



UNIVERSITY OF
LIVERPOOL

Ronan Keegan CCP₄ Group

Molecular Replacement model selection, preparation and assessing the solution

CCP₄-DLS School in Macromolecular Crystallography - 2021

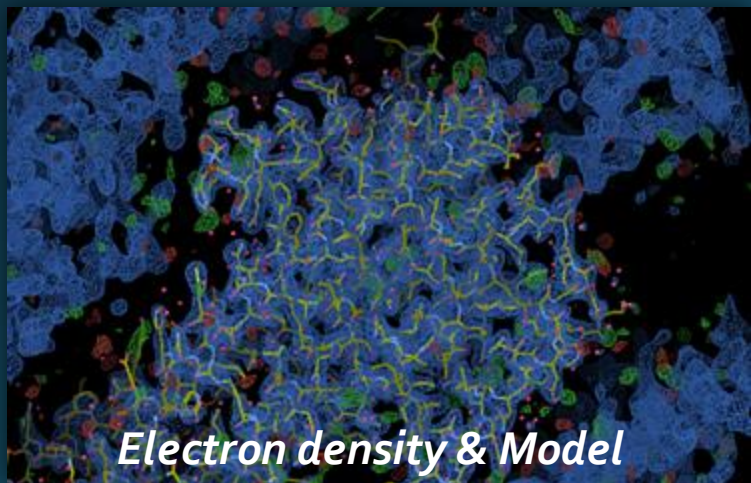
Introduction: *Basics of Molecular Replacement*

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

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)

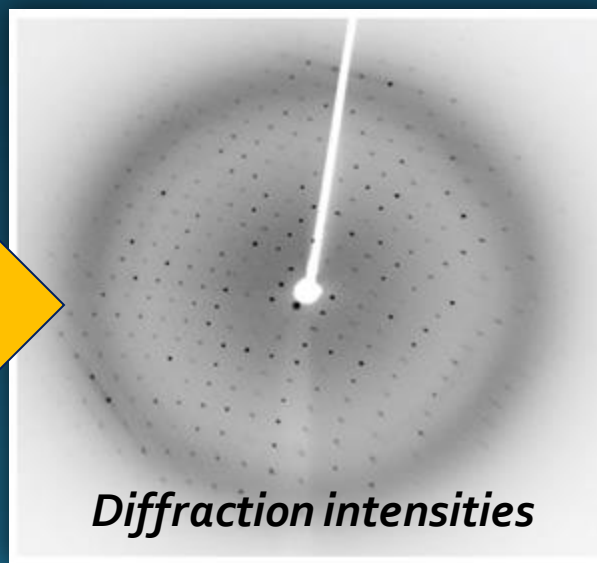
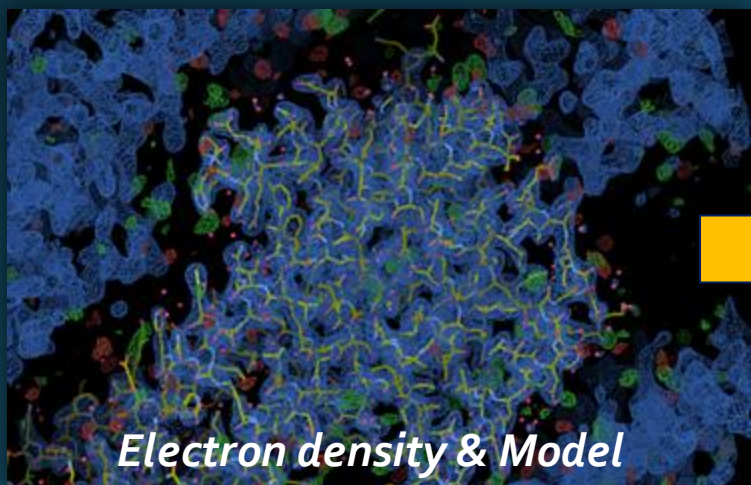


Introduction: *Basics of Molecular Replacement*

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

Electron density
at (x,y,z)

