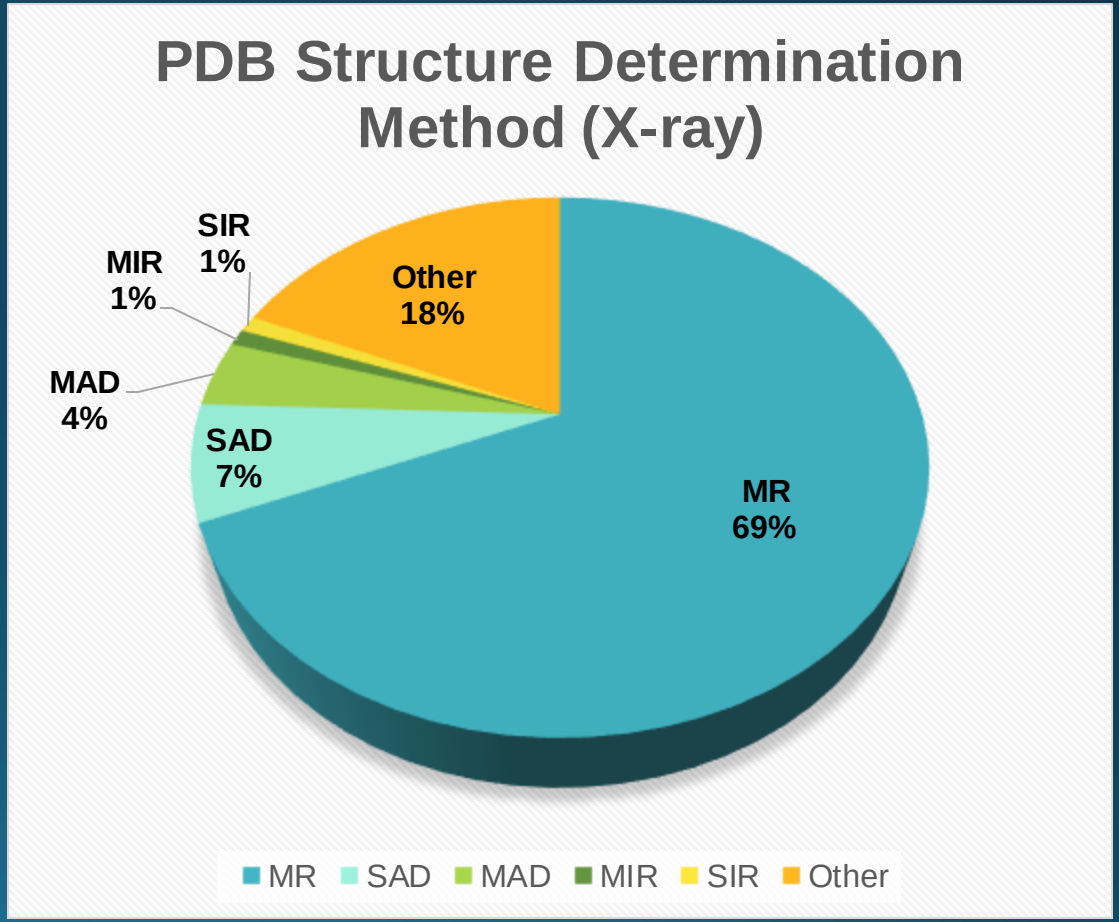
Introduction: *Basics of Molecular Replacement*



(slide from Gabor Bunkoczi)

Introduction: *Molecular Replacement and the PDB*

- Today, the vast majority of X-ray 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

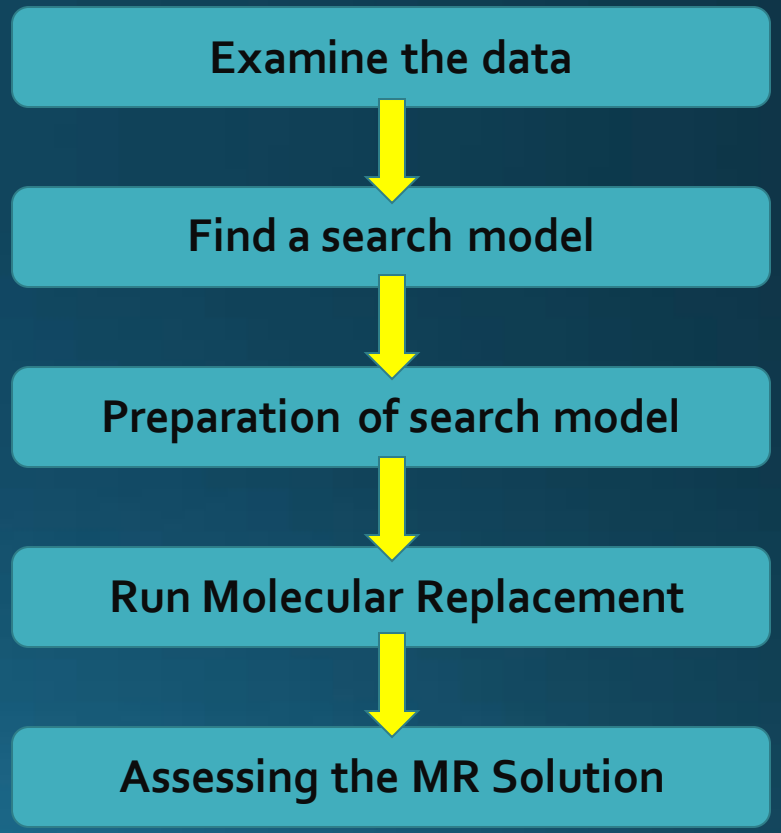


CCP4

MrBUMP

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?
    - Estimate through Matthews Coefficient
  - What's the resolution of the data?
  - Self-rotation function – signs of NCS?

Examine the data

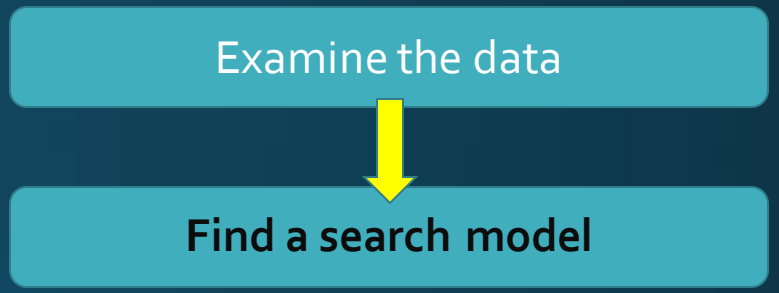
Step-by-Step MR: *Examining the data*

- Data issues can have an impact on how well MR will work
- Things to think about:
  - How many copies in the asymmetric unit?
    - Estimate through Matthews Coefficient
  - 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

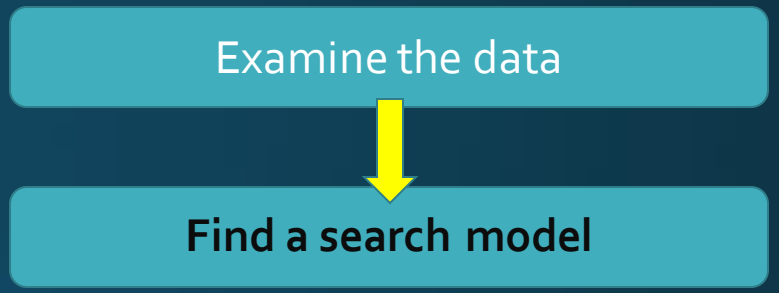
Step-by-Step MR: *Finding a search model*

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



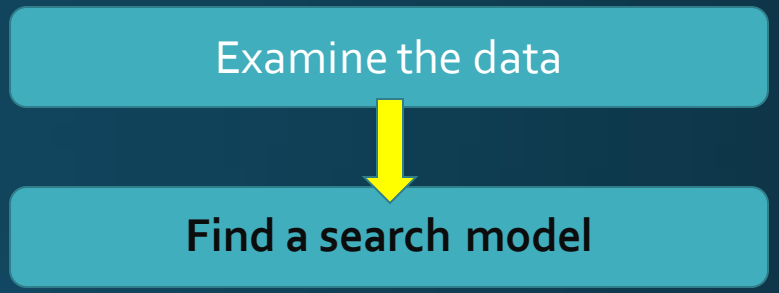
Step-by-Step MR: *Finding a search model*

- 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



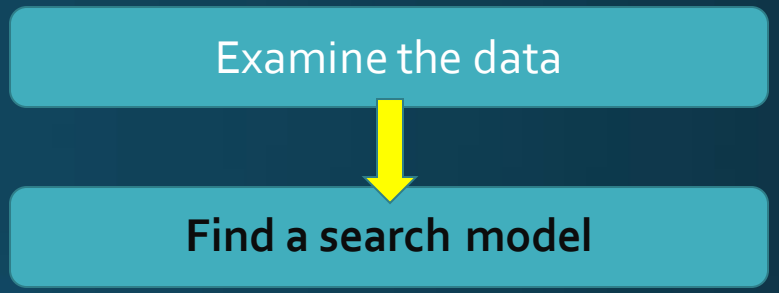
Step-by-Step MR: *Finding a search model*

- 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



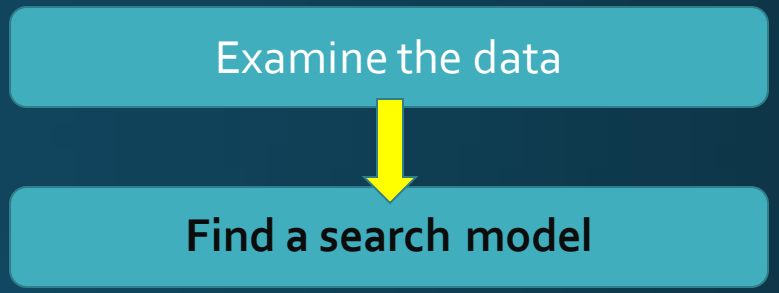
Step-by-Step MR: *Finding a search model*

- Considerations
  - Target Size:
    - The bigger the structure the more likely there will be sizeable conformational changes
  - Homologue Sequence identity < 30%
    - Suitable search models may exist – sequence identity calculation is difficult
    - eLLG from Phaser is a good measure of whether or not a search model has a chance to be placed correctly
  - Data Resolution



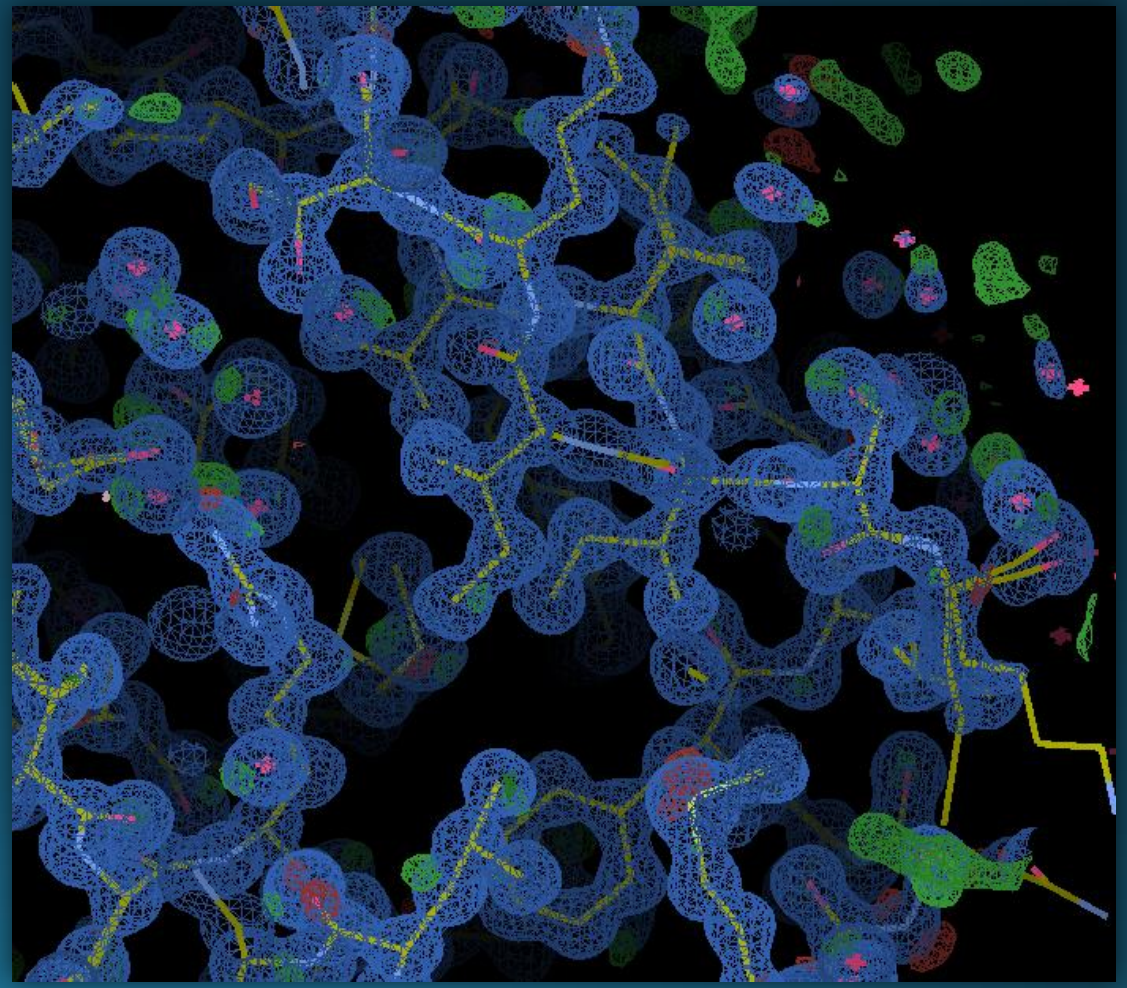
Step-by-Step MR: *Resolution and MR*

- How does data resolution impact on molecular replacement?



Step-by-Step MR: *Resolution and MR*

- Better than 1.0 Å – search model can be a small fragment or even a single atom in *Phaser*
  (*McCoy et al. 2017, PNAS*)

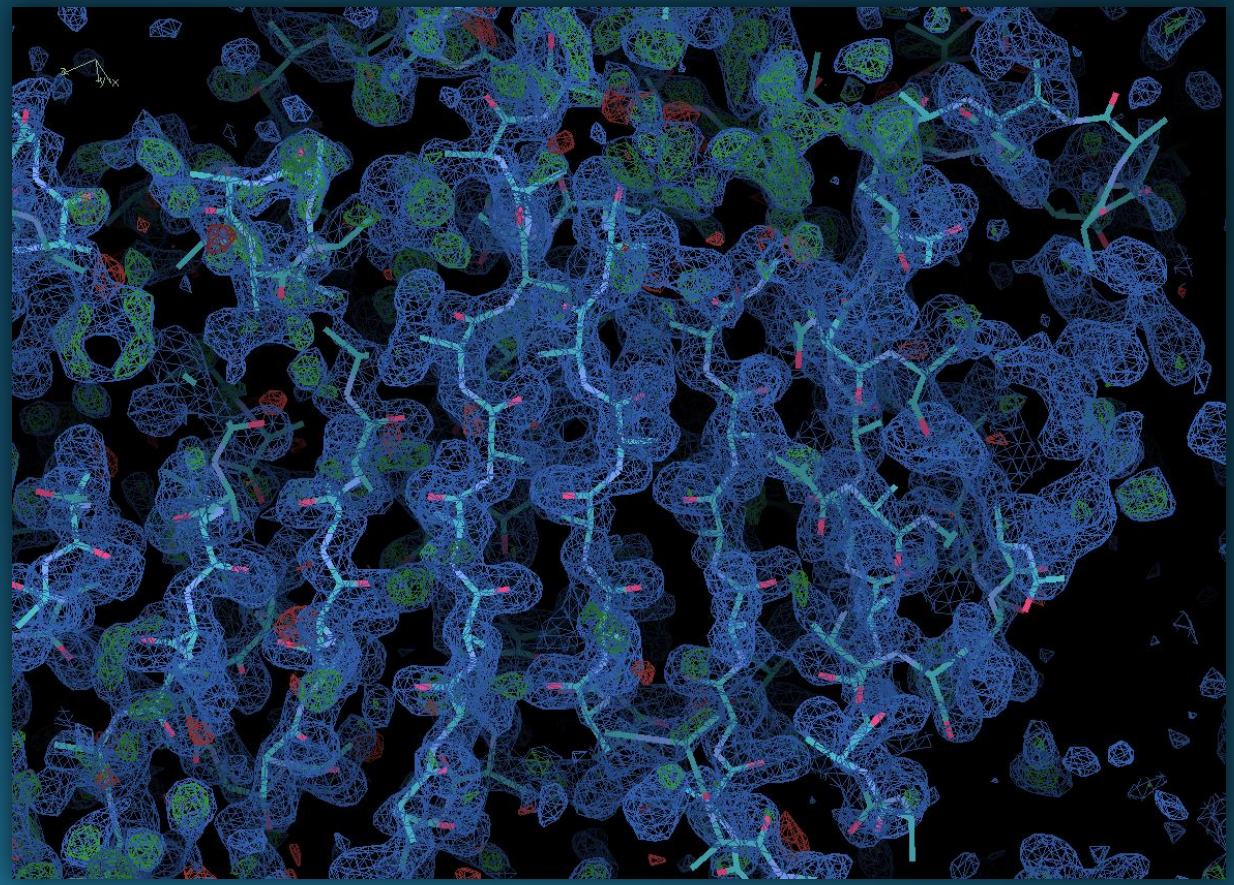


CCP4

MrBUMP

Step-by-Step MR: *Resolution and MR*

- At resolutions above ~ 2.5 Å
  - *Phaser* can place small fragments of total scattering
  - Applications like *SHELXE* and *Acorn* can improve through density modification (DM) and model building to a correct solution

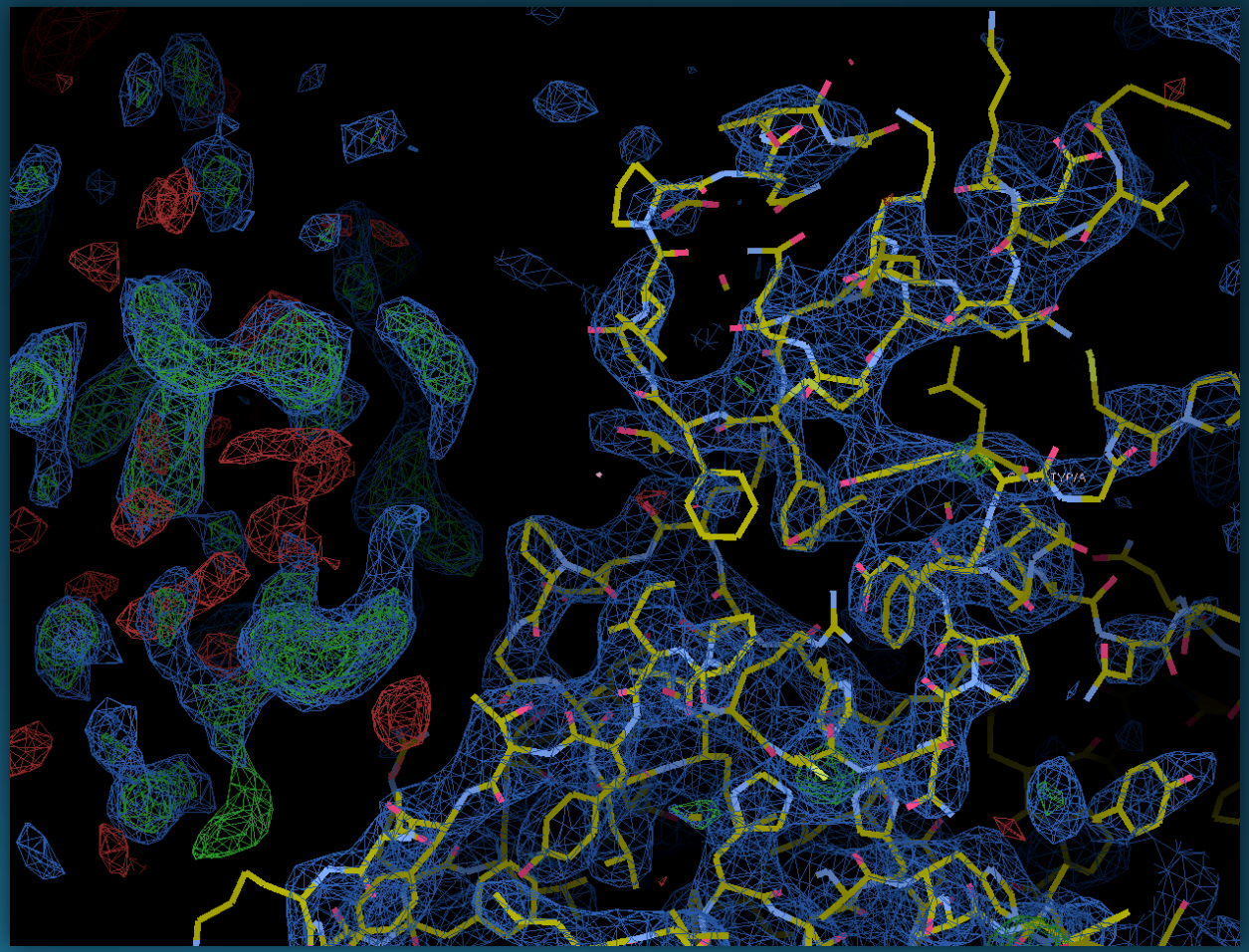


CCP4

MrBUMP

Step-by-Step MR: *Resolution and MR*

- 2.5Å or worse
  - DM techniques and fragment placing is less effective
  - Larger search models are required

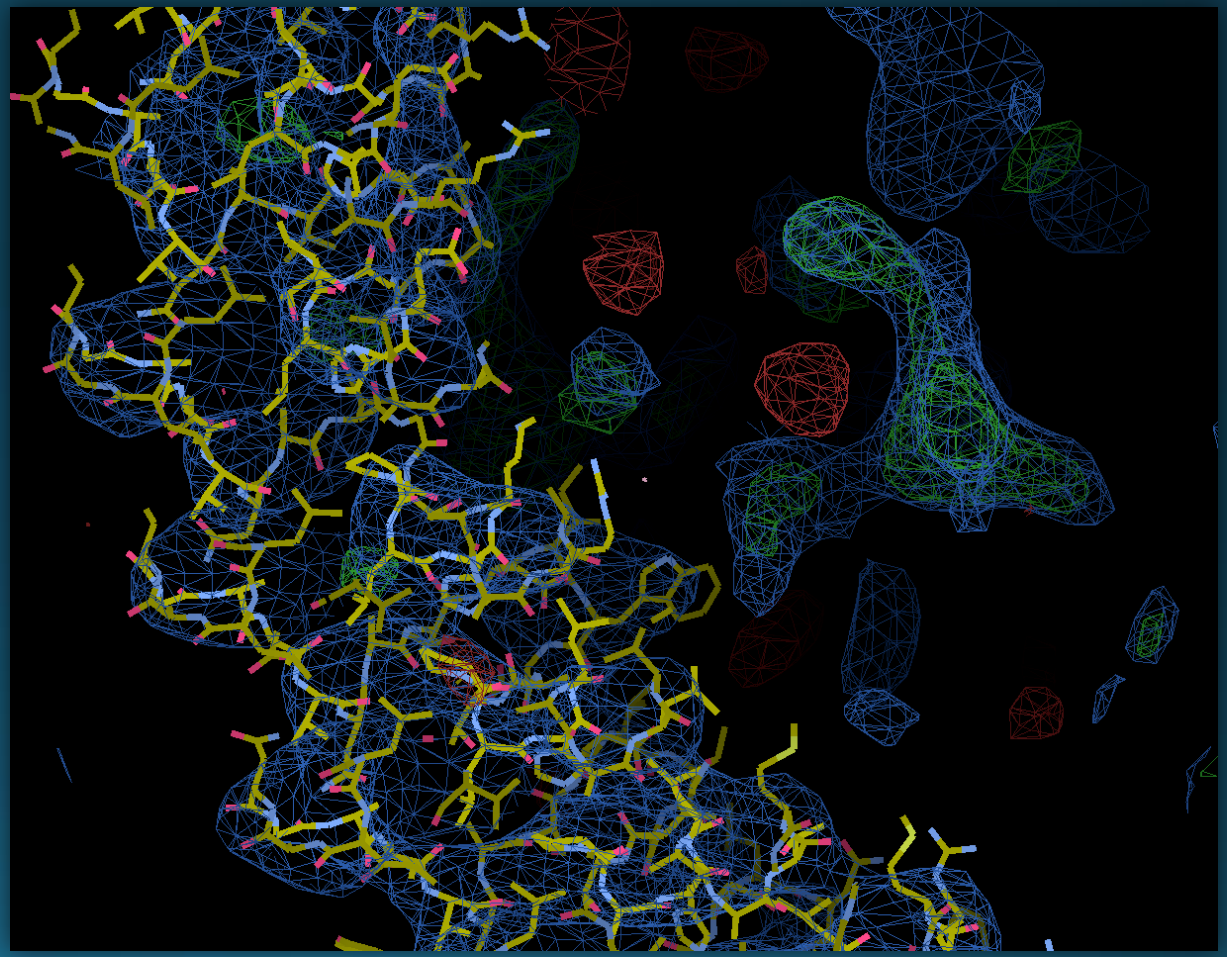


CCP4

MrBUMP

Step-by-Step MR: *Resolution and MR*

- Below 4Å automated model building becomes difficult

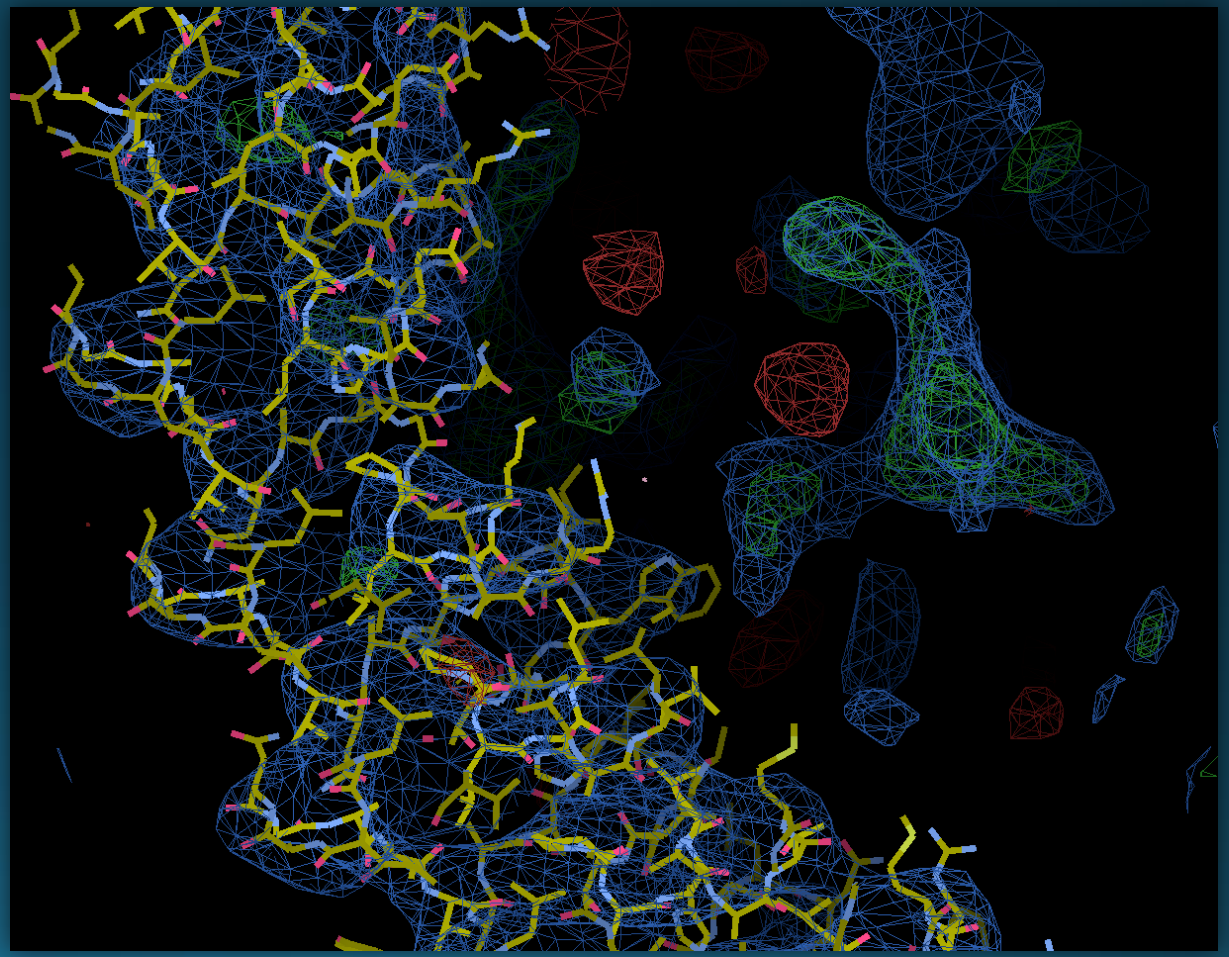


CCP4

MrBUMP

Step-by-Step MR: *Resolution and MR*

- Below 4Å automated model building becomes difficult
- However, given suitable search models, MR can be used to build up a model structure