Amplitudes



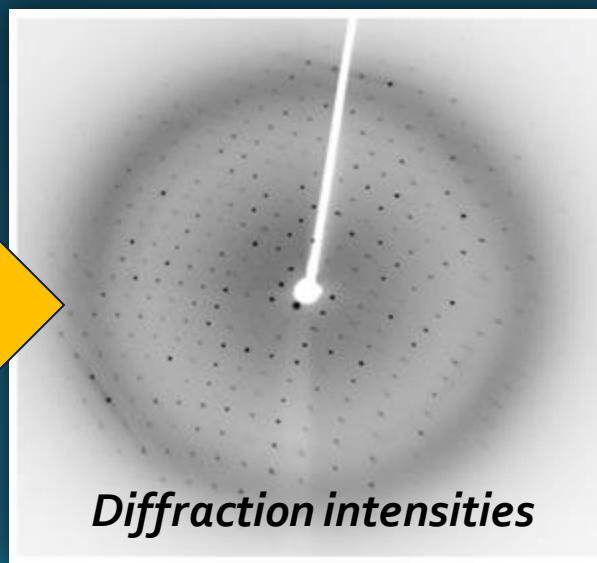
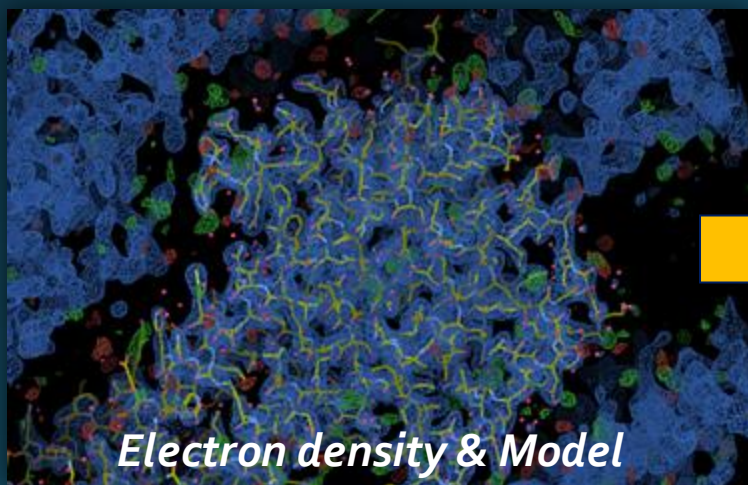
Introduction: *Basics of Molecular Replacement*

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

Electron density
at (x,y,z)

Amplitudes

Phases



?

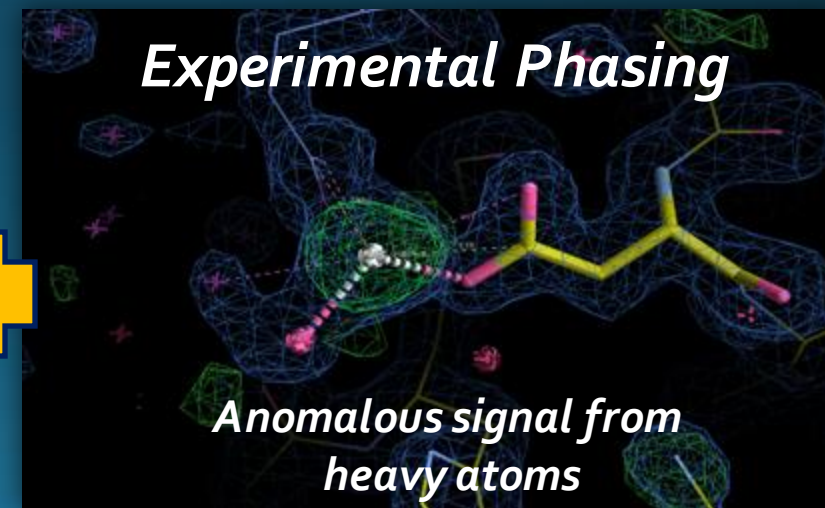
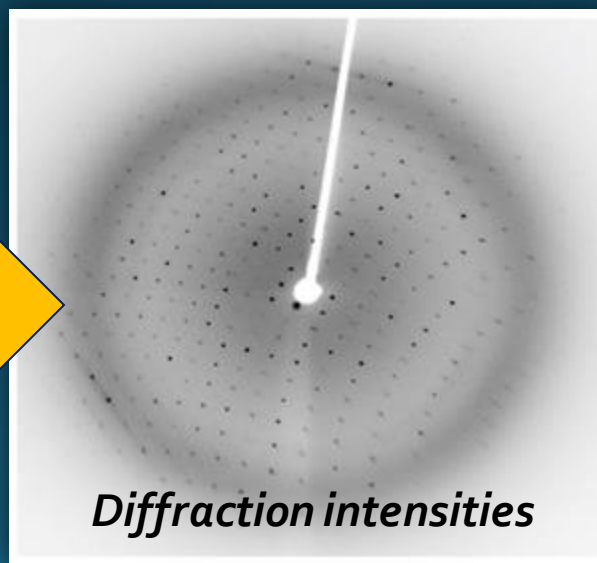
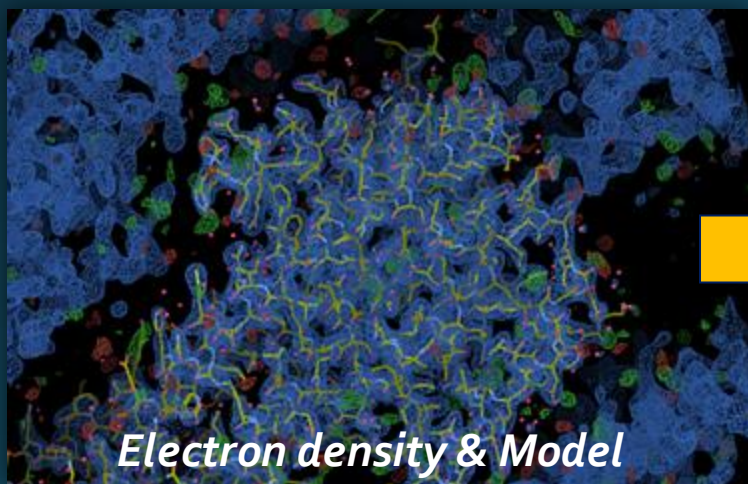
Introduction: *Basics of Molecular Replacement*