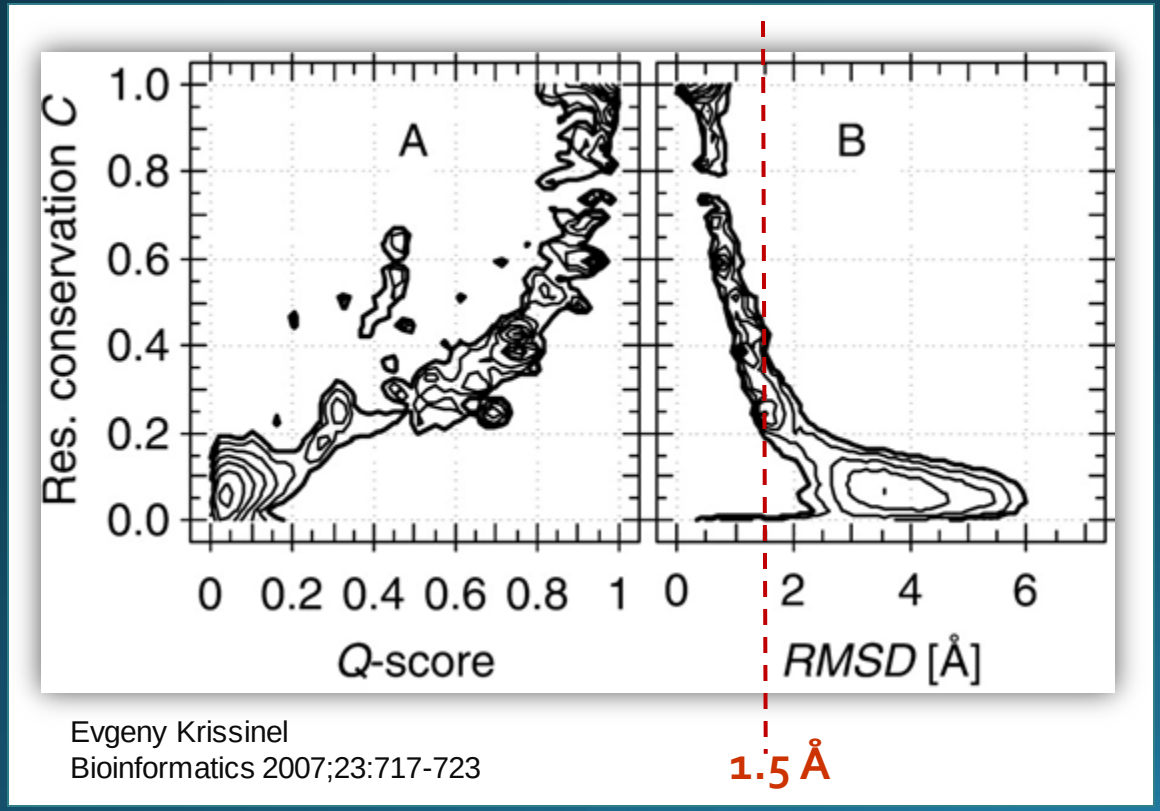
Step-by-Step MR: *Finding a search model*

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

Examine the data

↓

Find a search model



Evgeny Krissinel
Bioinformatics 2007;23:717-723

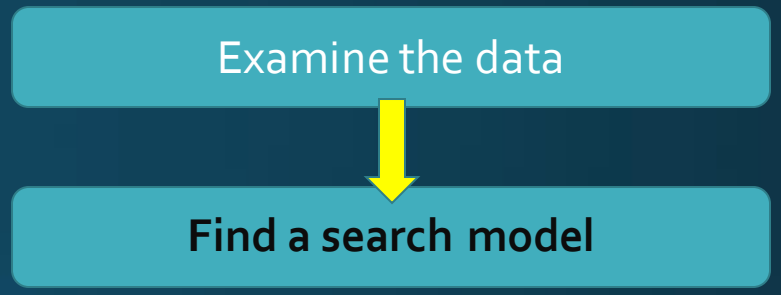
Krissinel et al. looked at structure-sequence relationship across entire PDB

CCP4

MrBUMP

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

- Bioinformatics gives us sensitive sequence-based search tools for finding potential homologues from which we can derive search models for MR



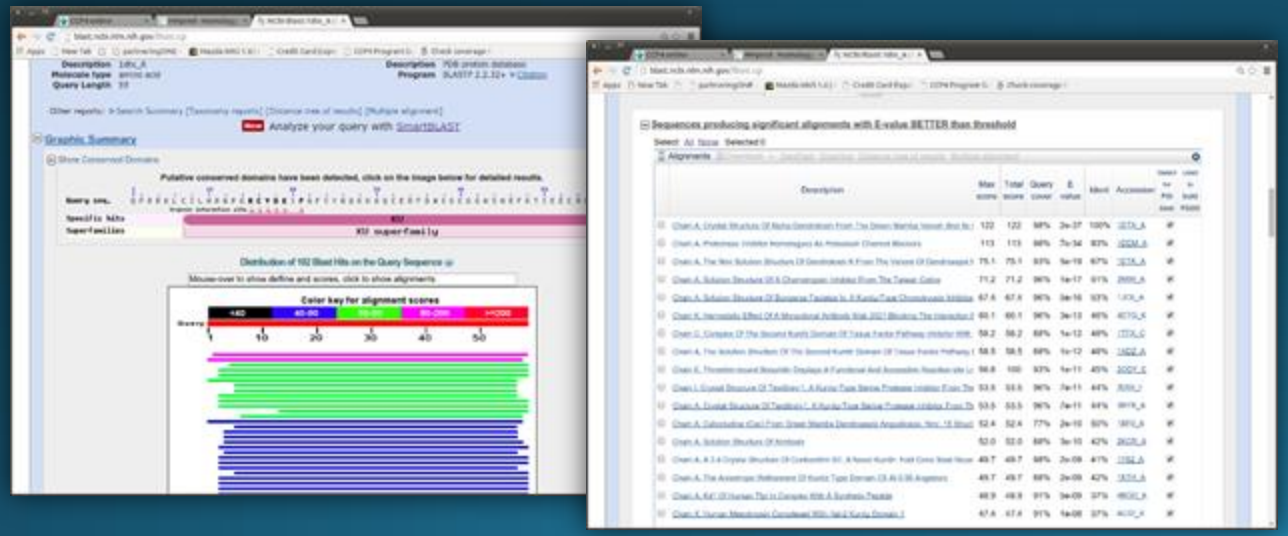
Step-by-Step MR: *Finding a search model*

- 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



CCP4

MrBUMP

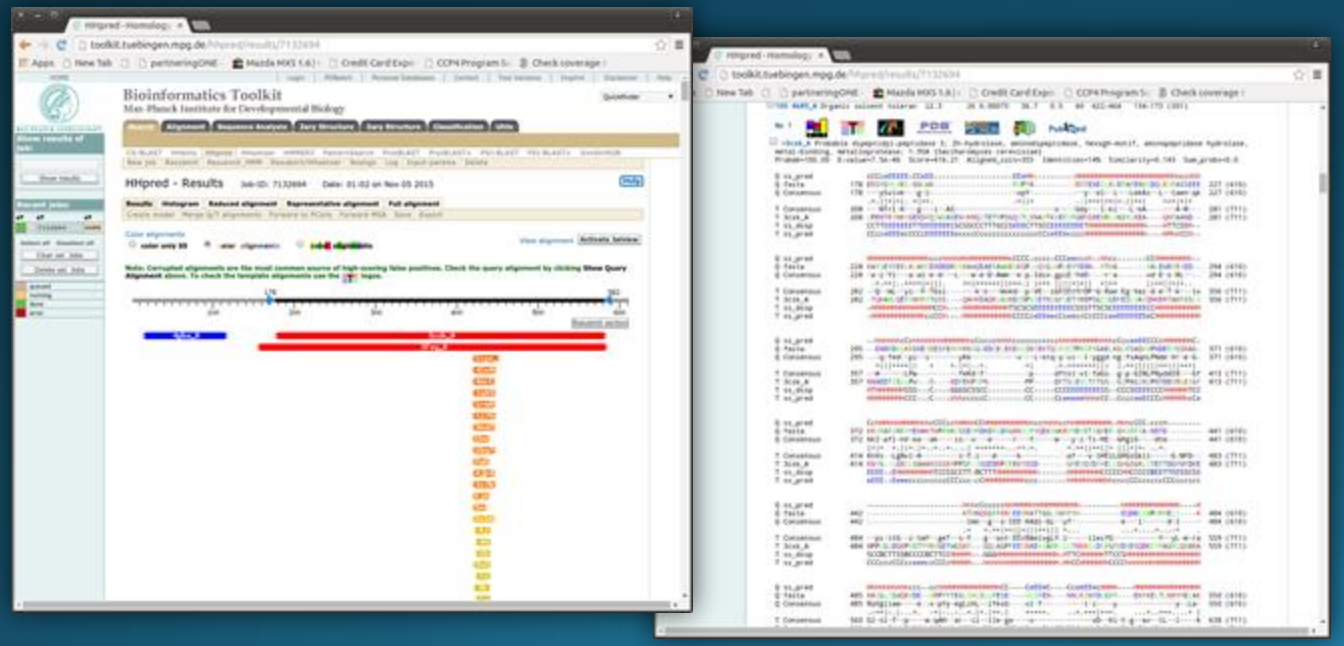
Step-by-Step MR: *Finding a search model*

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

Examine the data

↓

Find a search model

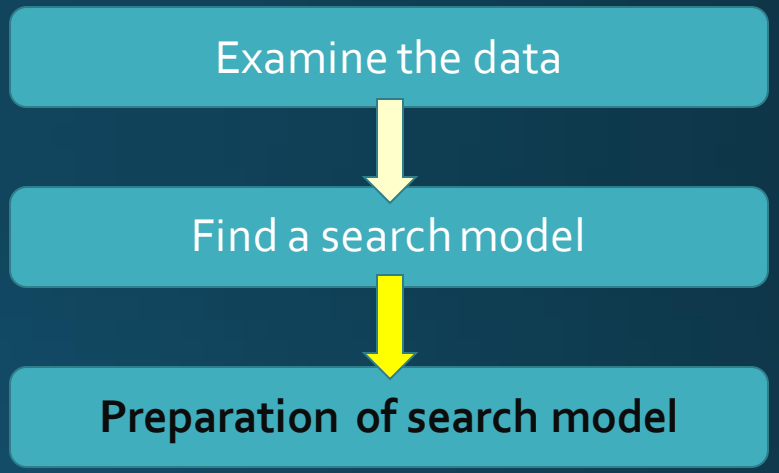


CCP4

MrBUMP

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

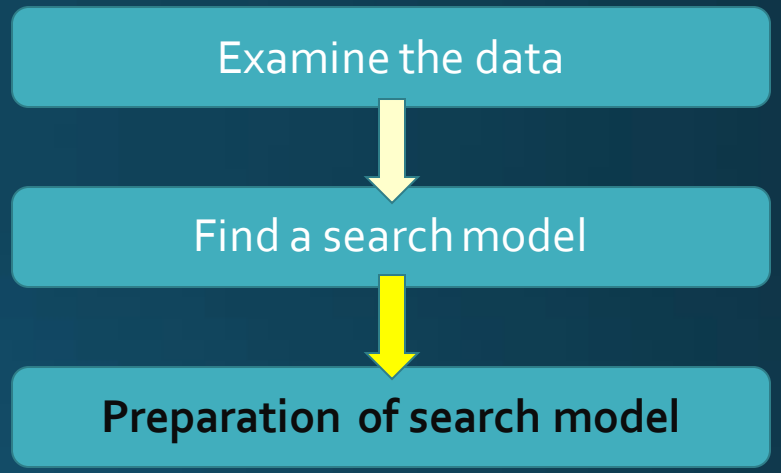
- Once a suitable homologue or set of homologues have been found they need to be prepared for use as search models



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

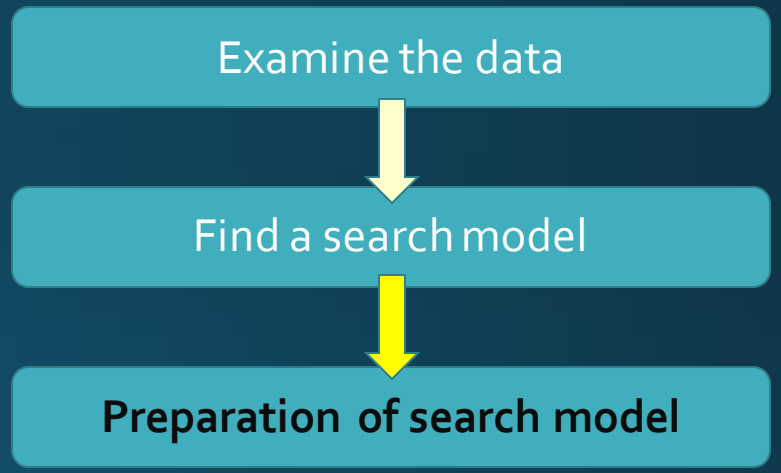


Step-by-Step MR: *Preparation of search model*

- 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

↓

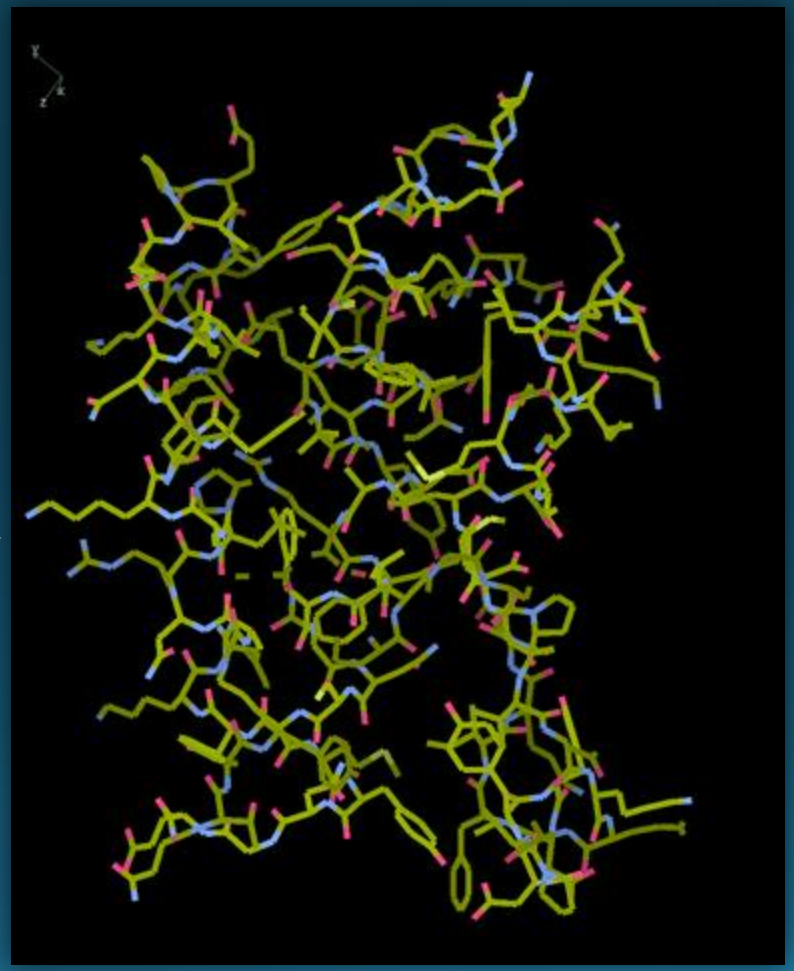
Preparation of search model

CCP4

MrBUMP

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

↓

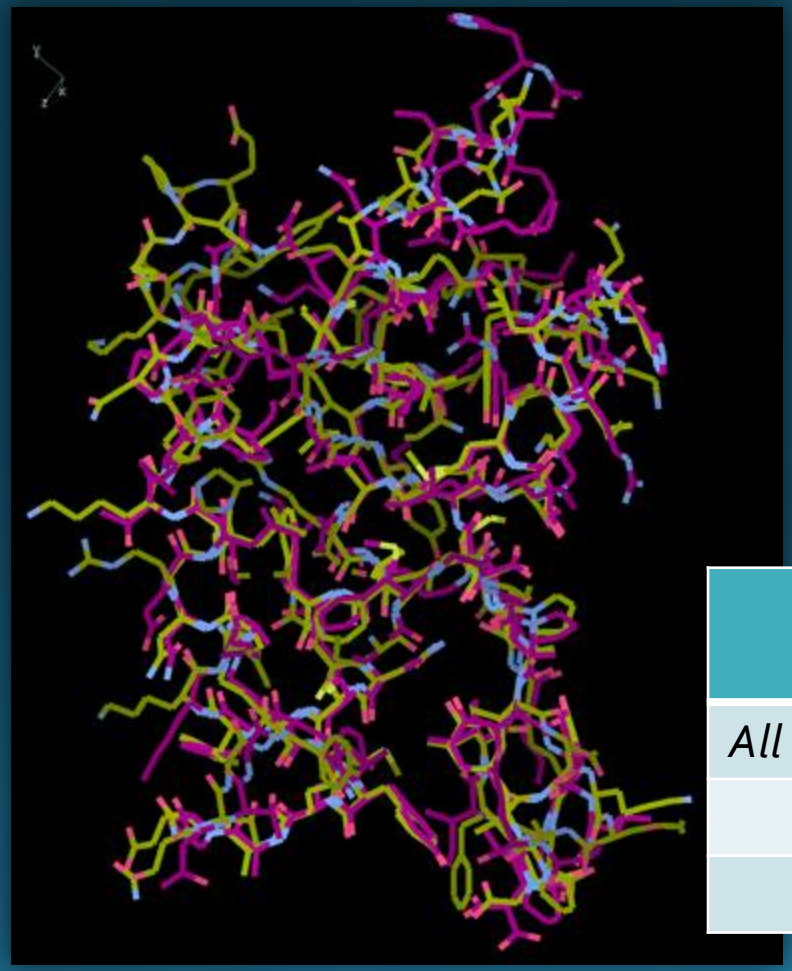
Preparation of search model

CCP4

MrBUMP

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

- Search Model preparation and MR:
  - All atoms?



Examine the data

↓

Find a search model

↓

Preparation of search model

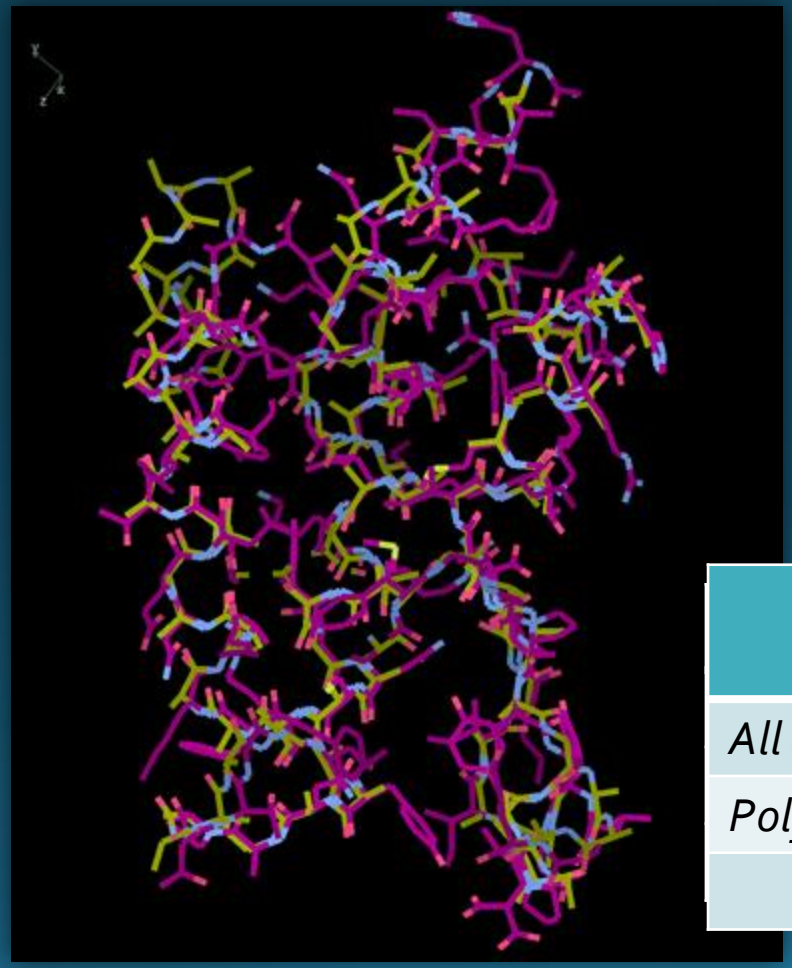
	Molrep TF/sigma	*Refmac* Rfactor	*Refmac* Rfree
All atoms	6.82	0.418	0.481

CCP4

MrBUMP

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

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



Examine the data

↓

Find a search model

↓

Preparation of search model

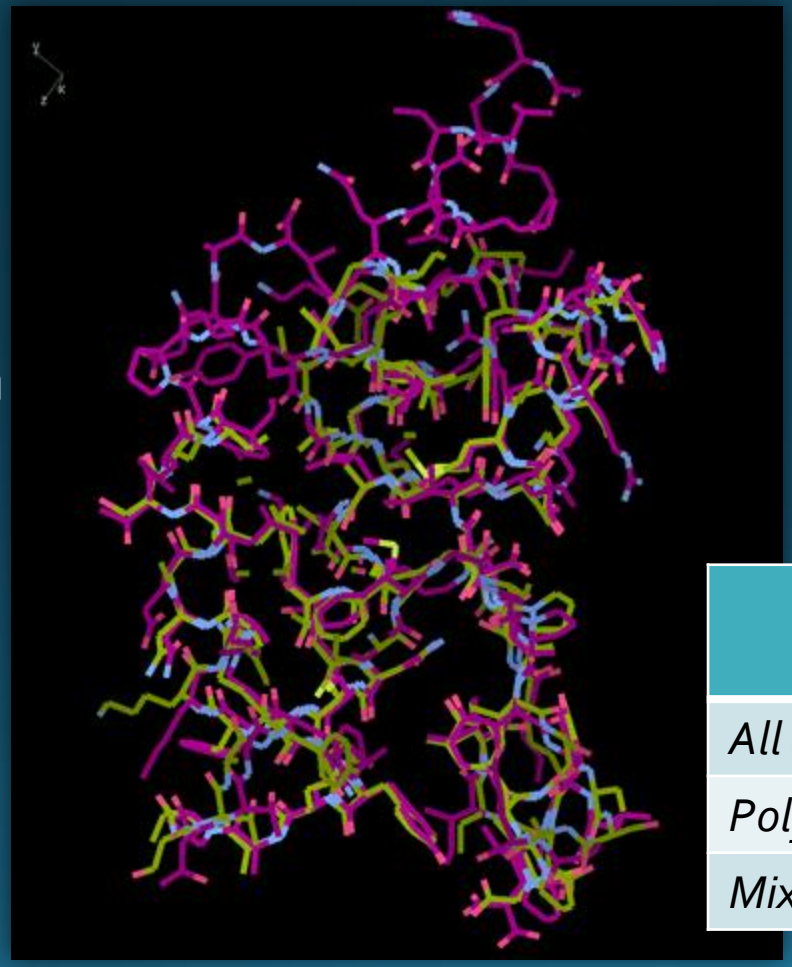
	Molrep TF/sigma	*Refmac* Rfactor	*Refmac* Rfree
All atoms	6.82	0.418	0.481
Poly	4.98	0.46	0.482

CCP4

MrBUMP

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



Examine the data

↓

Find a search model

↓

Preparation of search model

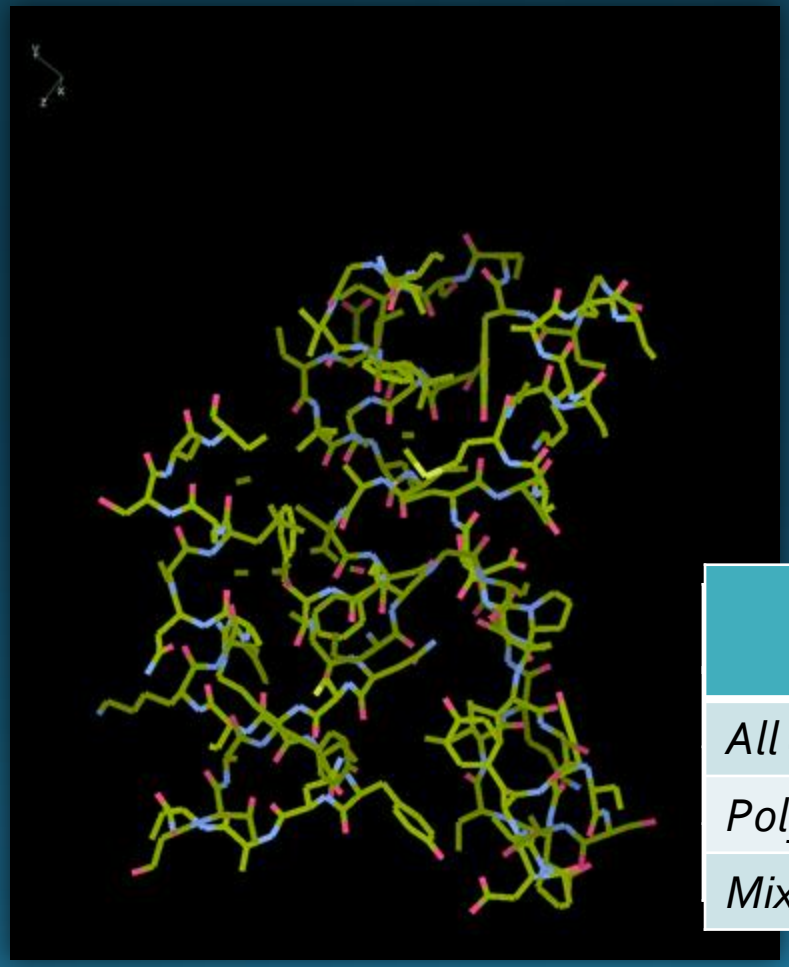
	Molrep TF/sigma	*Refmac* Rfactor	*Refmac* Rfree
All atoms	6.82	0.418	0.481
Poly	4.98	0.46	0.482
Mixed	10.27	0.379	0.418

CCP4

MrBUMP

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



Examine the data

↓

Find a search model

↓

Preparation of search model

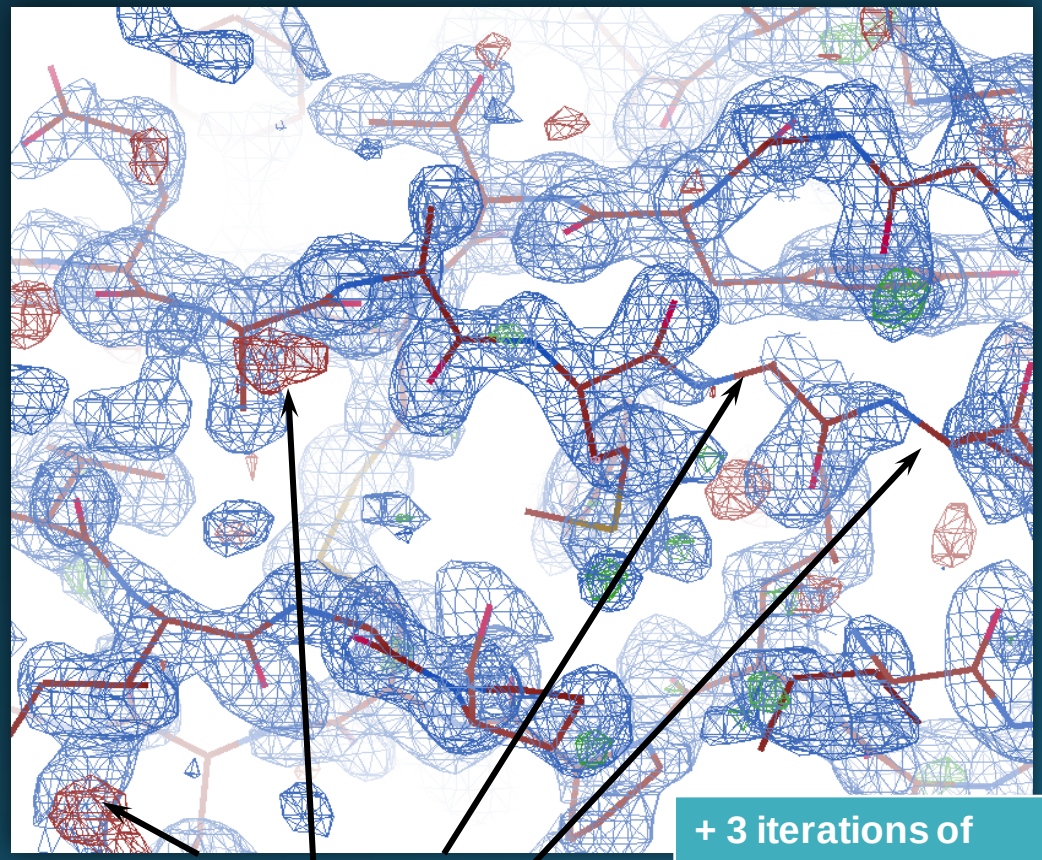
	Molrep TF/sigma	*Refmac* Rfactor	*Refmac* Rfree
All atoms	6.82	0.418	0.481
Poly	4.98	0.46	0.482
Mixed	10.27	0.379	0.418

CCP4

MrBUMP

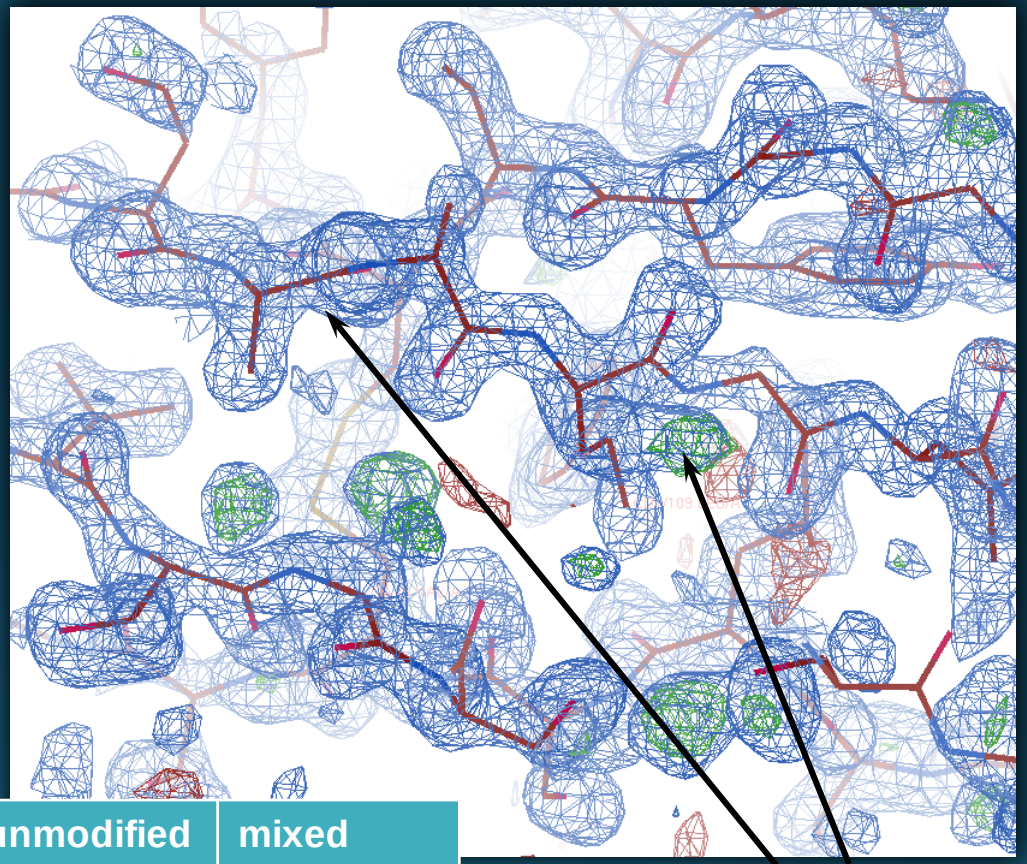
Step-by-Step MR: *Preparation of search model*

Unmodified model:



bad density features

Mixed model:



good density features

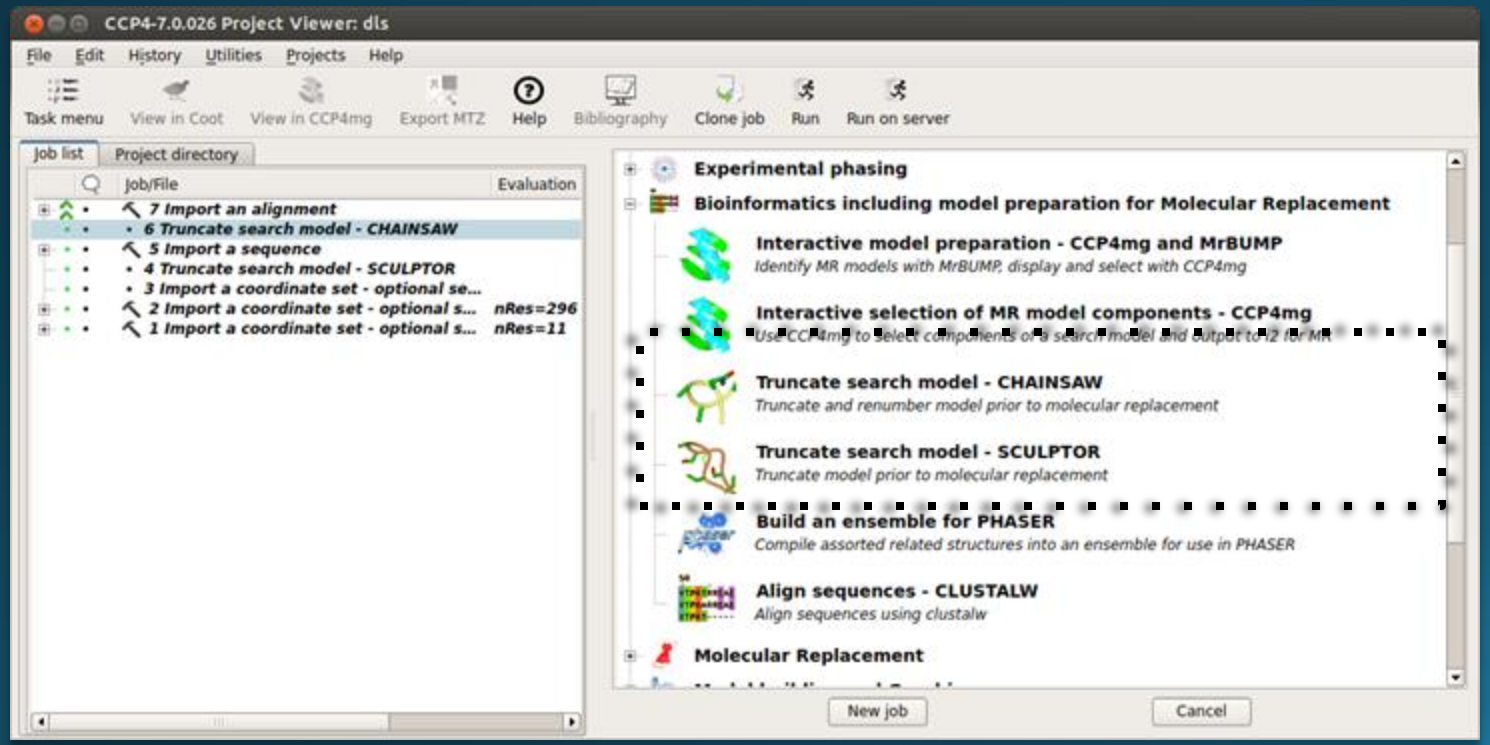
+ 3 iterations of buccaneer & refmac	unmodified	mixed
R	32	30
R-free	36	33

CCP4

MrBUMP

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

- Mixed-model generation in CCP4
  - Chainsaw & Sculptor



Examine the data

↓

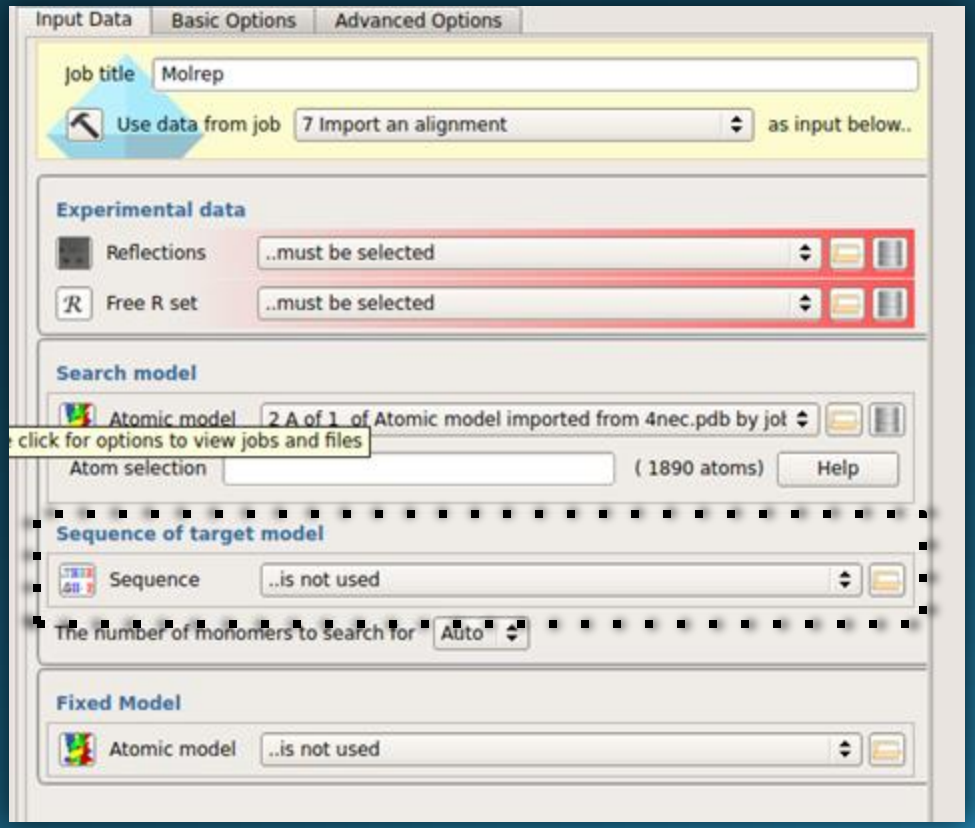
Find a search model

↓

Preparation of search model

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

- Mixed-model generation in CCP4
  - Molrep



Examine the data

↓

Find a search model

↓

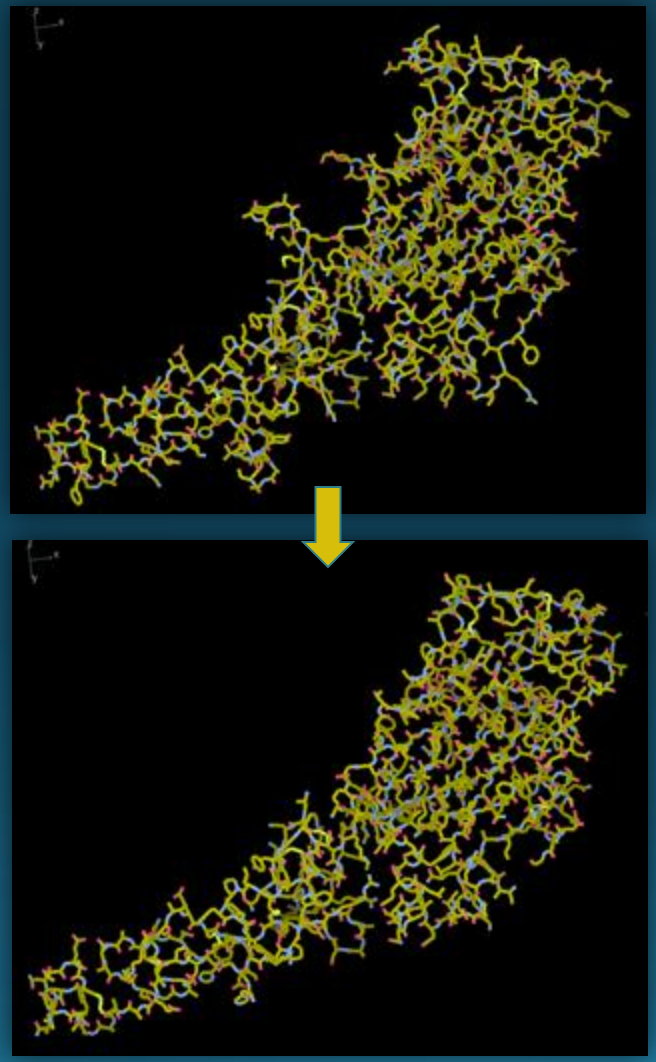
Preparation of search model

CCP4

MrBUMP

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"



Examine the data

↓

Find a search model

↓

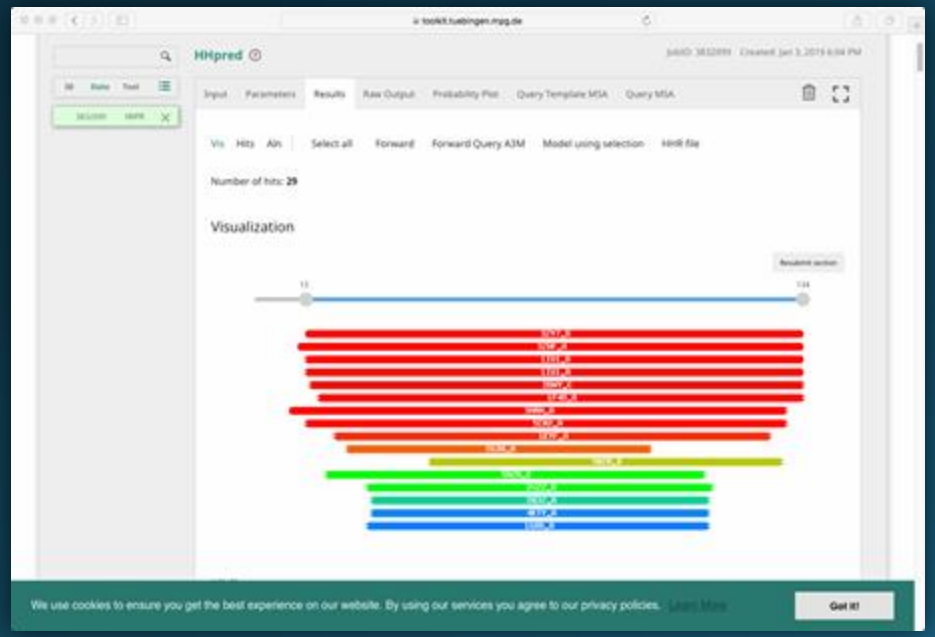
Preparation of search model

CCP4

MrBUMP

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

- Ensembles: alignment of search models : several homologues available



Examine the data

↓

Find a search model

↓

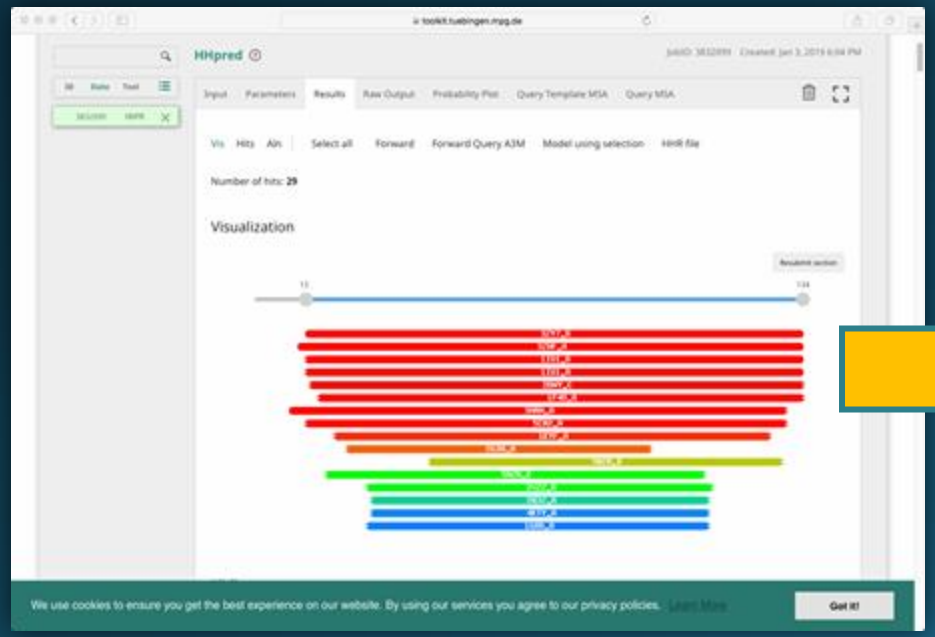
Preparation of search model

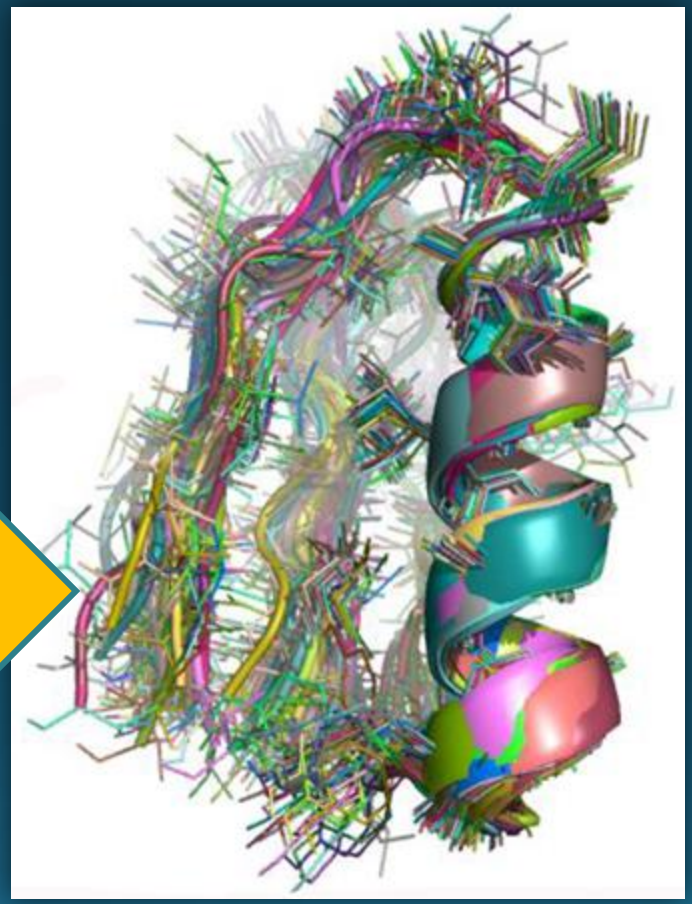
CCP4

MrBUMP

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

- Ensembles: alignment of search models





Examine the data

Find a search model

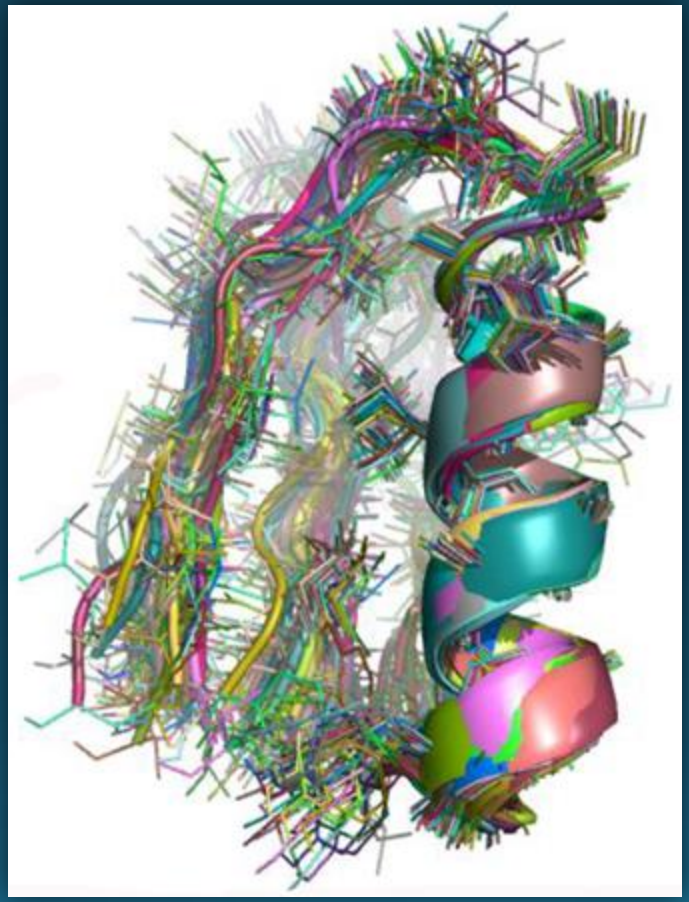
Preparation of search model

CCP4

MrBUMP

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

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



Examine the data

↓

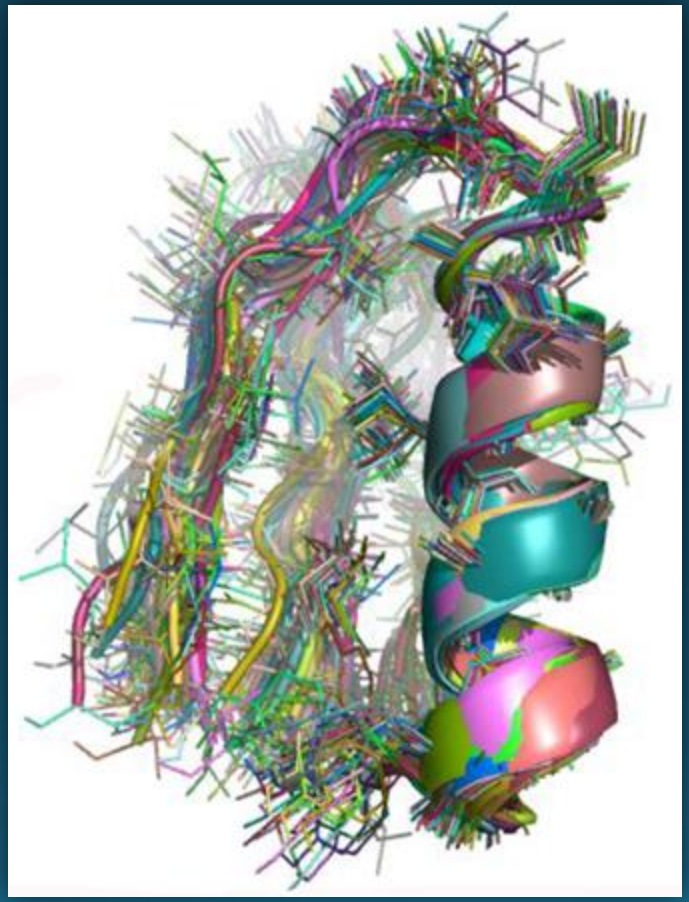
Find a search model

↓

Preparation of search model

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



Examine the data

↓

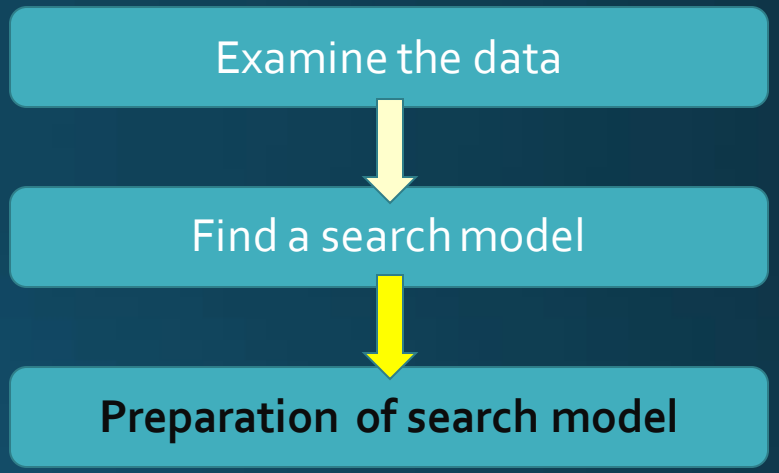
Find a search model

↓

Preparation of search model

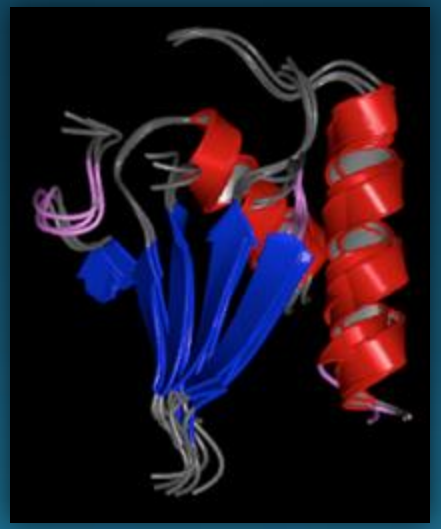
Step-by-Step MR: *Preparation of search model*