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

Electron density
at (x,y,z)

Amplitudes

Phases



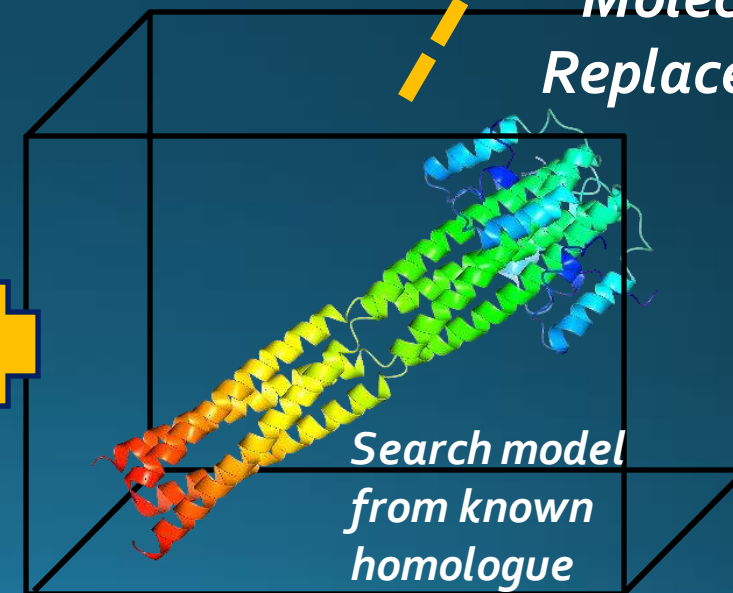
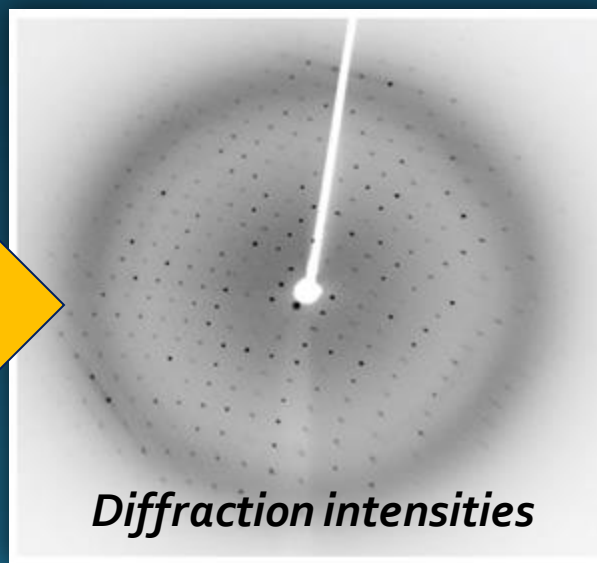
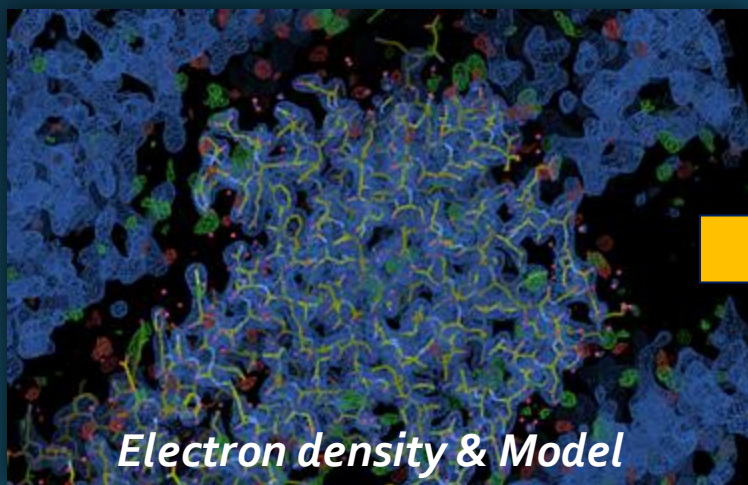
Introduction: *Basics of Molecular Replacement*

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

Electron density
at (x,y,z)

Amplitudes

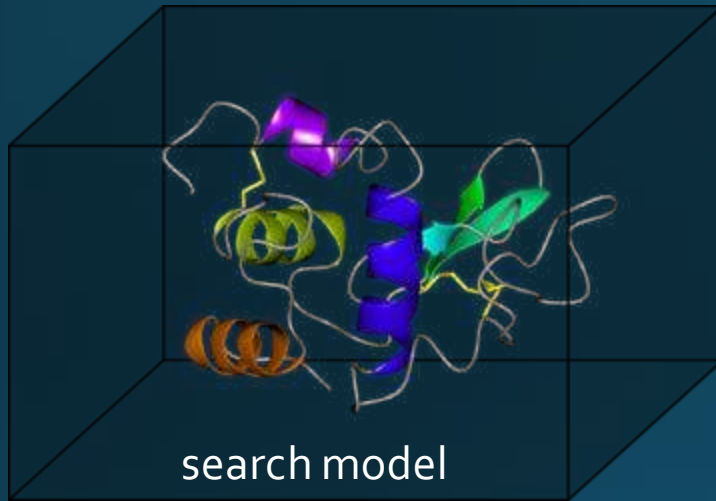
Molecular
Replacement



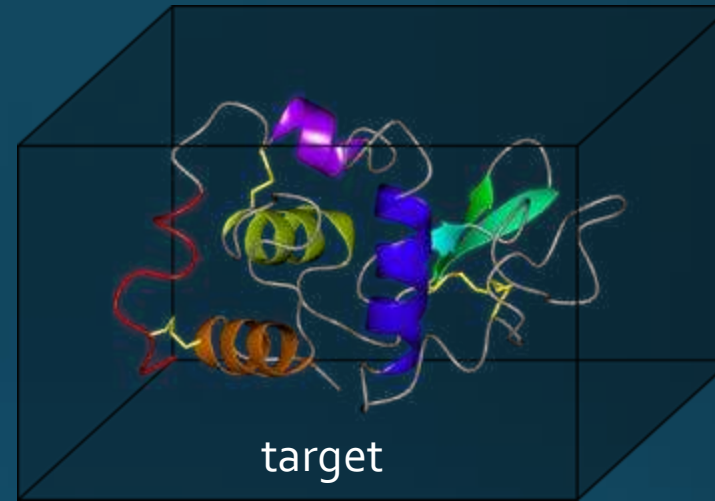
What is Molecular Replacement?

- We can take advantage of the conservation of structural form of proteins related through evolution
- A known structure of a protein related to our target (a homologue) could provide a good first approximation to the phases of the target when placed correctly in the target unit cell
- Molecular Replacement is the method of finding the best placement for a homologue such that it gives us this initial approximation to the target's phases

Introduction: *Basics of Molecular Replacement*



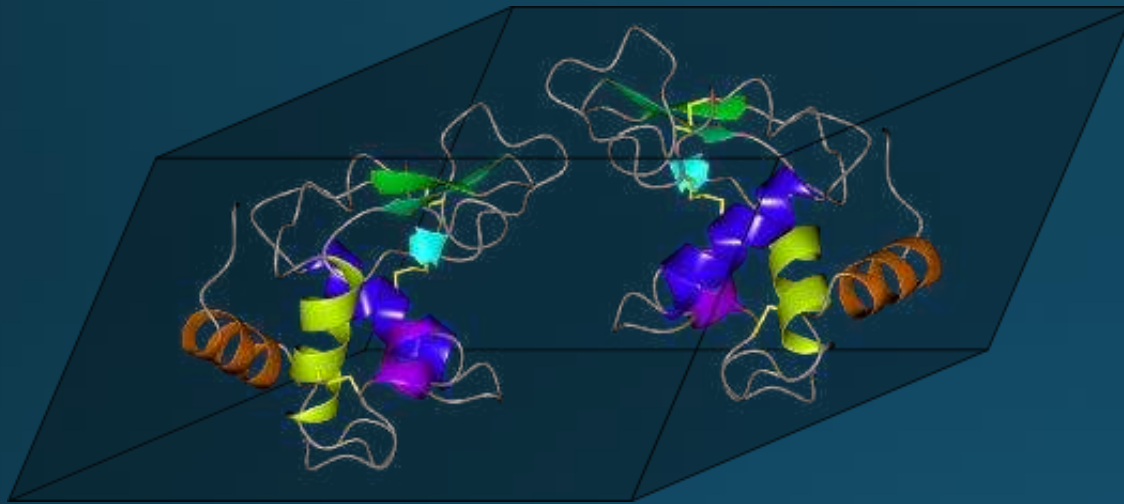
h	k	l	F	ϕ
0	0	1	12.6	123
0	0	2	2.1	12
0	0	3	69.9	287



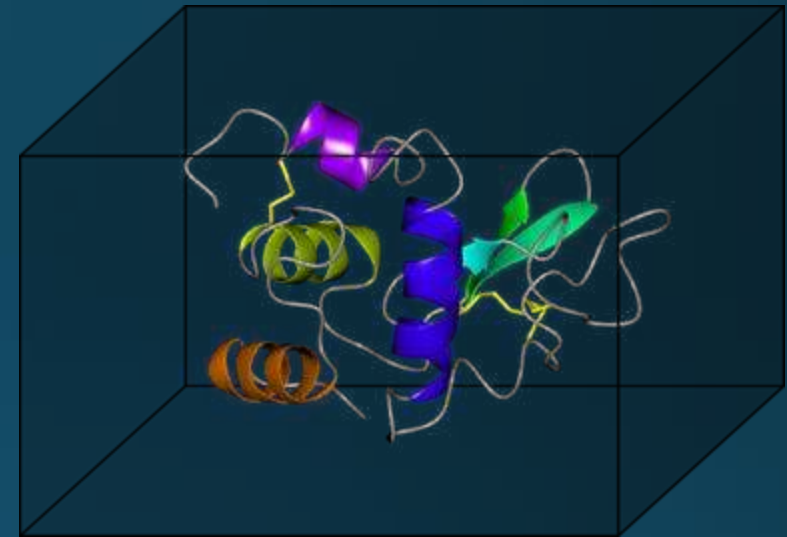
h	k	l	F	ϕ
0	0	1	10.4	113
0	0	2	3.5	18
0	0	3	57.2	265