- 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



Examine the data

↓

Find a search model

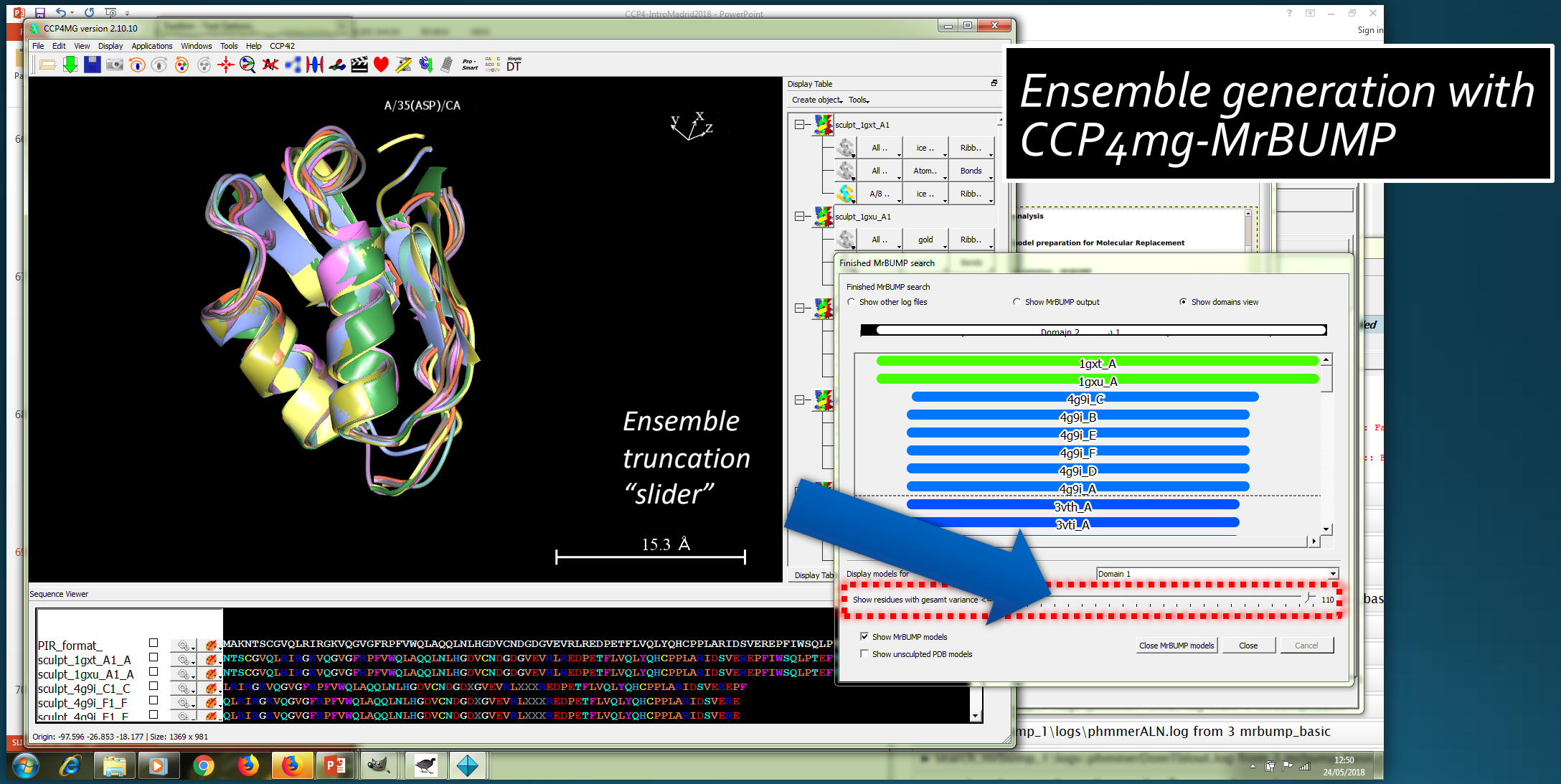
↓

Preparation of search model

CCP4

MrBUMP

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



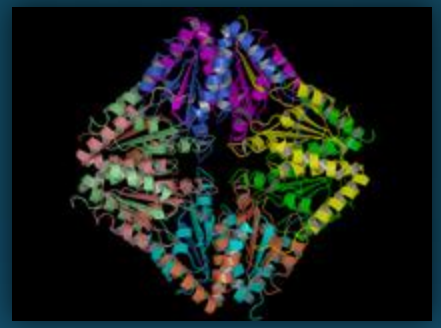
Ensemble generation with CCP4mg-MrBUMP

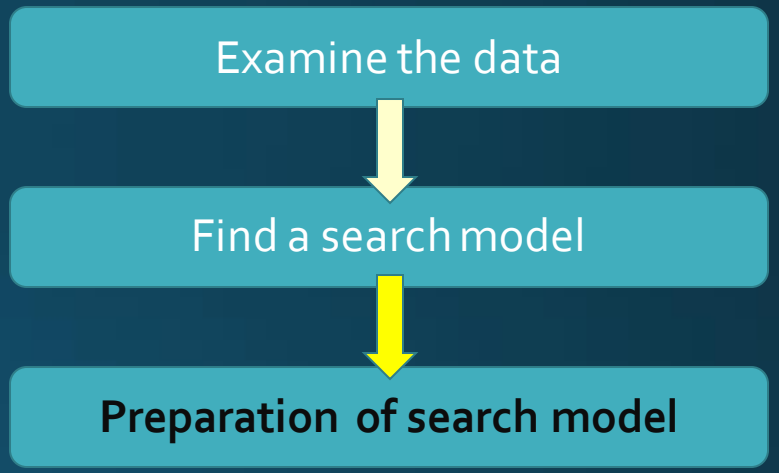
CCP4

MrBUMP

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

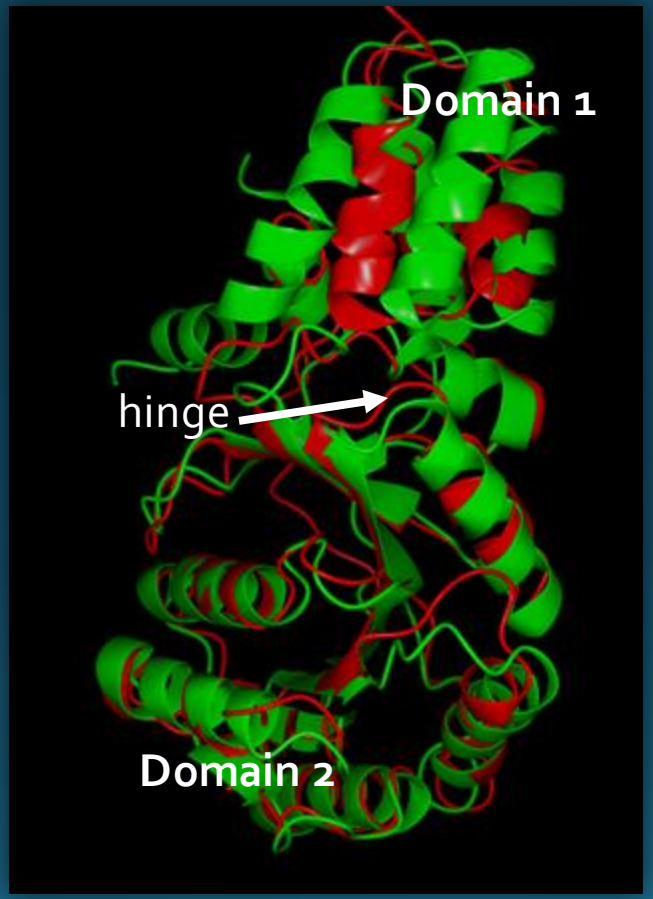
- Multimers as search models
  - A single chain search model can be too small if the target has crystallised in multimeric form
  - The signal for the correct position is too weak against the background noise of incorrect positions
  - Particularly a problem at lower resolutions and crystals with high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



Examine the data

↓

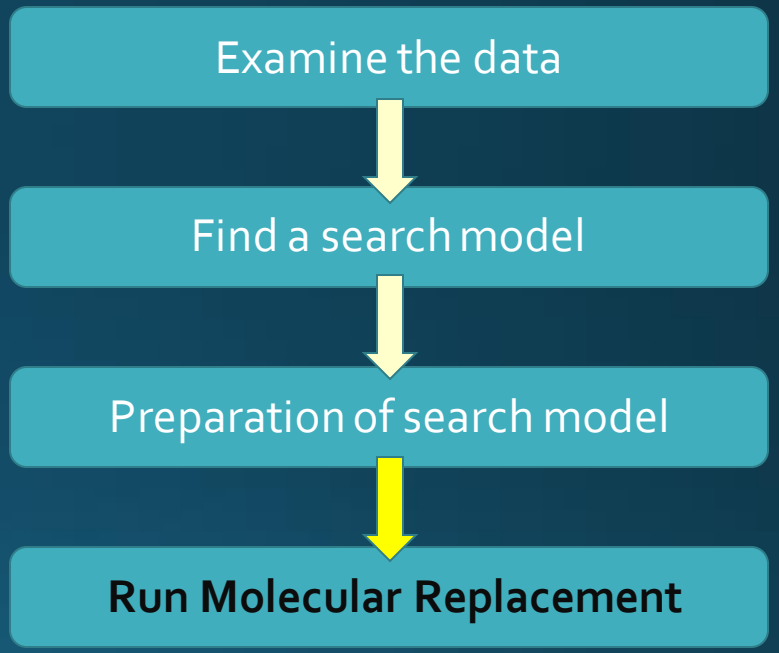
Find a search model

↓

Preparation of search model

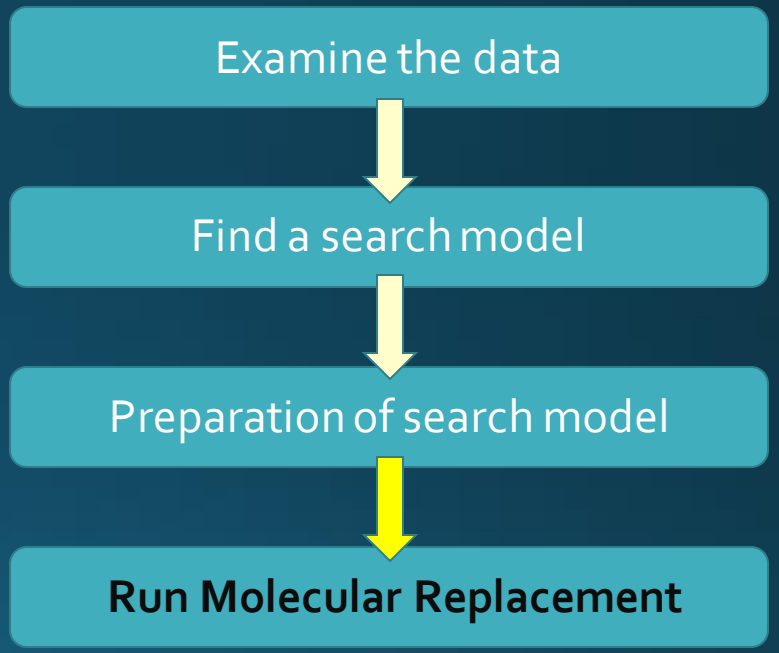
Step-by-Step MR: *Run Molecular Replacement*

- 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



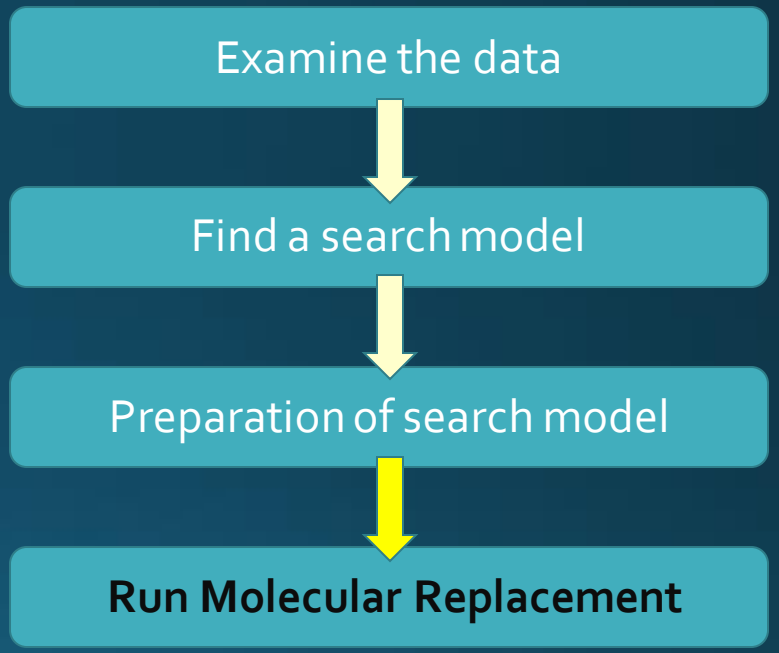
Step-by-Step MR: *Run Molecular Replacement*

- Important points on using Phaser



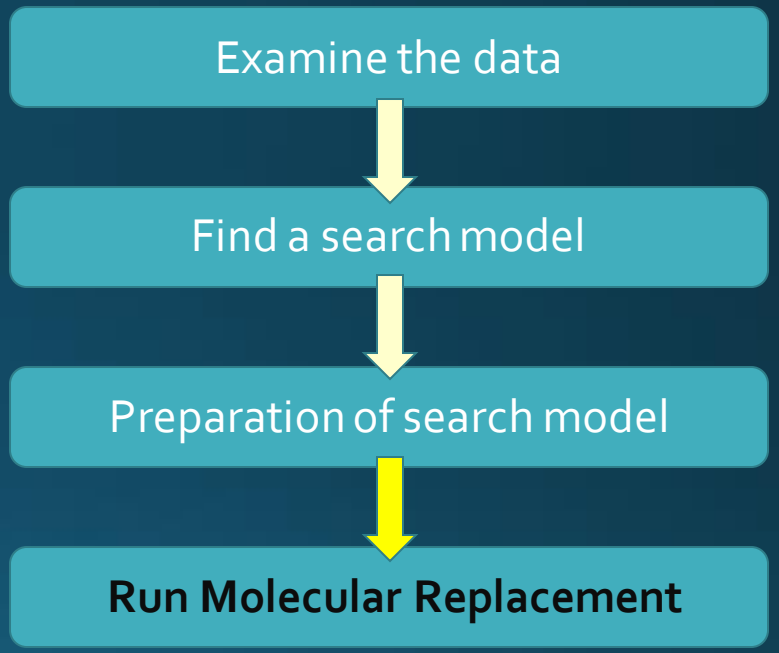
Step-by-Step MR: *Run Molecular Replacement*

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



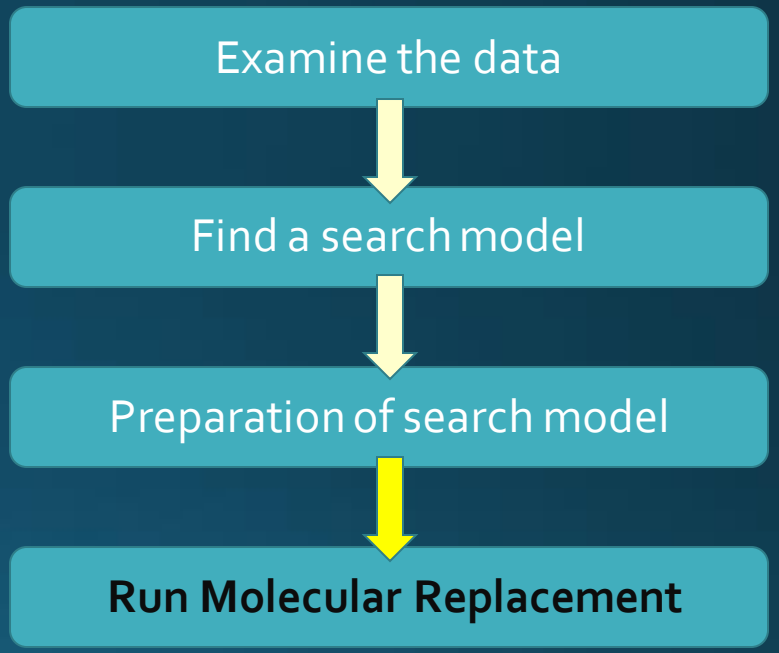
Step-by-Step MR: *Run Molecular Replacement*

- 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



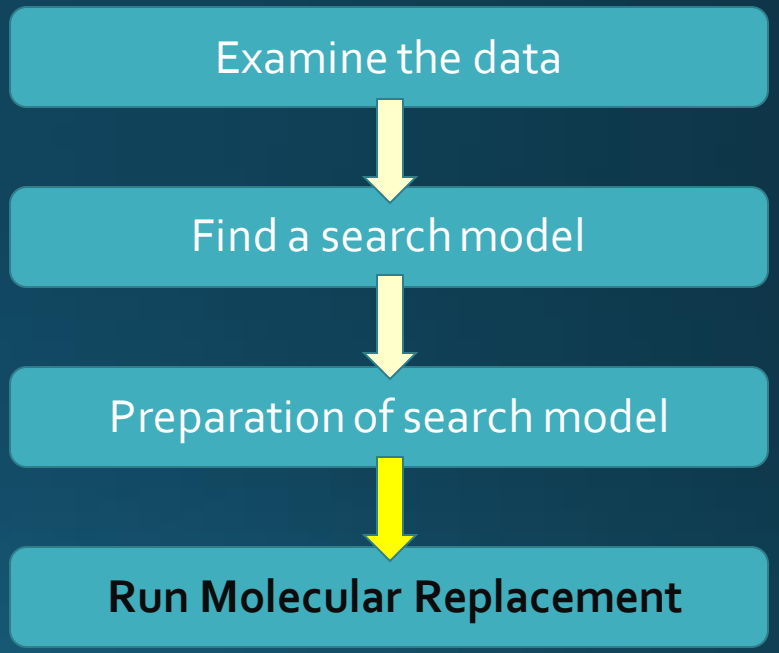
Step-by-Step MR: *Run Molecular Replacement*

- 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

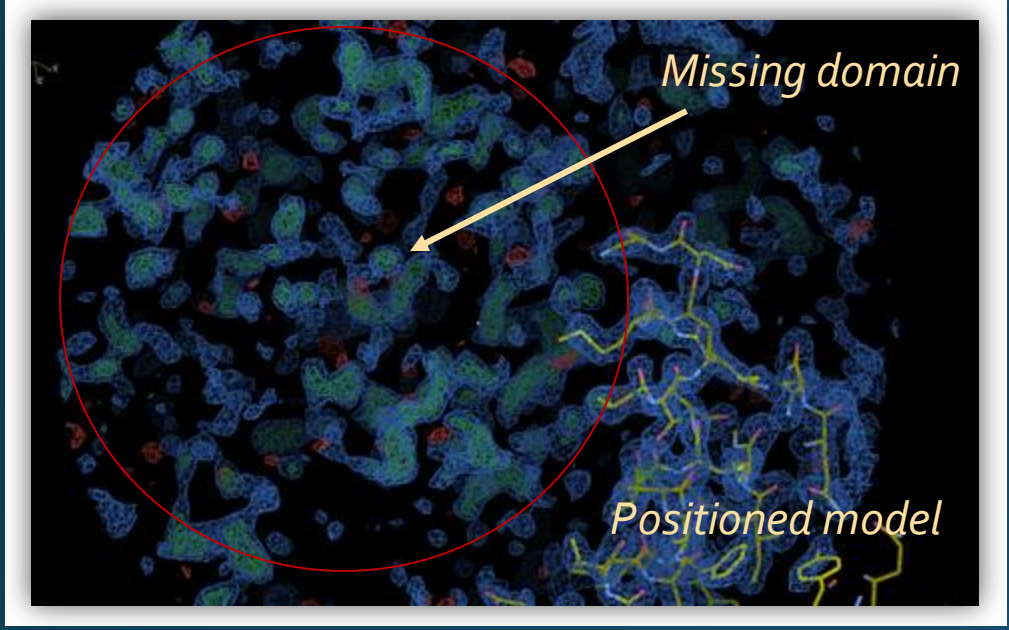


Step-by-Step MR: *Run Molecular Replacement*

- 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

Examine the data

Find a search model

Preparation of search model

Run Molecular Replacement

CCP4

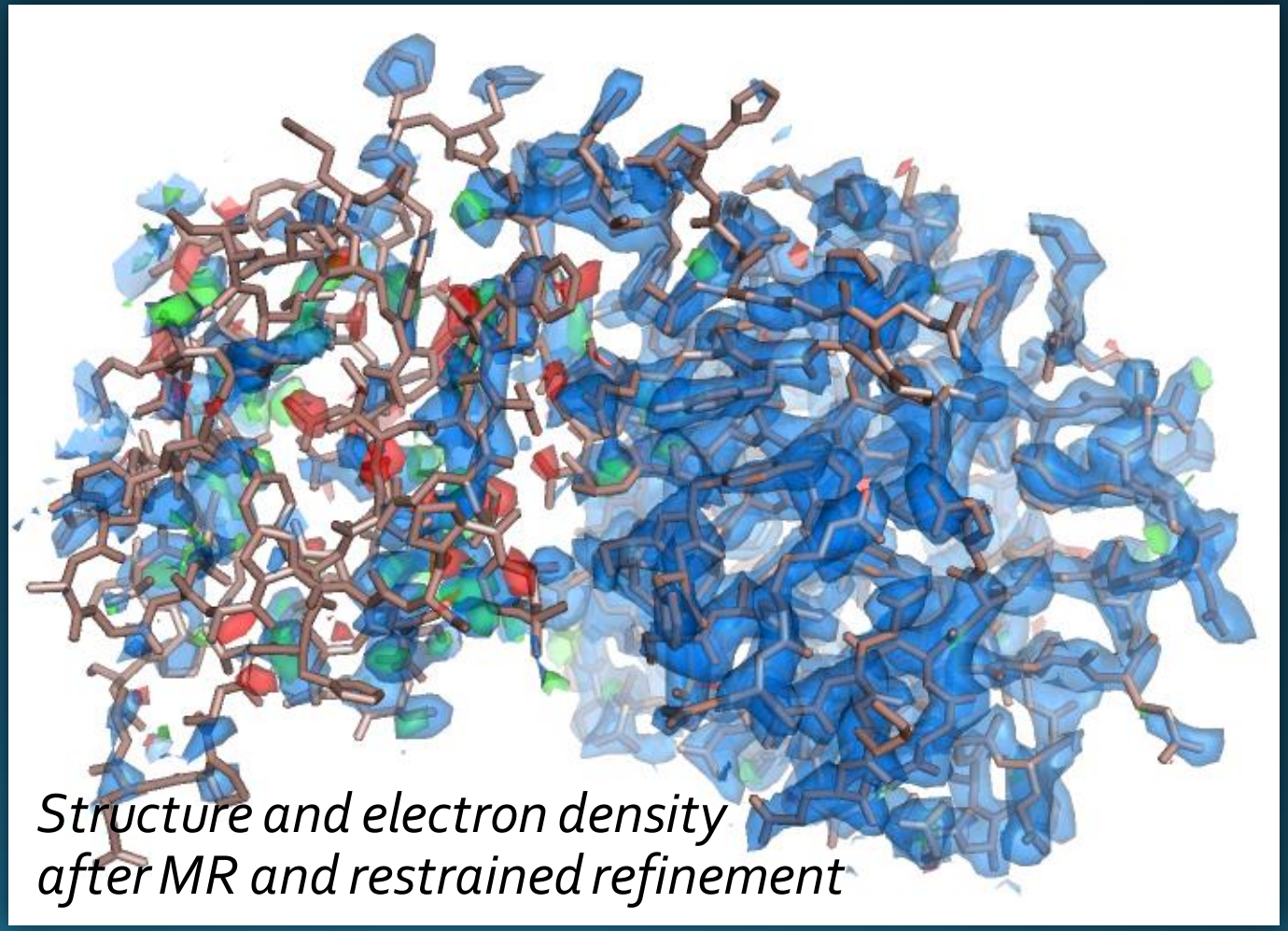
MrBUMP

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

Phased translation search:

Example: ***1tj3***

Search model: 1s20, chain A



Structure and electron density after MR and restrained refinement

CCP4

MrBUMP

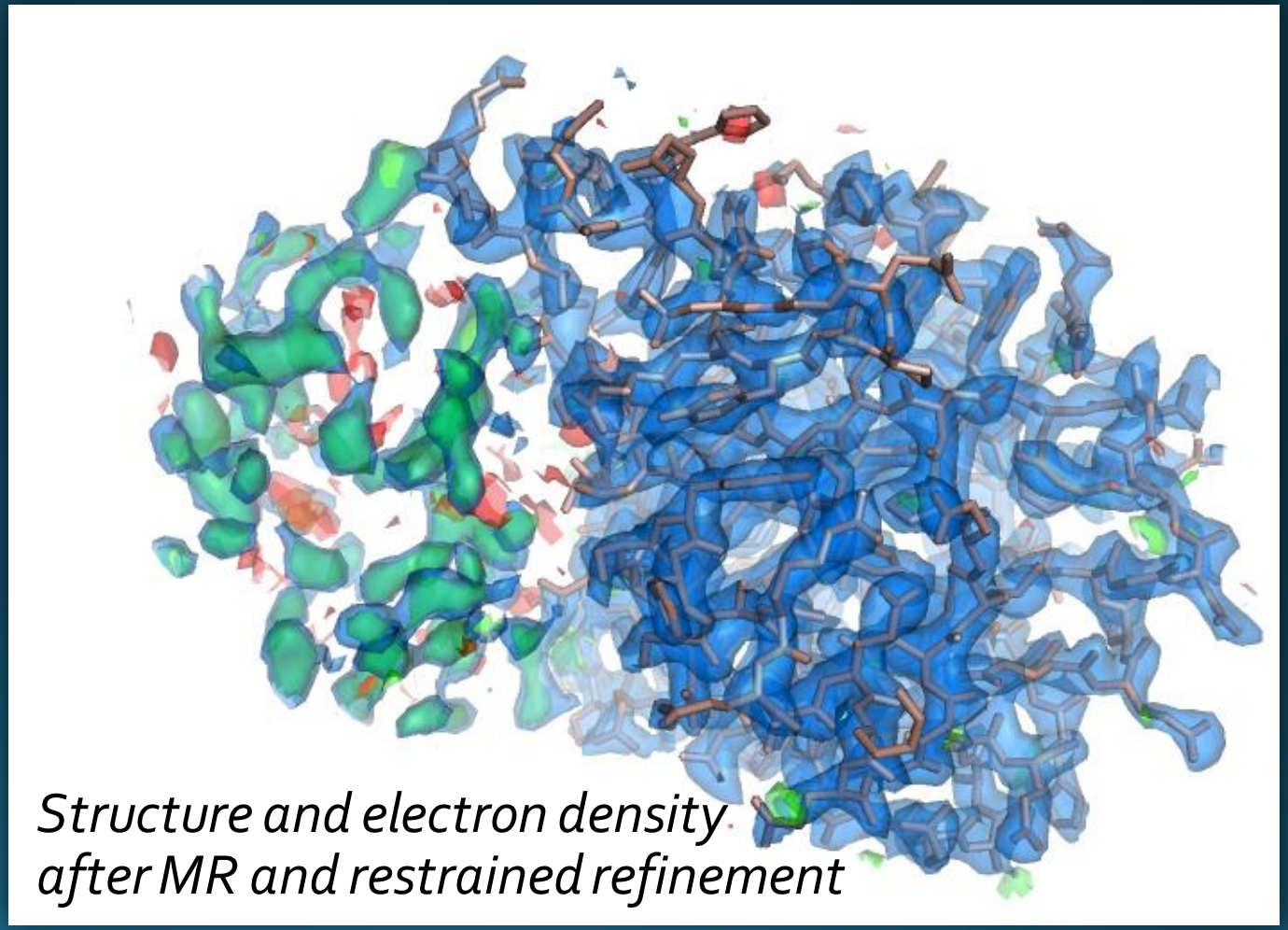
Step-by-Step MR: *Run Molecular Replacement*

Phased translation search:

Example: ***1tj3***

Search model: 1s20, chain A

Large Domain only



Structure and electron density after MR and restrained refinement

CCP4

MrBUMP

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

Phased translation search:

Example: ***1tj3***

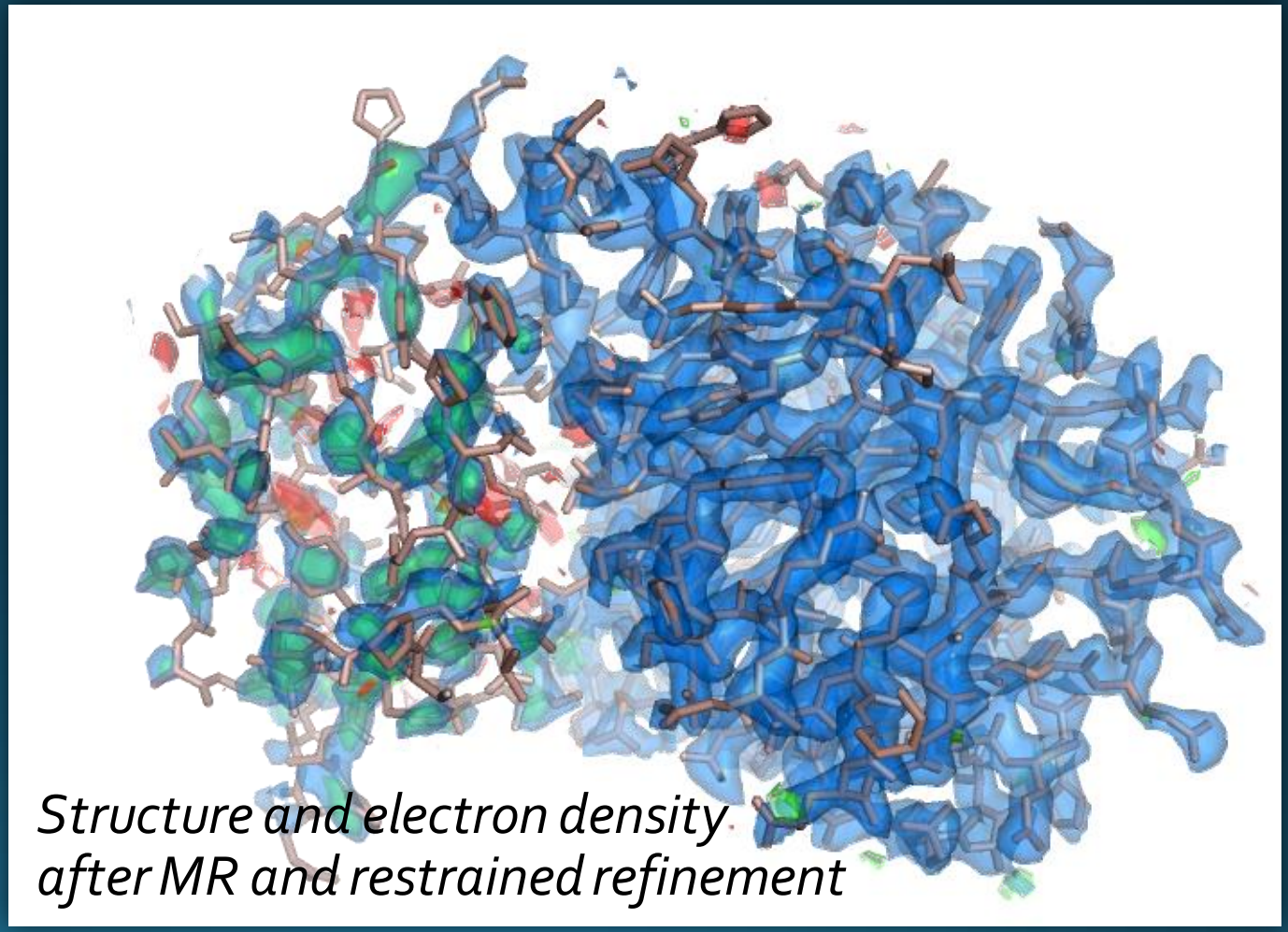
Search model: 1s20, chain A

Large Domain fixed

Small Domain search

Fit into density

(Molrep & Phaser)



Structure and electron density after MR and restrained refinement

CCP4

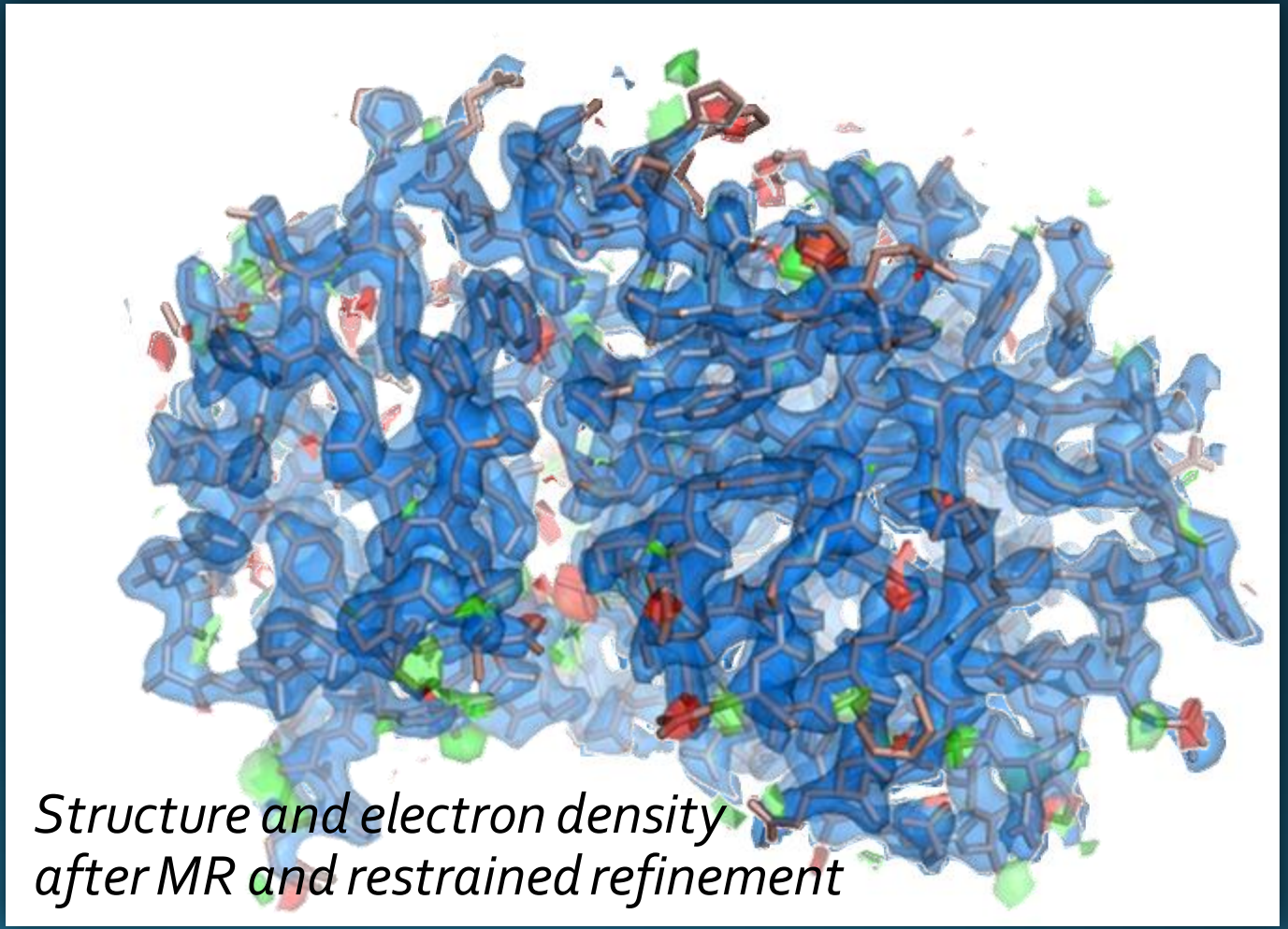
MrBUMP

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

Phased translation search:

Example: ***1tj3***

Complete structure and electron density after restrained refinement

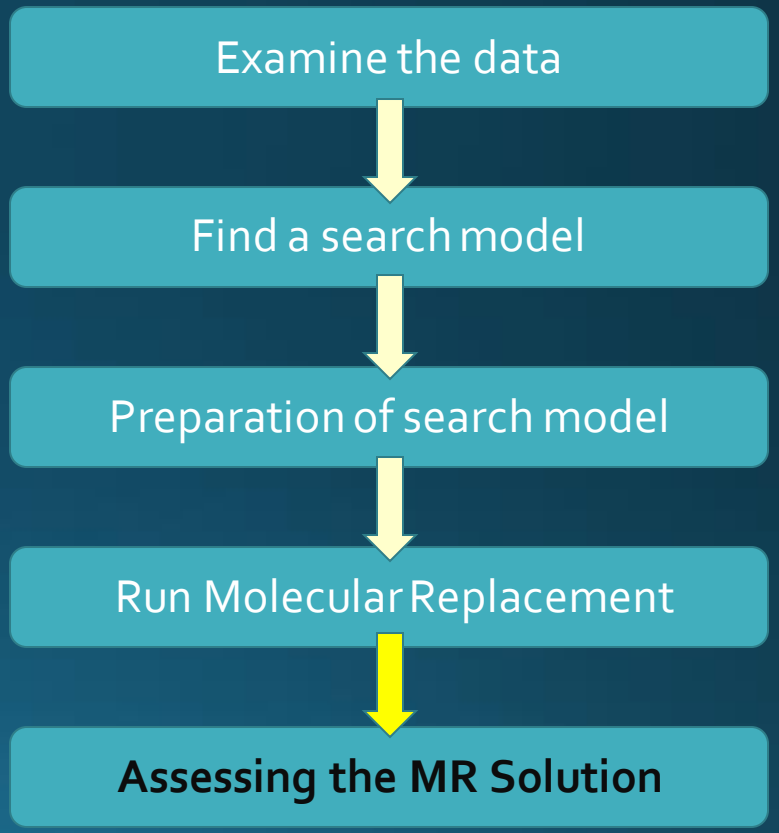


Structure and electron density after MR and restrained refinement

How do I know my MR solution is successful?

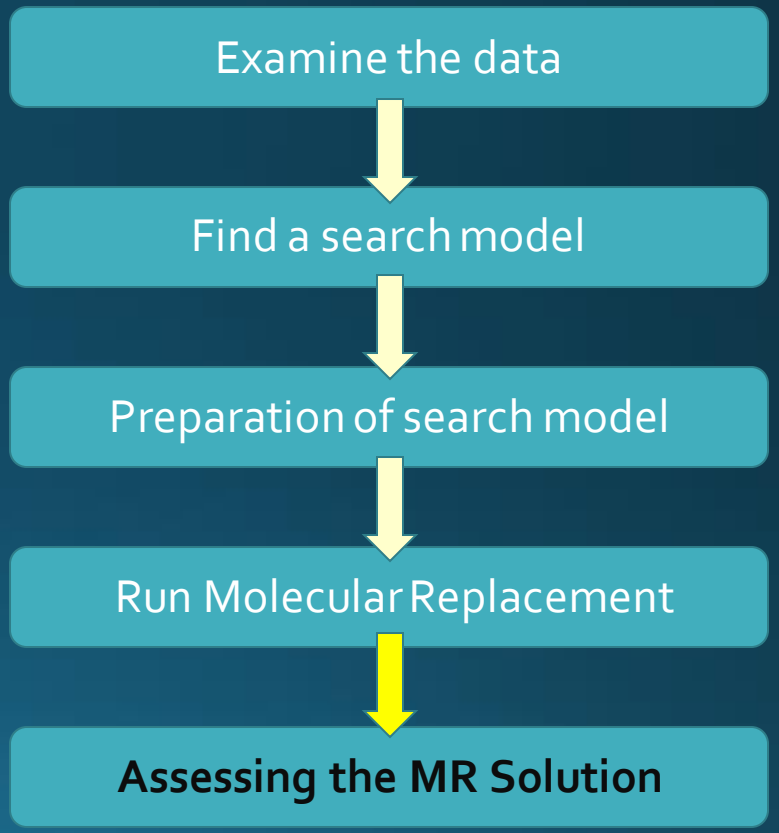
Step-by-Step MR: *Assessing the MR solution*

- Terminology
  - MR solution
  - Successful MR solution
- What is a successful MR solution?
  - A search model placed in the target unit cell such that it will provide us with sufficiently accurate phase estimates for the target
  - Enables us to complete the model through model building and refinement

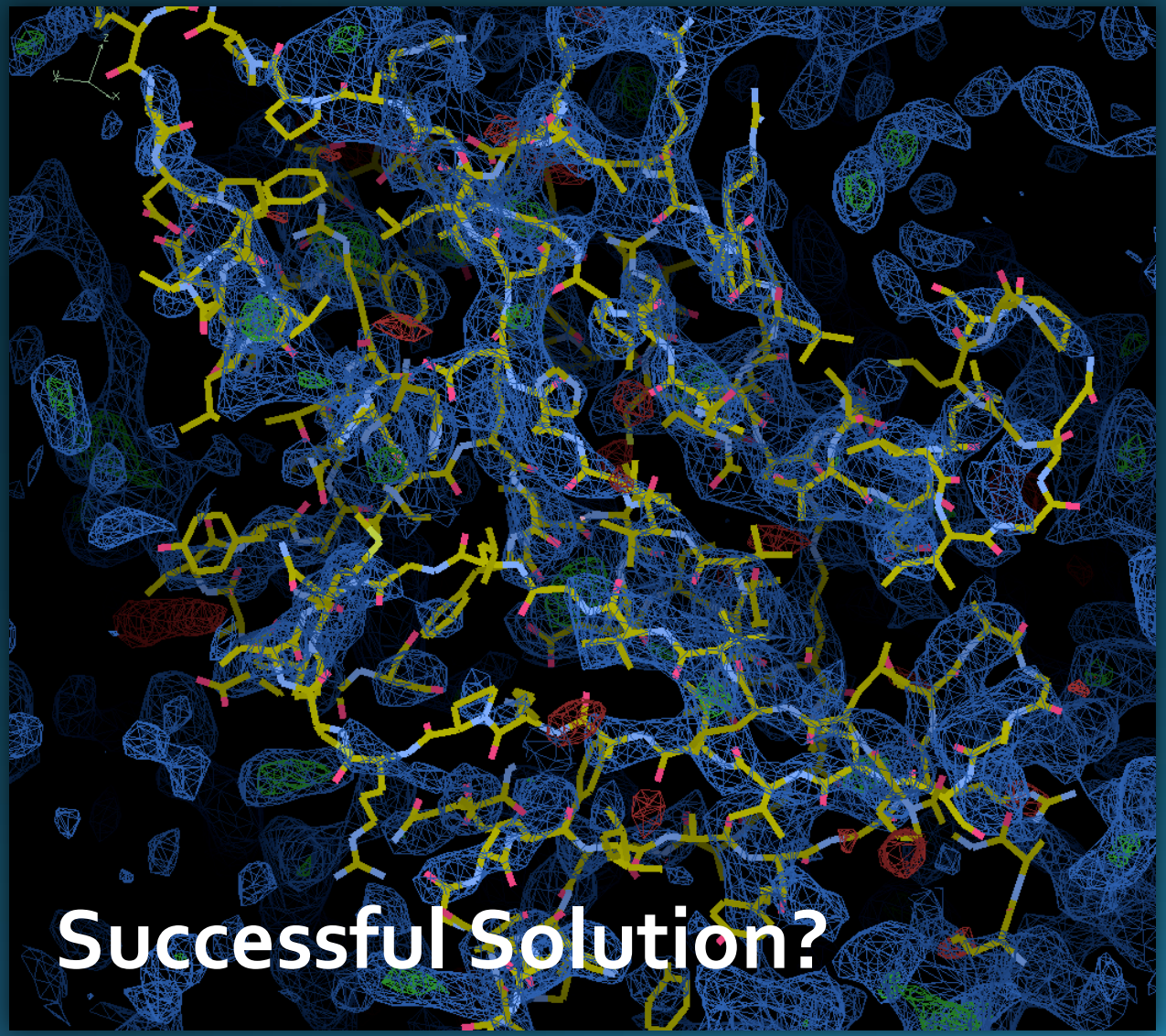


Step-by-Step MR: *Assessing the MR solution*

- In difficult cases the position may be correct but getting from the MR solution to a complete model may not be straight forward
- Assessment often involves performing additional structure solutions steps such as refinement and density modification



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

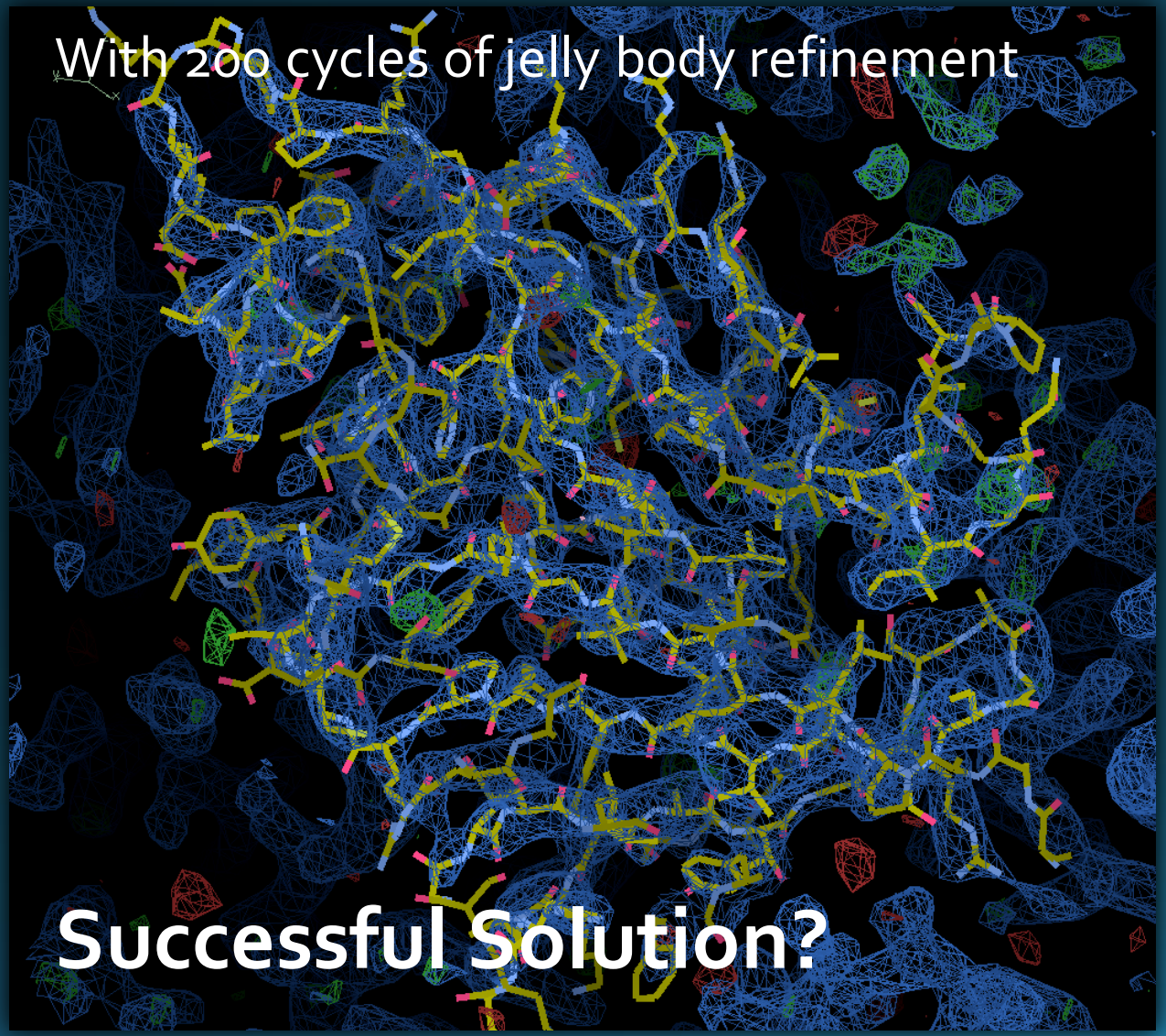


- Examine the data
- Find a search model
- Preparation of search model
- Run Molecular Replacement
- **Assessing the MR Solution**