(slide from Gabor Bunkoczi)

Introduction: *Basics of Molecular Replacement*



Search model crystal form

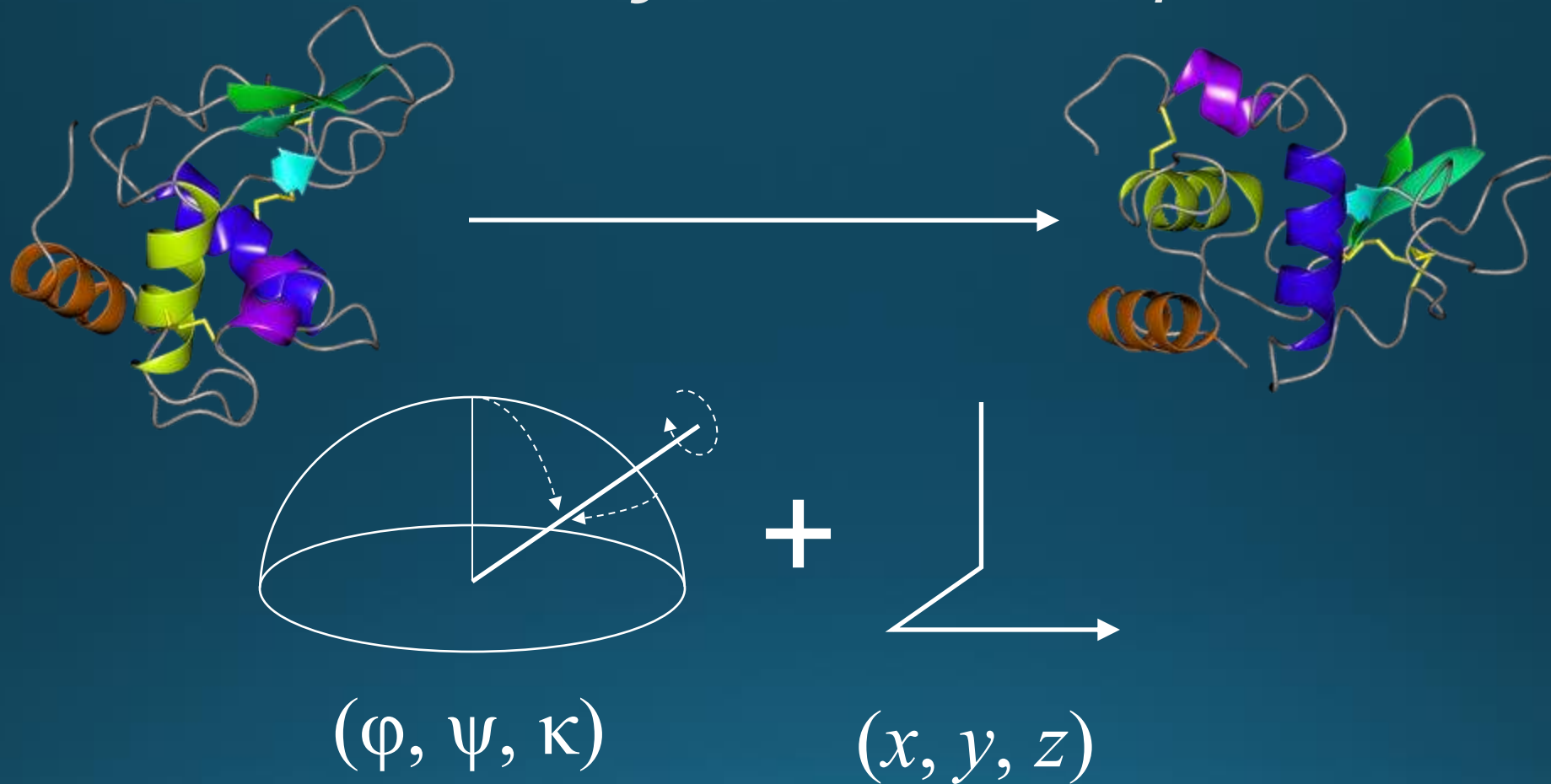


Target crystal form

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

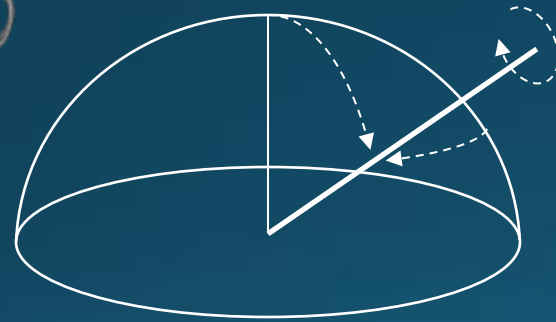
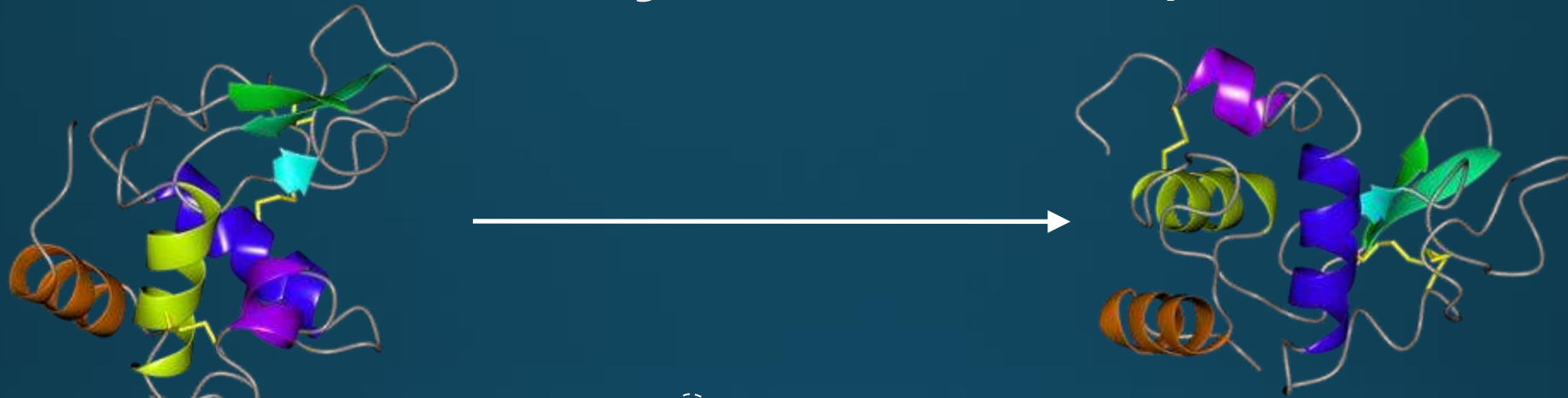
(slide from Gabor Bunkoczi)

Introduction: *Basics of Molecular Replacement*



(slide from Gabor Bunkoczi)

Introduction: *Basics of Molecular Replacement*



(φ, ψ, κ)

*Rotation
Function*

+



(x, y, z)

*Translation
Function*

(slide from Gabor Bunkoczi)

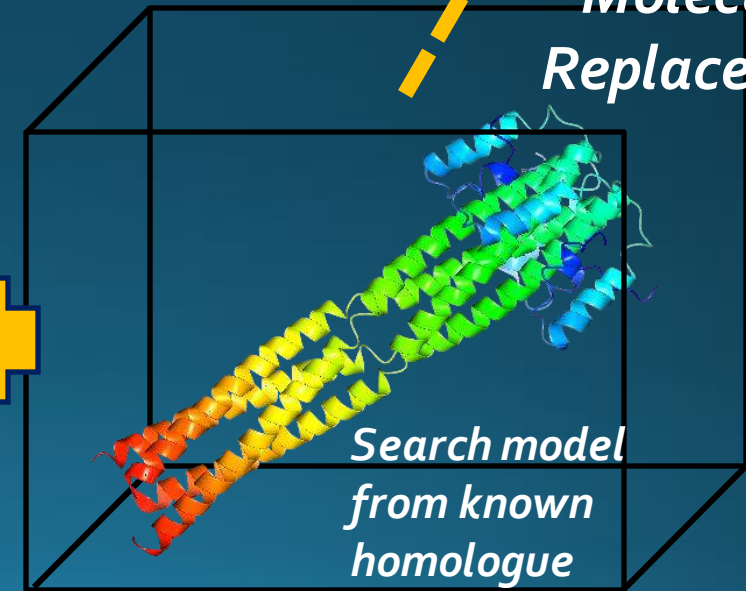
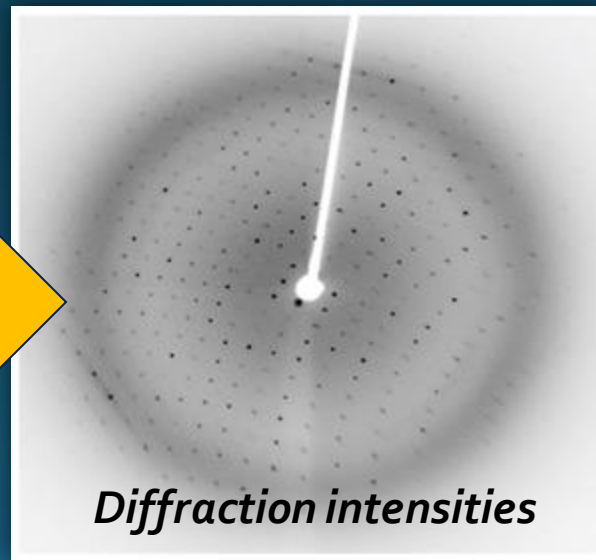
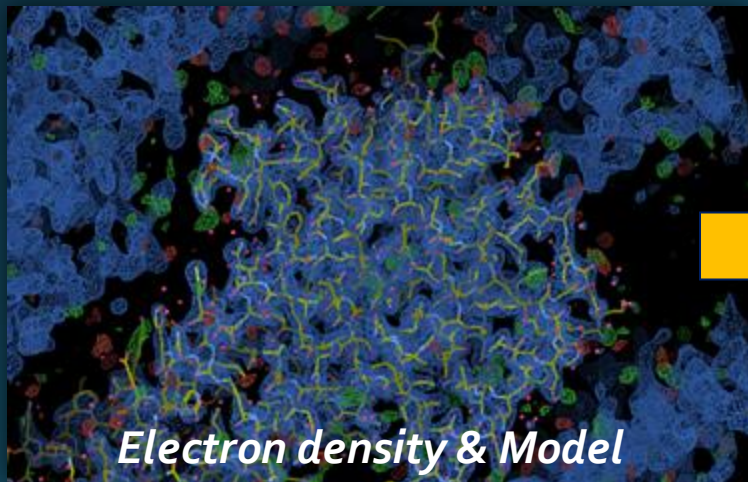
How do we find the correct positioning?