CCP4

MrBUMP

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

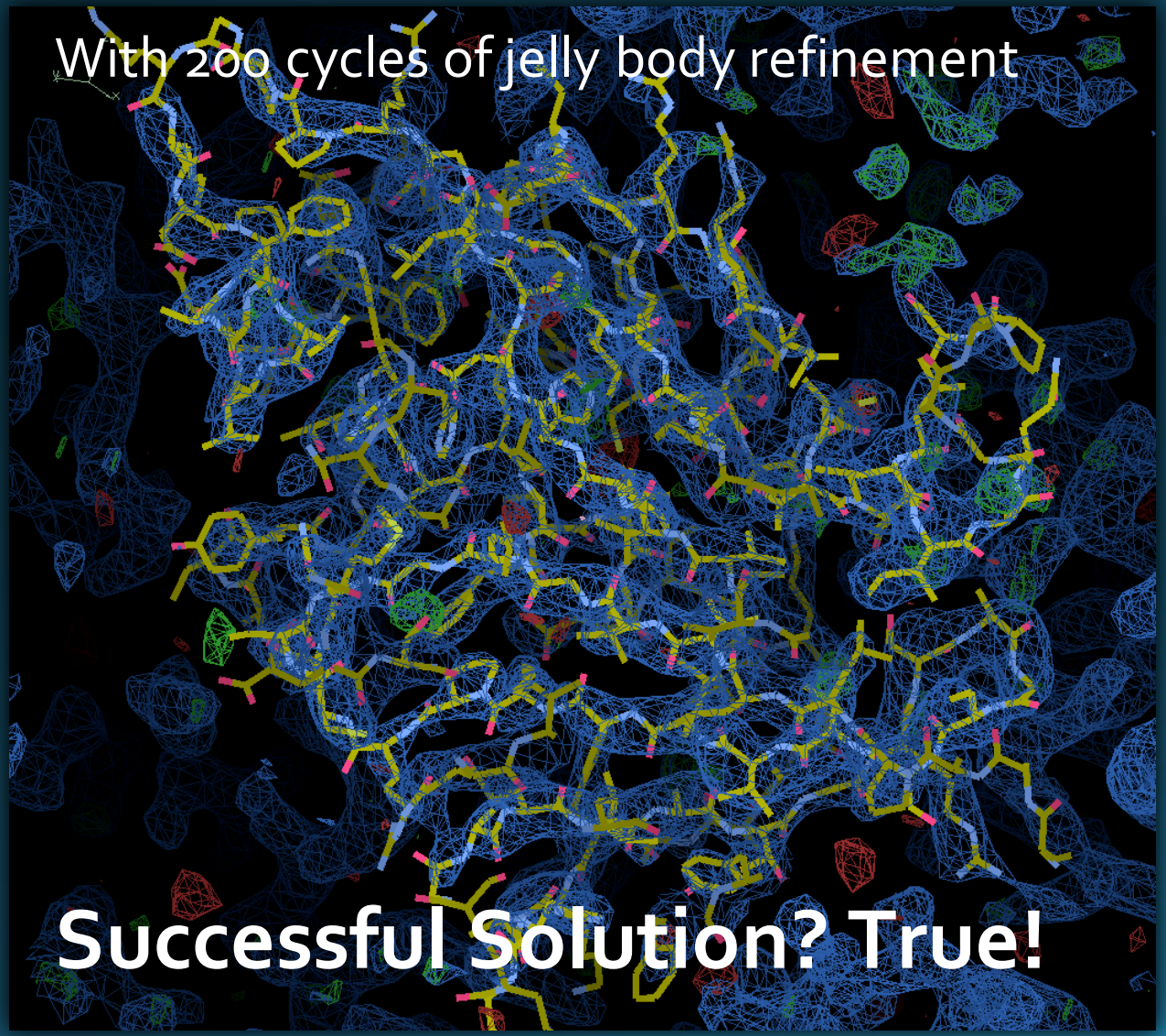


- Examine the data
- Find a search model
- Preparation of search model
- Run Molecular Replacement
- **Assessing the MR Solution**

CCP4

MrBUMP

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



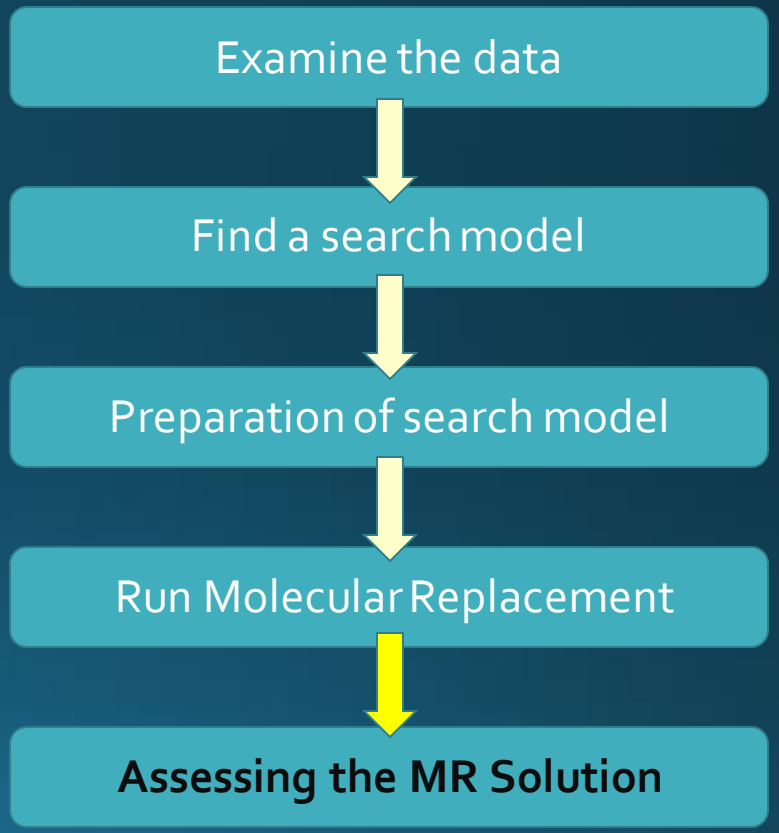
- Examine the data
- Find a search model
- Preparation of search model
- Run Molecular Replacement
- **Assessing the MR Solution**

CCP4

MrBUMP

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

- Rough guide to MR program scoring

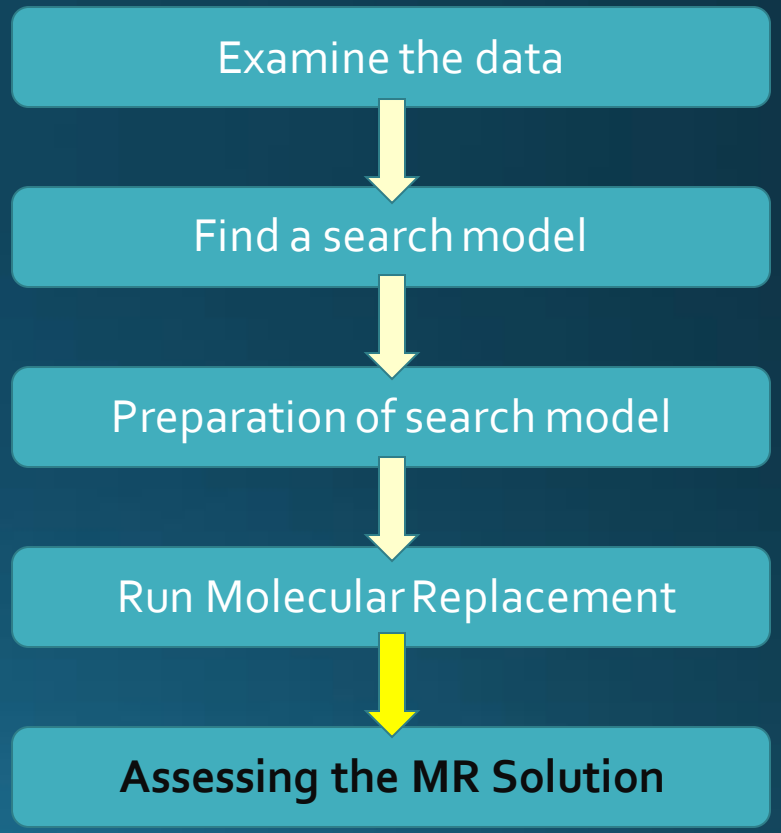


CCP4

MrBUMP

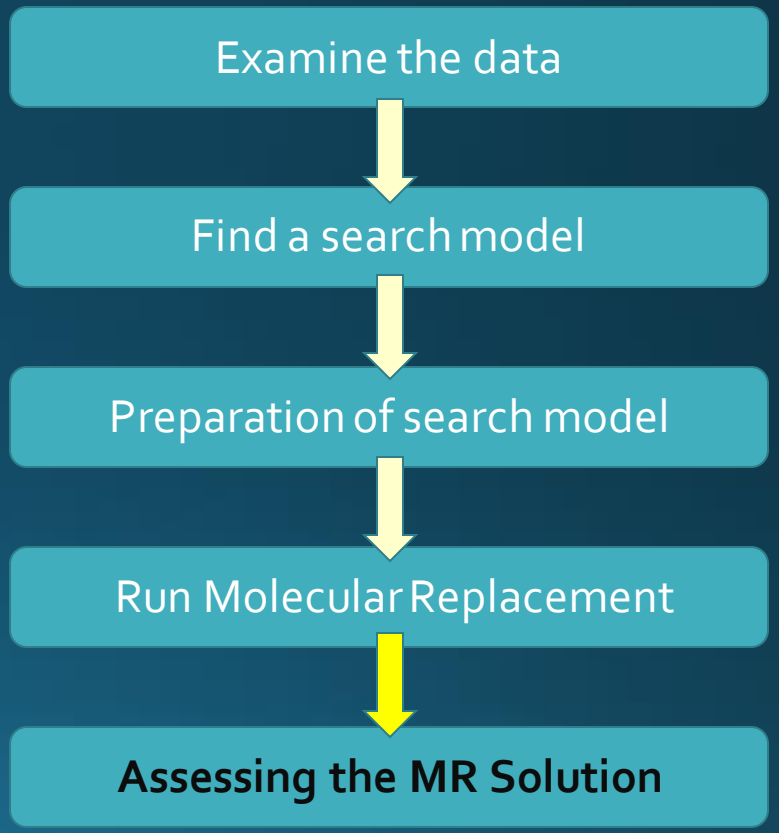
Step-by-Step MR: *Assessing the MR solution*

- Rough guide to MR program scoring
  - Phaser scores
    - LLG scores – has it increased by 6σ 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



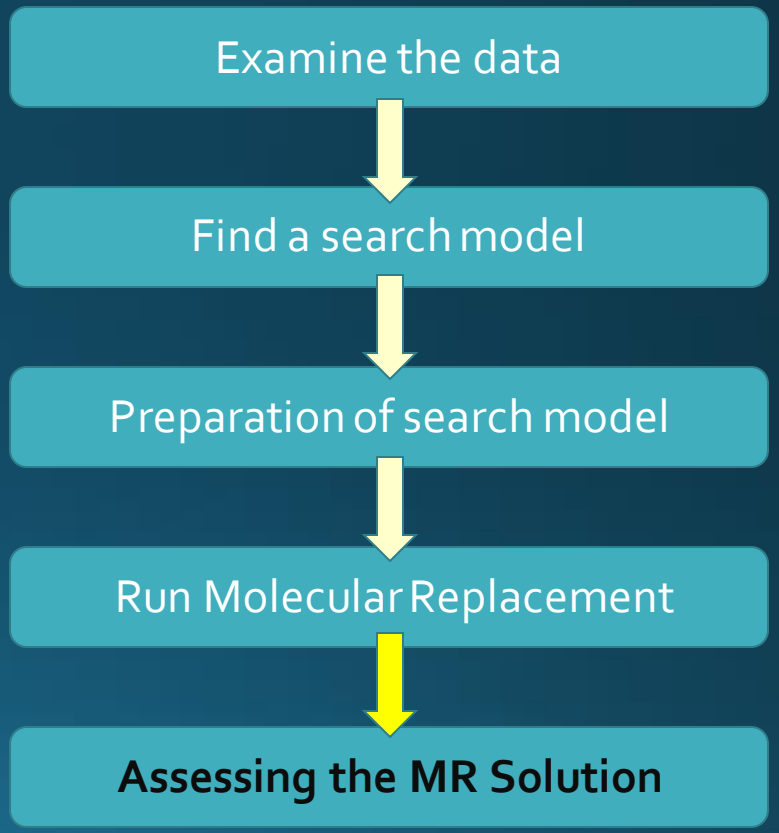
Step-by-Step MR: *Assessing the MR solution*

- 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

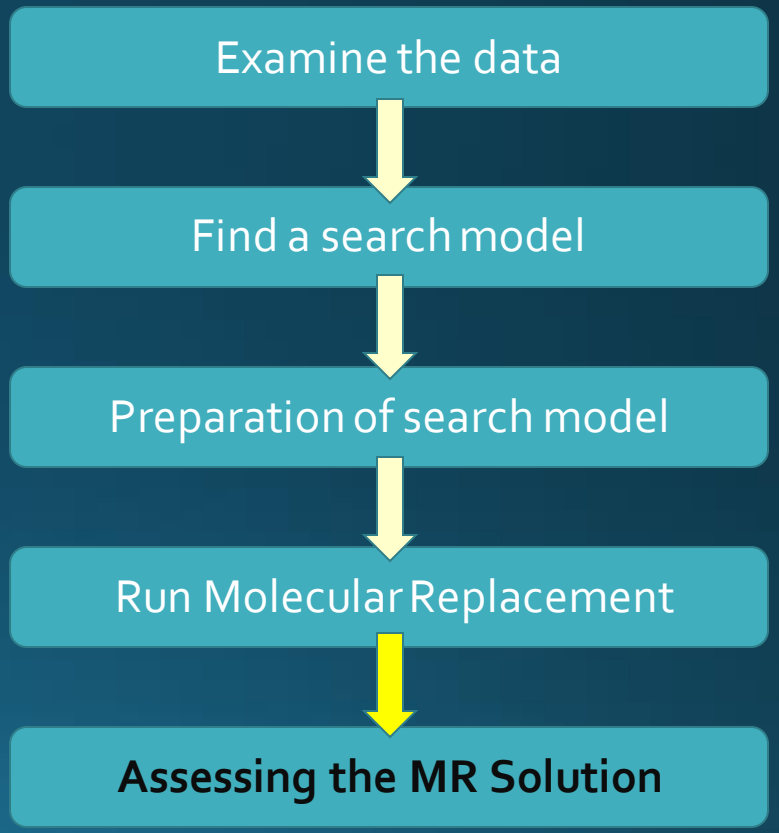


CCP4

MrBUMP

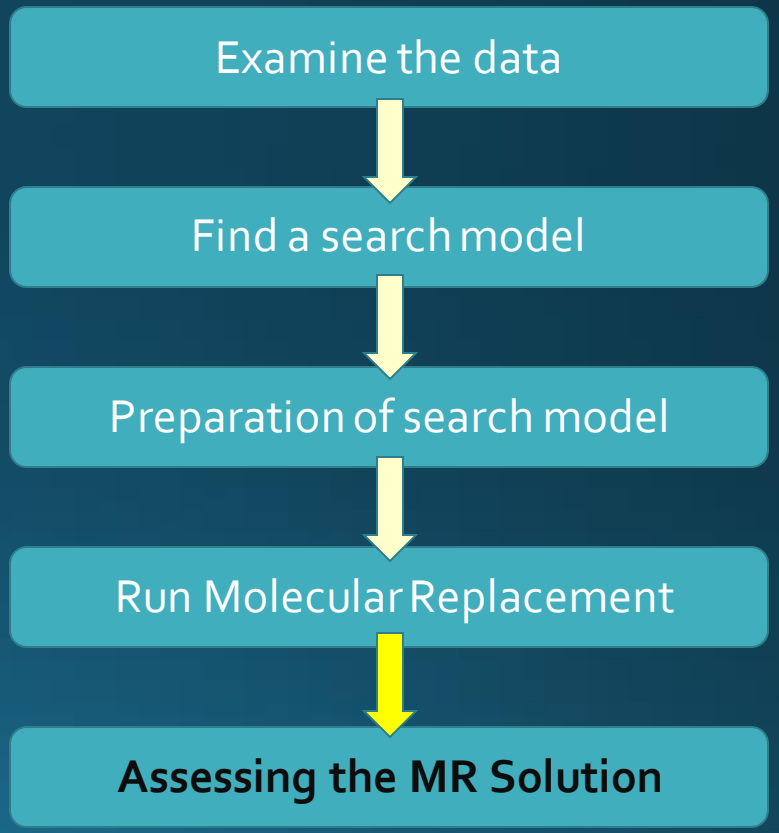
Step-by-Step MR: *Assessing the MR solution*

- 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

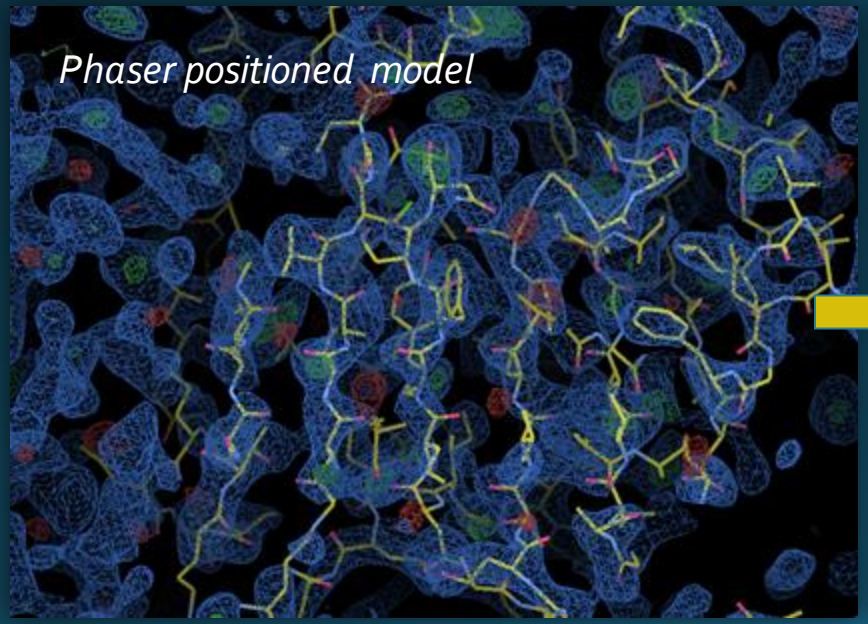


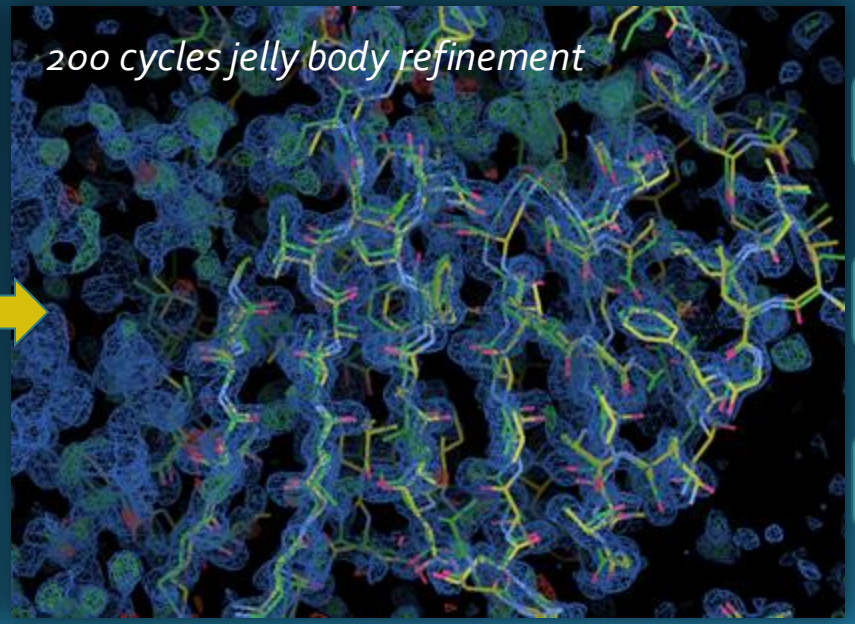
CCP4

MrBUMP

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





Examine the data

Find a search model

Preparation of search model

Run Molecular Replacement

Assessing the MR Solution

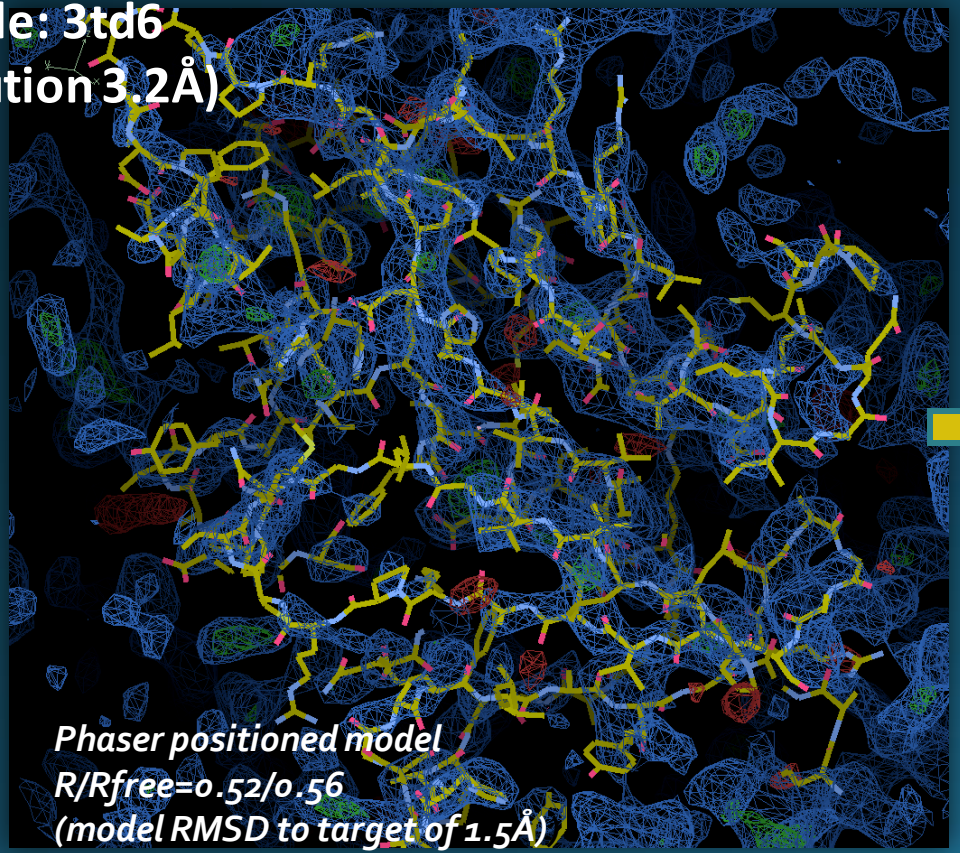
CCP4

MrBUMP

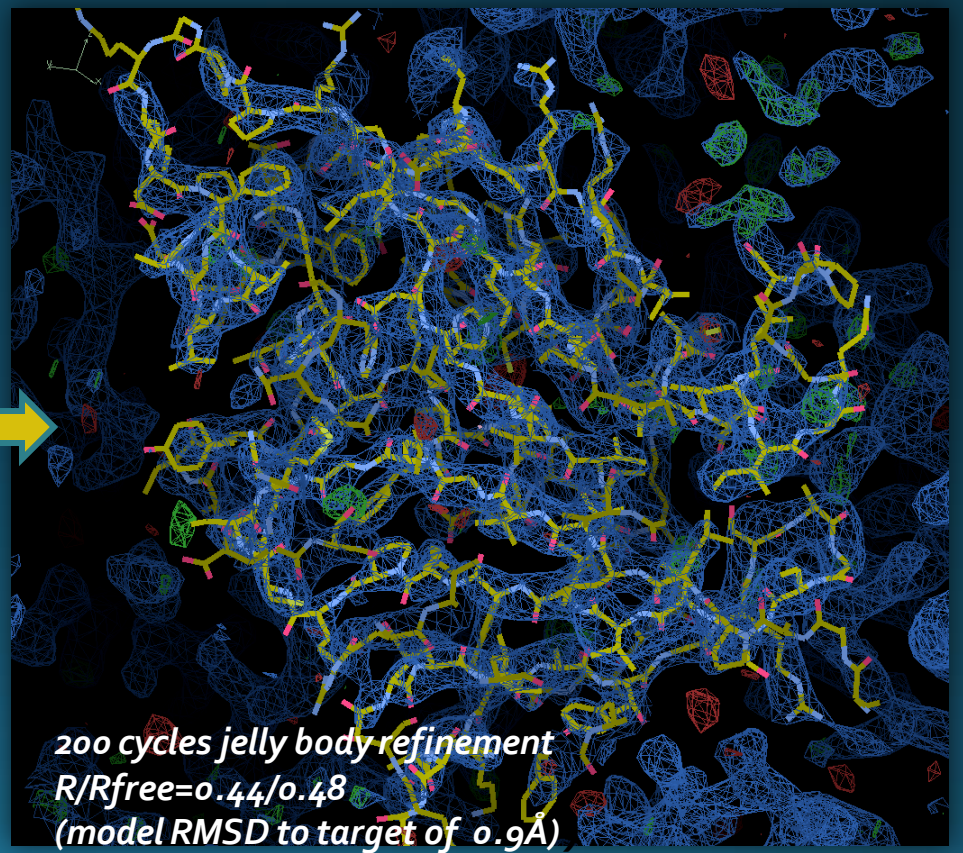
Step-by-Step MR: *Assessing the MR solution*

- Refinement
  - Use 200 cycles of jelly body refinement option in Refmac post MR

Example: 3td6
(resolution 3.2Å)



Phaser positioned model
R/Rfree=0.52/0.56
(model RMSD to target of 1.5Å)