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

Electron density
at (x,y,z)

Amplitudes

Molecular
Replacement



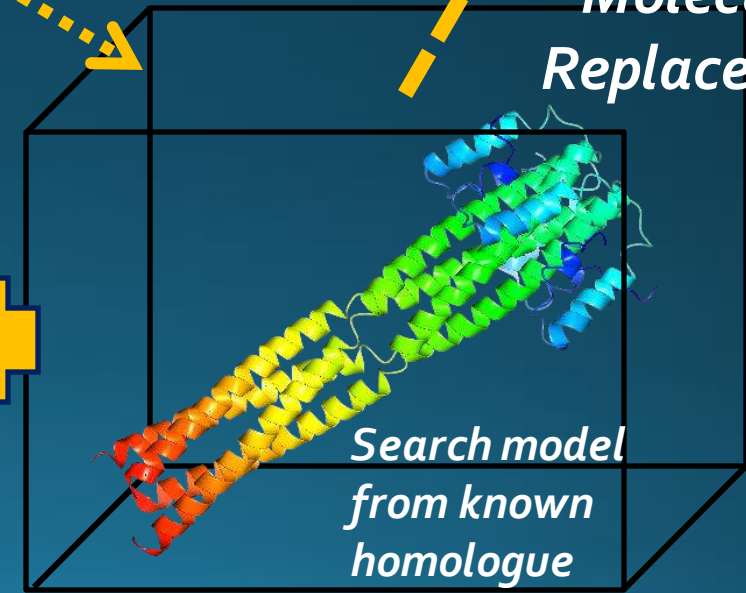
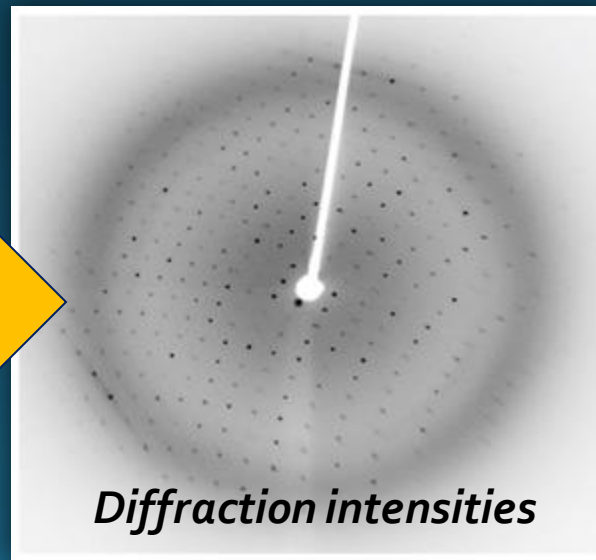
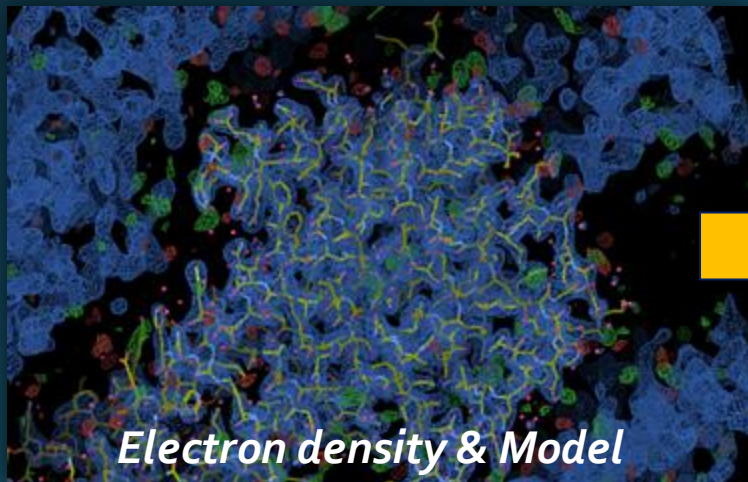
How do we find the correct positioning?