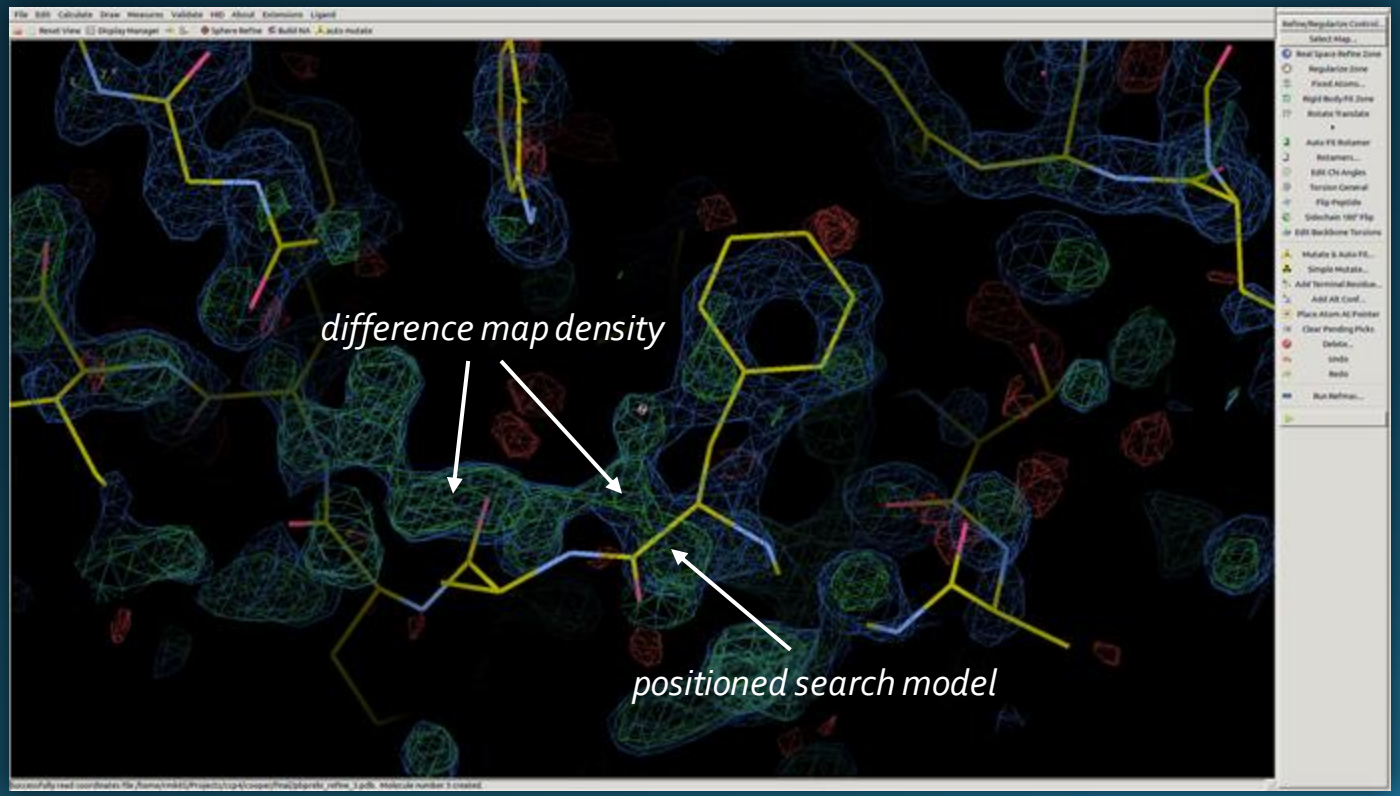
200 cycles jelly body refinement
R/Rfree=0.44/0.48
(model RMSD to target of 0.9Å)

CCP4

MrBUMP

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

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



Examine the data

↓

Find a search model

↓

Preparation of search model

↓

Run Molecular Replacement

↓

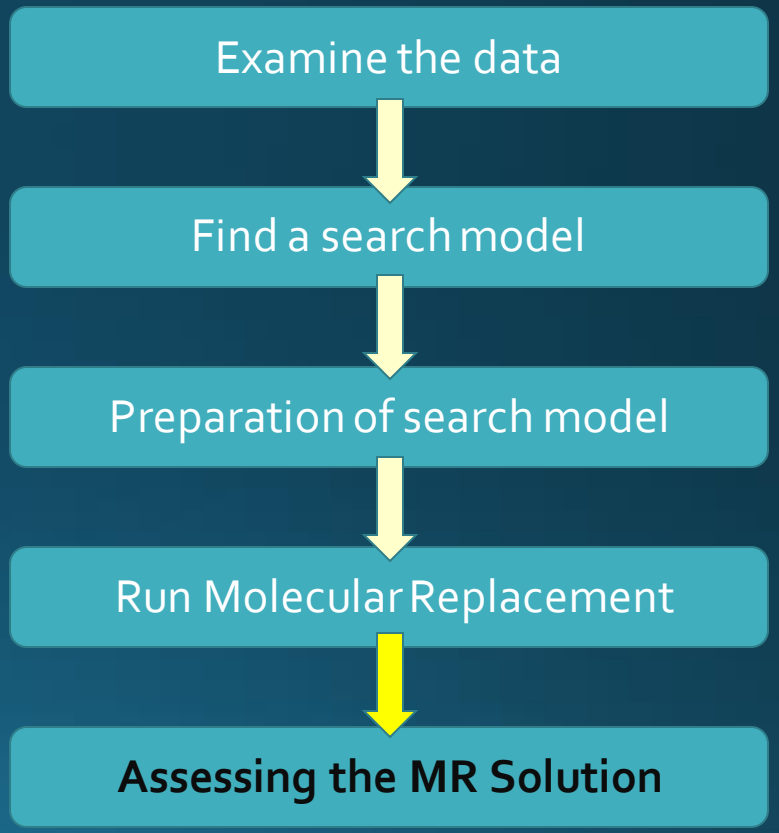
Assessing the MR Solution

CCP4

MrBUMP

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

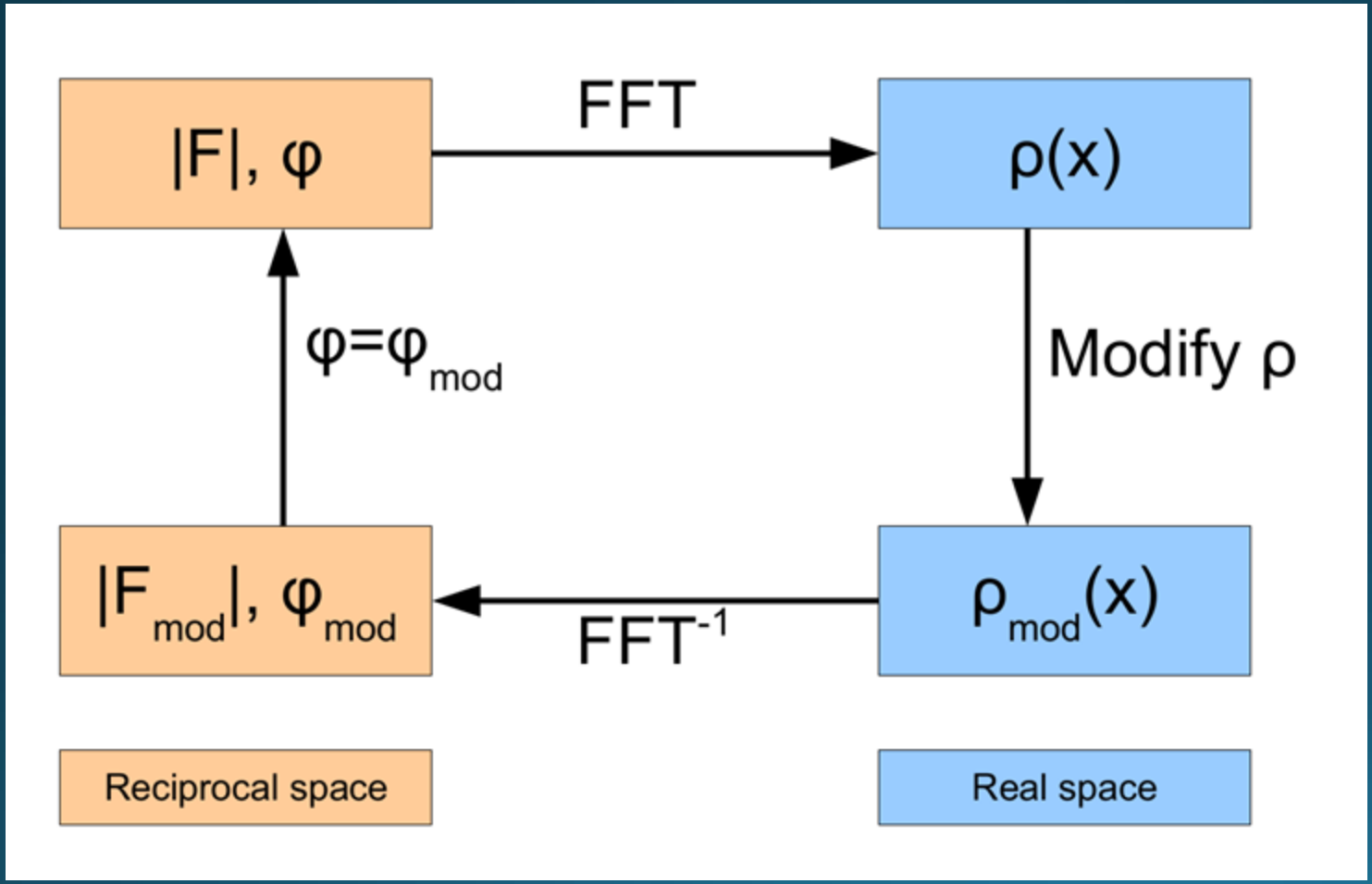
- Density Modification & C-alpha tracing



CCP4

MrBUMP

Density Modification



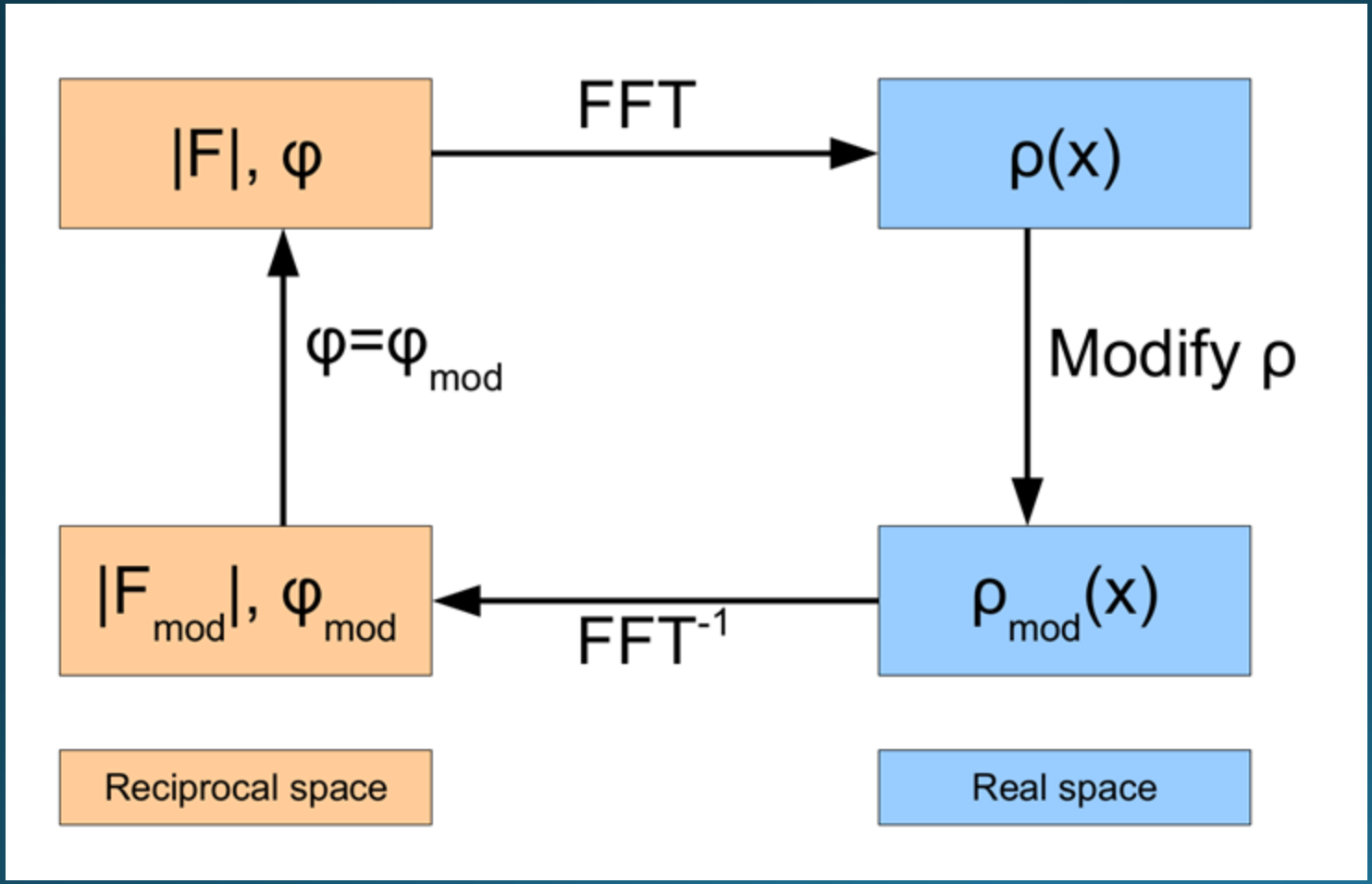
Main techniques:

1. Solvent flattening
2. Histogram matching
3. NCS averaging
4. C-alpha tracing

(slide from Kevin Cowtan)

CCP4

Density Modification



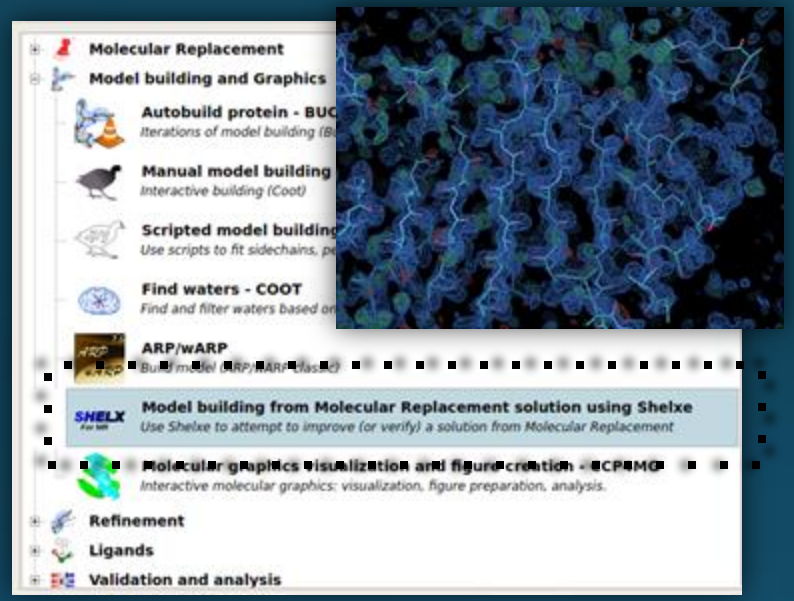
CCP4 Applications:

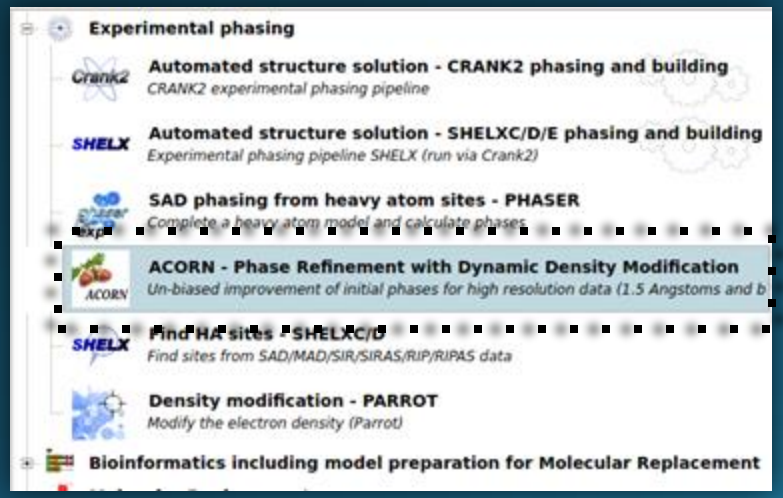
- Parrot
- SHELXE
- ACORN
- Pirate
- Solomon
- DM

(slide from Kevin Cowtan)

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

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





- ACORN (*resolution better than 1.7Å*)

Examine the data

↓

Find a search model

↓

Preparation of search model

↓

Run Molecular Replacement

↓

Assessing the MR Solution

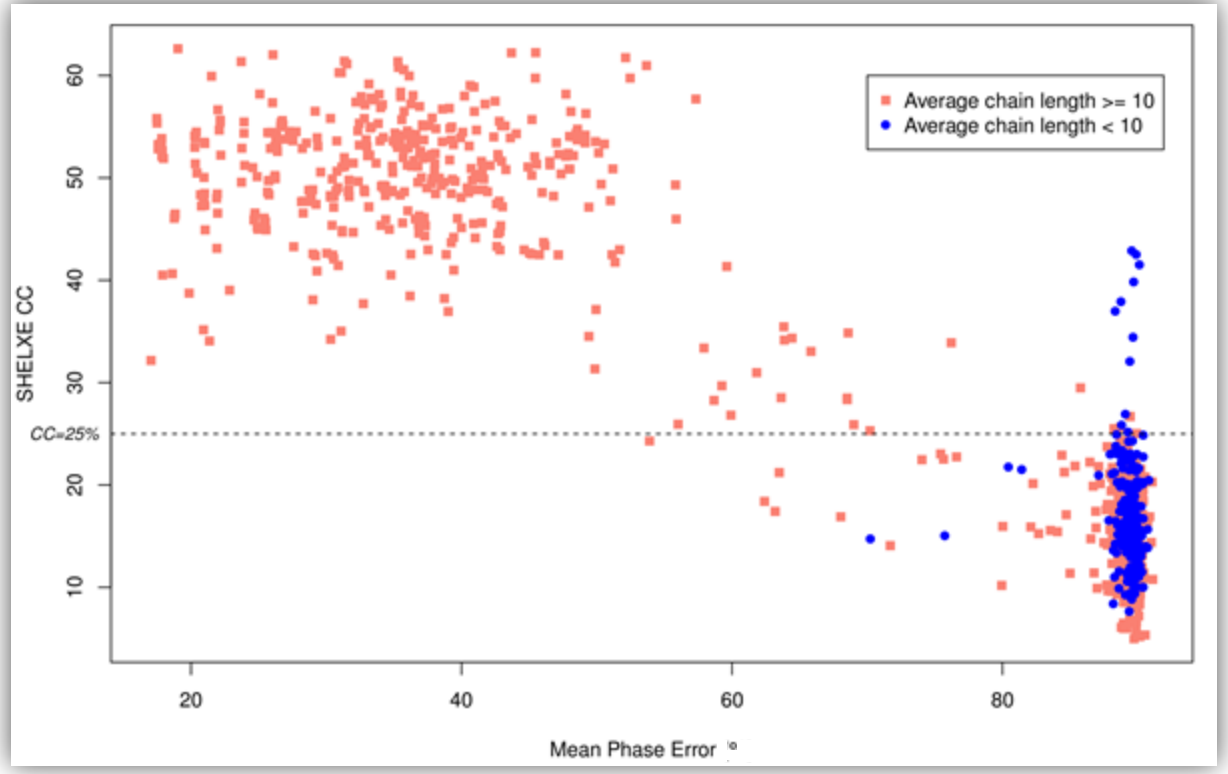
CCP4

MrBUMP

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

- SHELXE scoring: CC >= 25 %
  (average c-alpha fragment length >= 10)

SHELXE scoring tested for 1000 separate MR solutions



*(2015) Acta Crystallogr D Struct Biol **71**: 338-343*

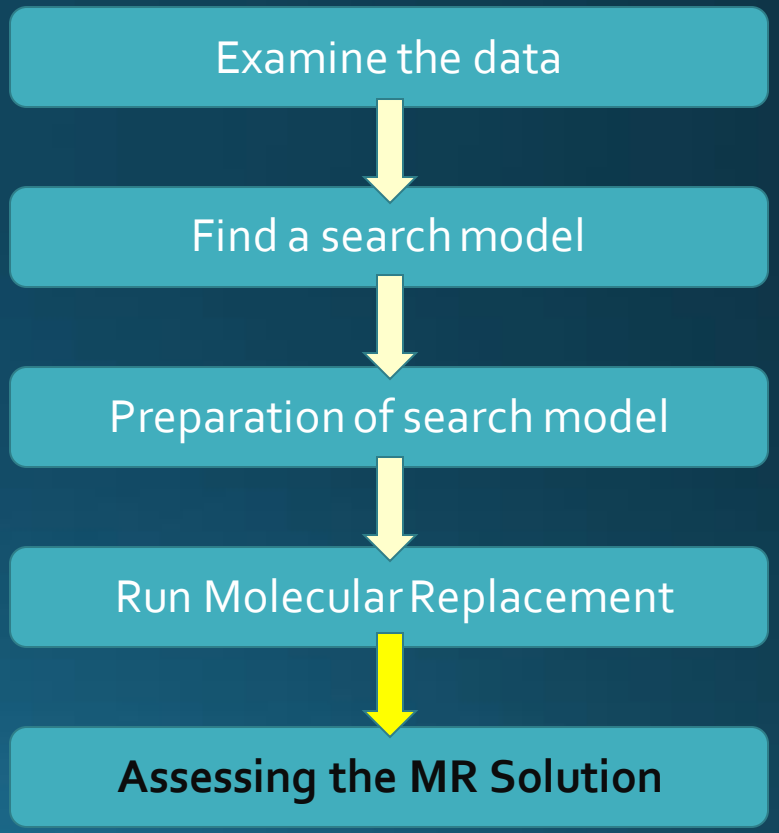
- Examine the data
- Find a search model
- Preparation of search model
- Run Molecular Replacement
- **Assessing the MR Solution**

CCP4

MrBUMP

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

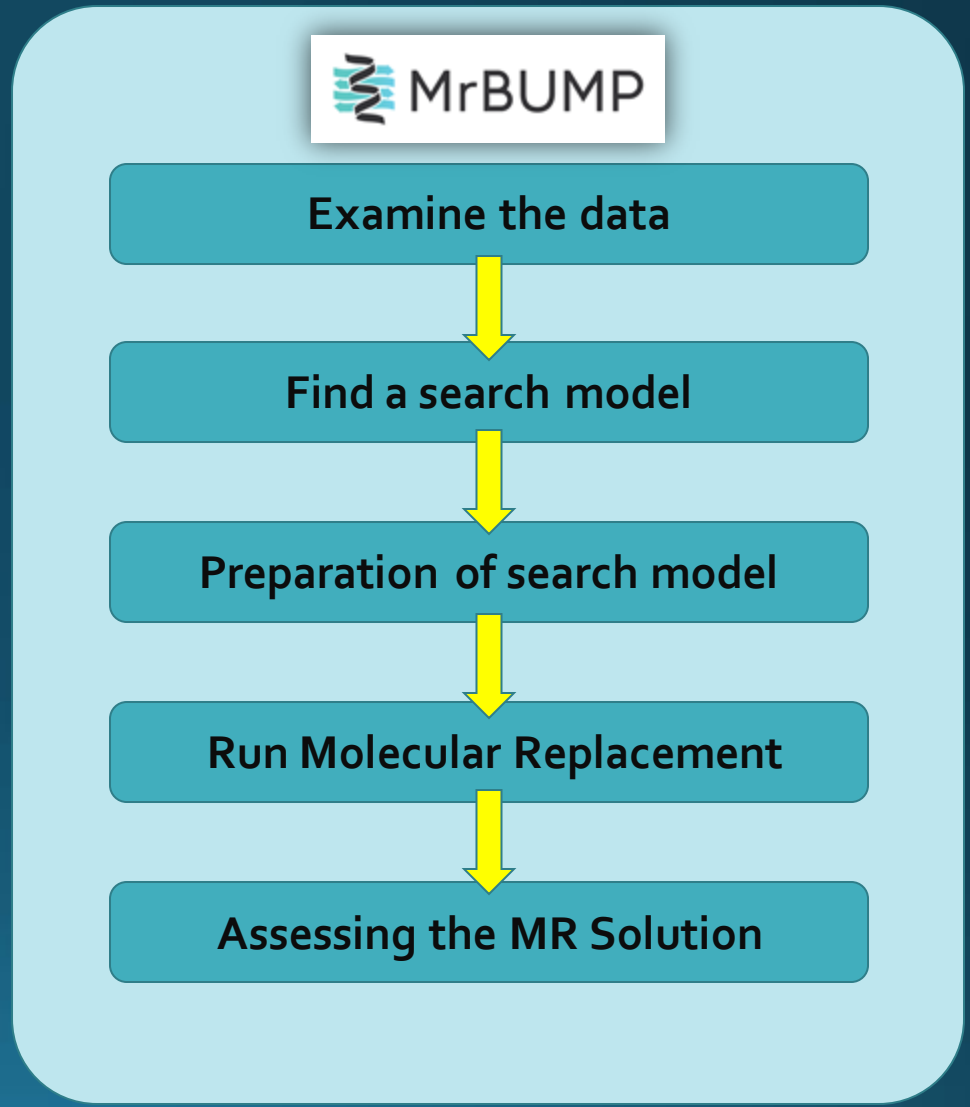
- Several automation pipelines for MR in CCP4:
  - ***MrBUMP*** – model search, preparation, MR and refinement
  - ***BALBES*** – model search in custom version of PDB database
  - ***MoRDa*** – similar to BALBES

CCP4

MrBUMP

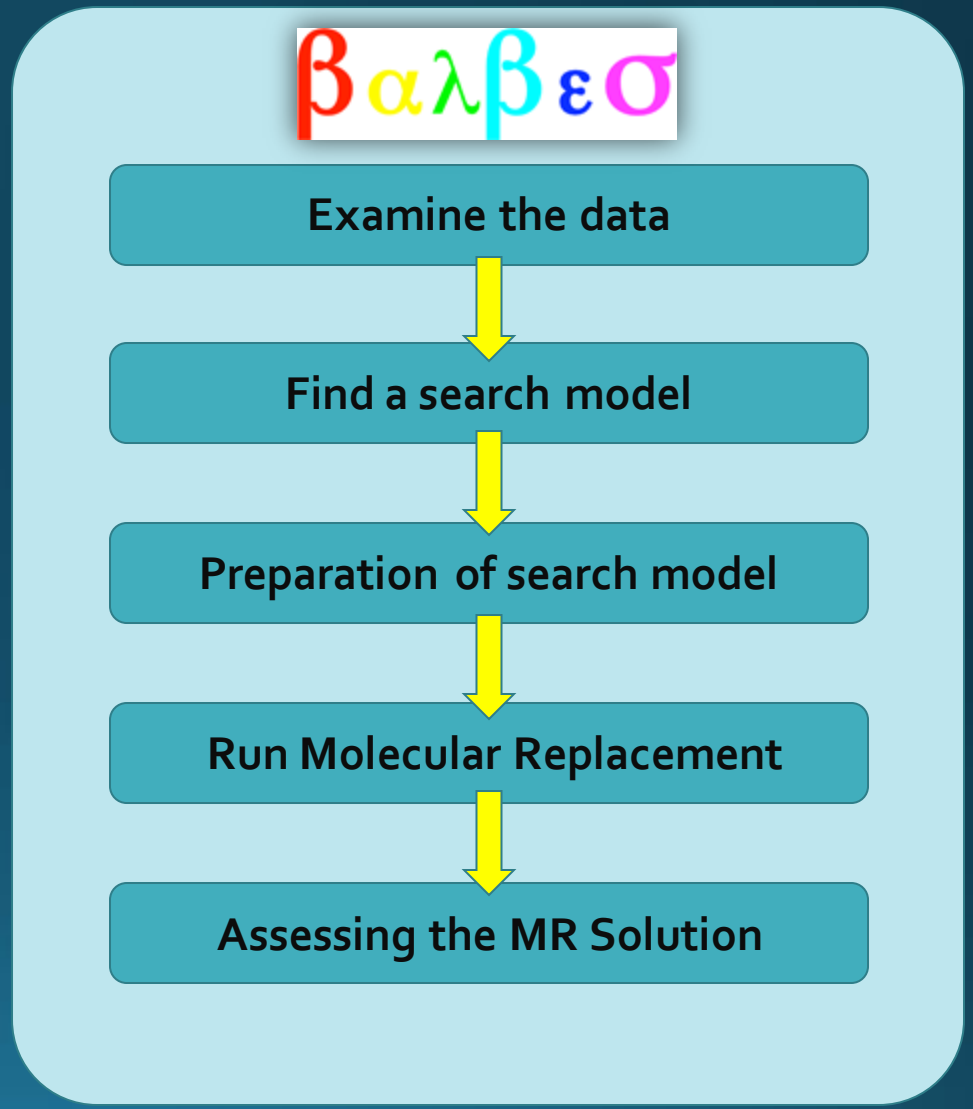
MrBUMP

- 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, CCP4 Cloud & CCP4 Online
  - CCP4mg – model generation step with graphical view



BALBES/MoRDa

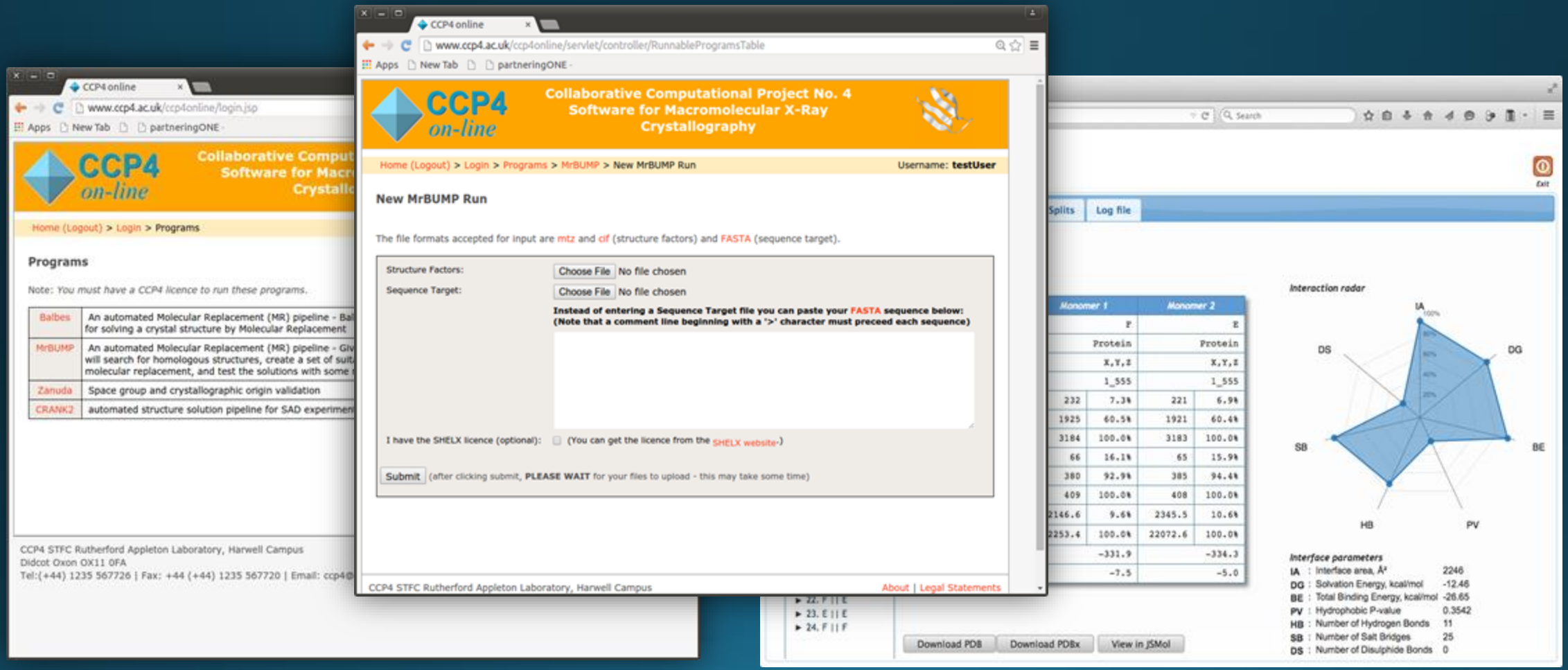
- Covers all stages from data examination through to refinement
- Custom database:
  1. Non-redundant version of PDB
  2. Basic entry is a domain with definitions for constructing full chains, multimers and ensembles
- Other features:
  - Search all spacegroups
  - Uses Molrep for MR and Refmac for refinement
  - Available in CCP4i2, CCP4 Cloud & on CCP4 Online



CCP4

MrBUMP

CCP4 Online: *MrBUMP, BALBES, MoRDa & more*



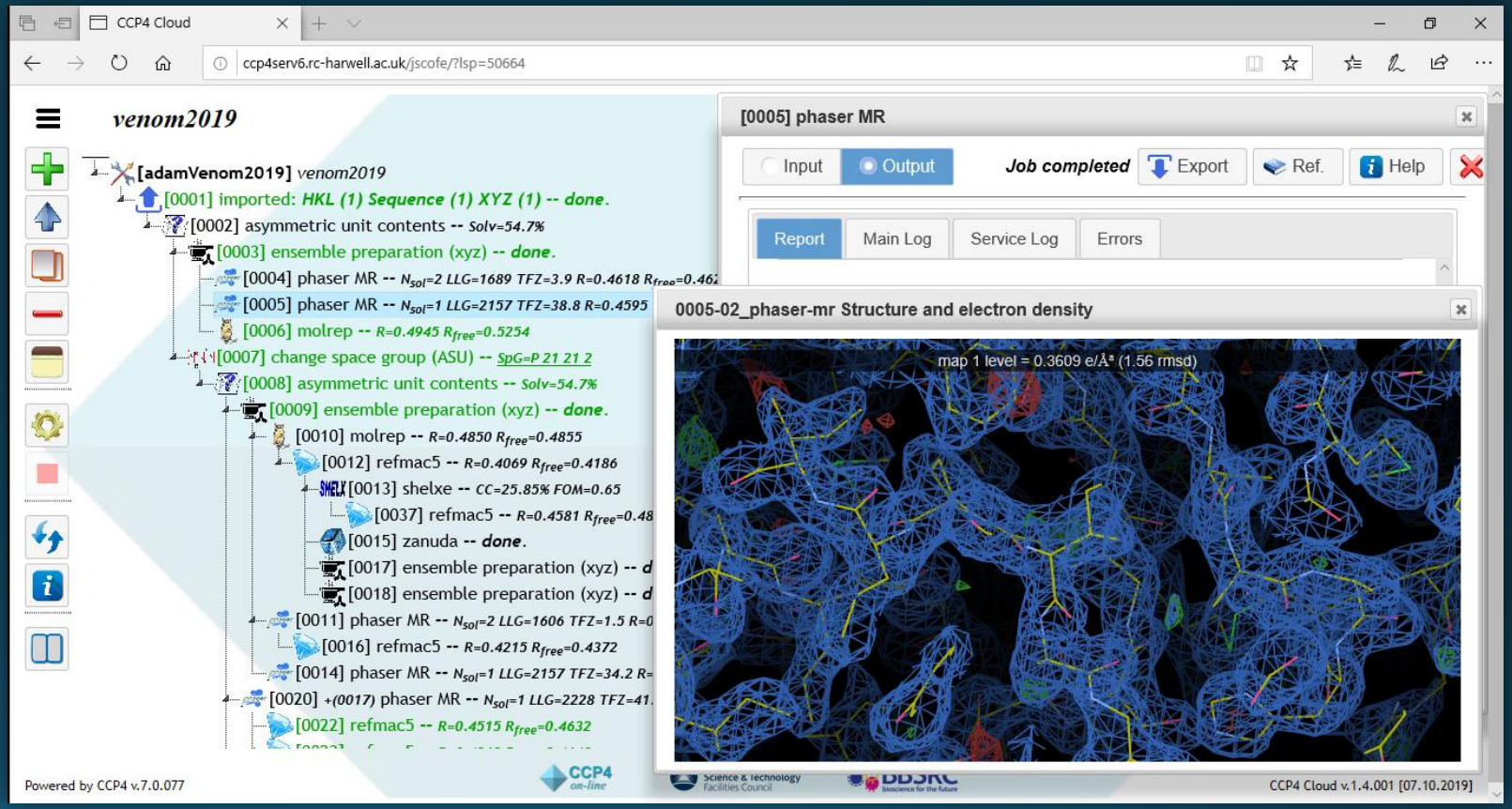
http://www.ccp4.ac.uk/ccp4online

CCP4

MrBUMP

MR on the CCP4 Cloud

- Excellent interface for tackling difficult MR problems
- Explore broad range of search models, preparation protocols and refinements
- Exploit vast computational resources
-



https://cloud.ccp4.ac.uk

CCP4

MrBUMP

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

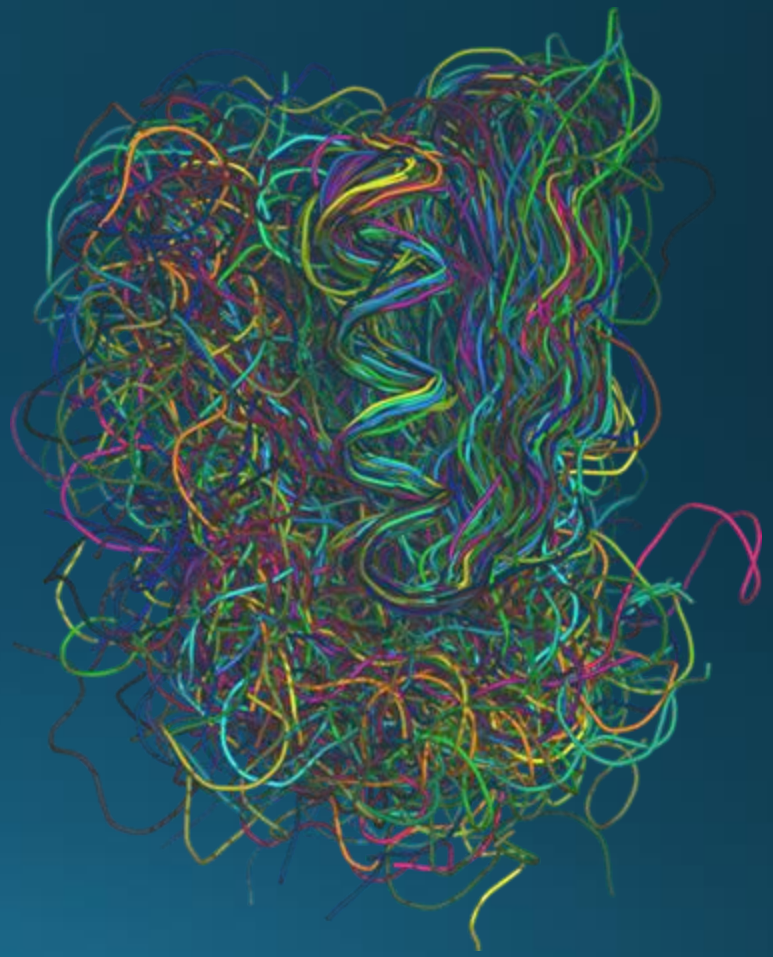
- Try different search models or different preparation methods for the homologues
- Try more distant homologues
  - Sequence similarity does not always imply structural similarity e.g. S100 family of proteins
  - Distant homologues can be structurally similar e.g. globular enzymes

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
  - Sulpher 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

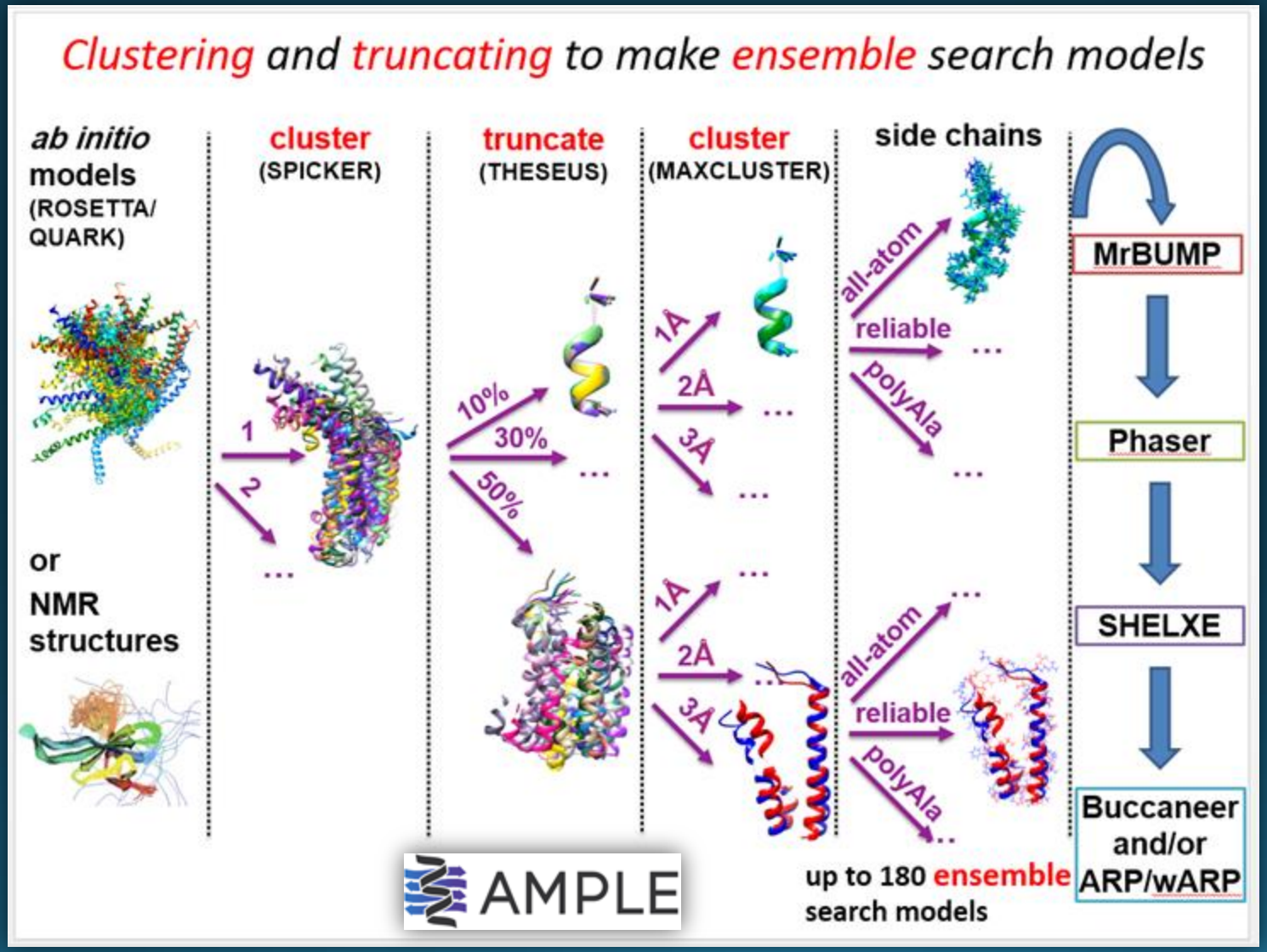


CCP4

MrBUMP

What to do if Molecular Replacement doesn't work

AMPLE Pipeline



(image from Jens Thomas)

CCP4

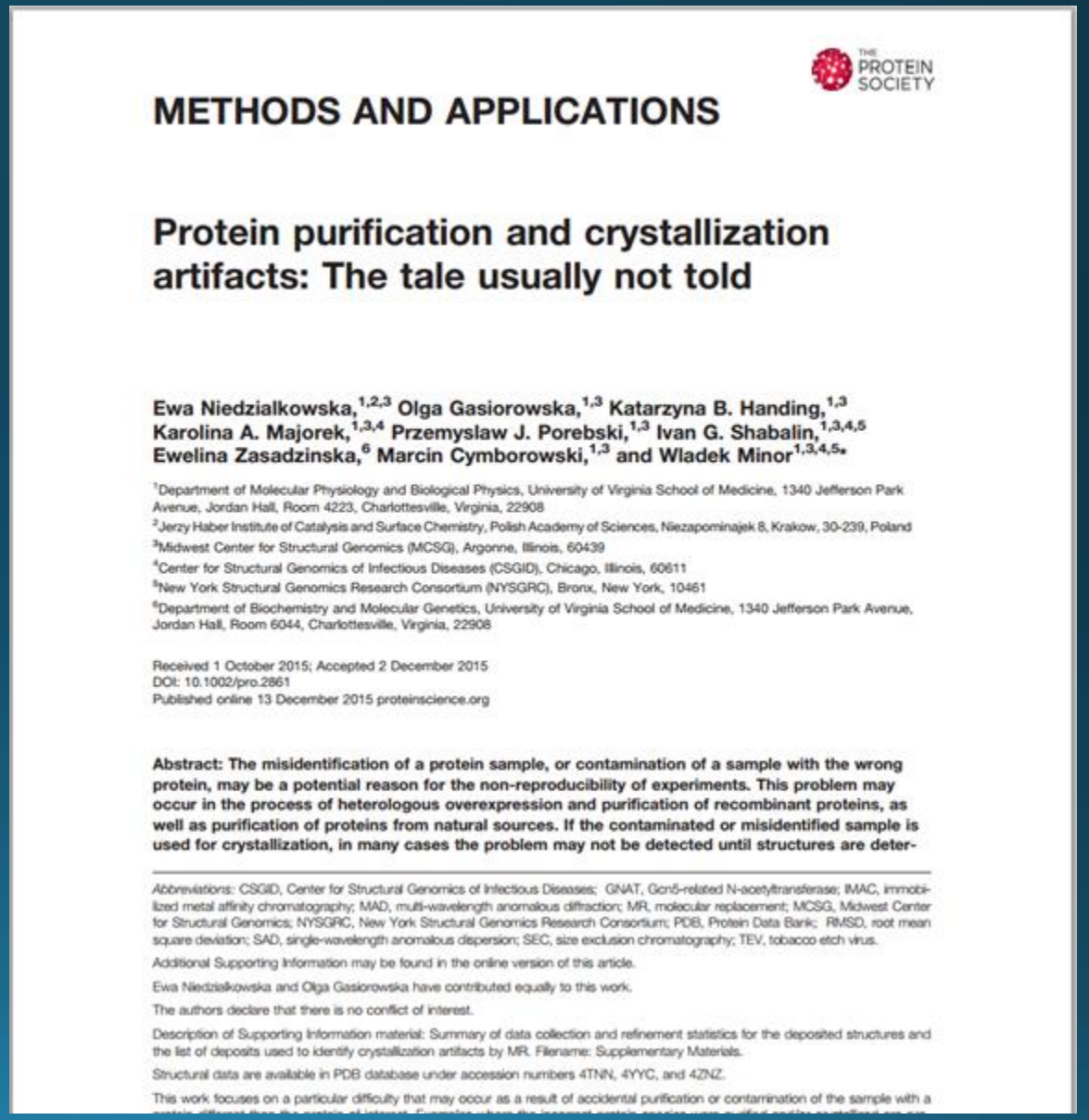
MrBUMP

What to do if Molecular Replacement doesn't work

- ***ARCIMBOLDO*** *(Isabel will present more)*
- Lite
  - Makes use of simple secondary structure elements such as helices in MR
  - Attempts to position fragments and build up the rest of the c-alpha backbone using SHELXE
- Borges
  - Similar to Lite but draws on a library of common motifs from the PDB as search models e.g. Zinc fingers or common beta sheet fragments
- Shredder
  - Cuts distant homologues into fragments and will use them as search models in ARCIMBOLDO
- **http://chango.ibmb.csic.es/ARCIMBOLDO/**

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 Contaminer



THE PROTEIN SOCIETY

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]*

[1]Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, 1340 Jefferson Park Avenue, Jordan Hall, Room 4223, Charlottesville, Virginia, 22908
[2]Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Niezapominajek 8, Krakow, 30-239, Poland
[3]Midwest Center for Structural Genomics (MCSG), Argonne, Illinois, 60439
[4]Center for Structural Genomics of Infectious Diseases (CSGID), Chicago, Illinois, 60611
[5]New York Structural Genomics Research Consortium (NYSGRC), Bronx, New York, 10461
[6]Department 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.

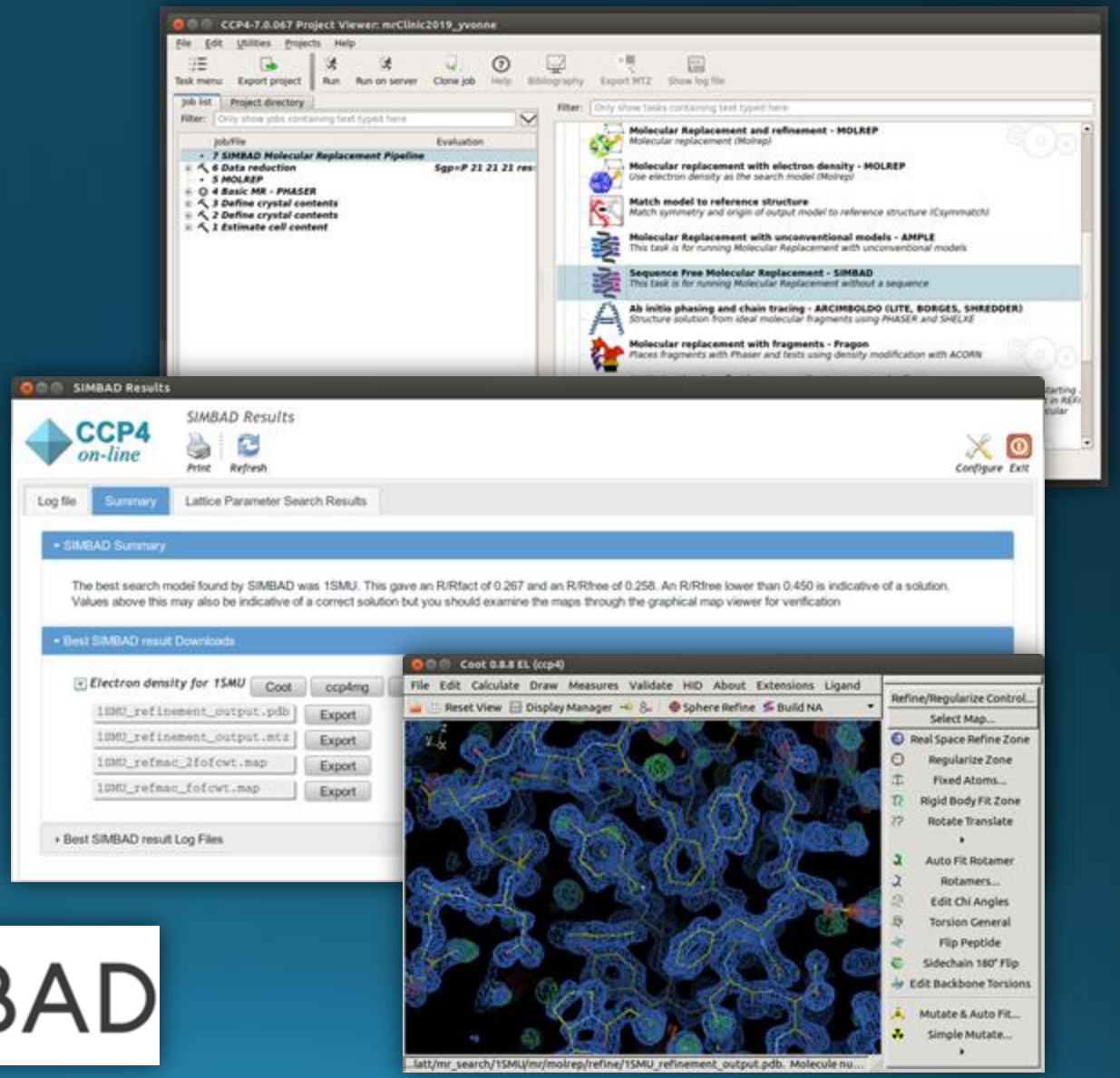
This work focuses on a particular difficulty that may occur as a result of accidental purification or contamination of the sample with a

CCP4

What to do if Molecular Replacement doesn't work

- **SIMBAD**: *Sequence-less MR*
  1. nearest cell (lattice)
  2. contaminants
  3. entire non-redundant domain database (90000 models)

- Accessible from *Molecular Replacement* menu
  - CCP4 Online
  - CCP4i2



SIMBAD

CCP4

Acknowledgements

- Alexey Vagin & Andrey Lebedev – ***MoRDa and Molrep***
- Daniel Rigden, Jens Thomas, Felix Simkovic & Adam Simpkin - ***University of Liverpool***
- Stuart McNicholas, Keith Wilson & Christian Roth - ***University of York***
- Randy Read, Airlie McCoy & Gabor Bunkozci - ***Phaser***
- Martyn Winn, Charles Ballard, Eugene Krissinel, Ville Uski & Kyle Stevenson - ***CCP4***
- Marcin Wojdyr - ***Global Phasing***
- Isabel Uson, Andrea Thorn, Tim Gruene & George Sheldrick – ***SHELXE & ARCIMBOLDO***
- Garib Murshudov, Rob Nicholls & Fei Long – ***Refmac and BALBES***
- Kevin Cowtan & Paul Bond – ***Buccaneer & DM***
- Victor Lamzin & Grzegorz Chojnowski - ***ARP/wARP***
- Thanks to authors of all other underlying programs