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

Electron density
at (x,y,z)

Amplitudes

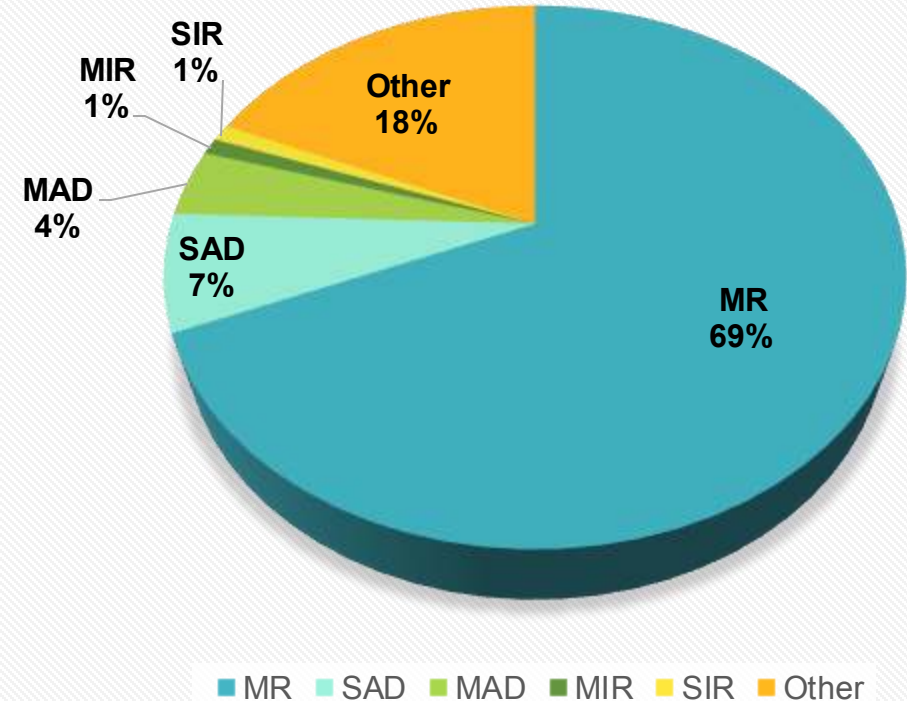
Molecular
Replacement



Molecular Replacement and the PDB

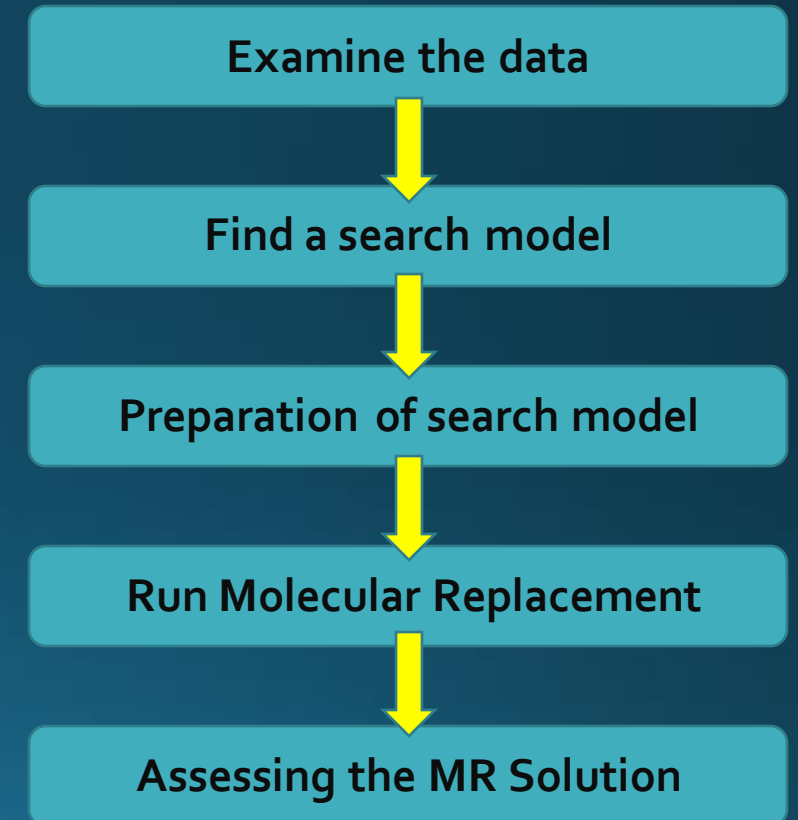
- Today, the vast majority of X-ray entries in the PDB have been phased using MR
- Increasing success rate of MR has been due to:
 - Improving methods in software e.g. maximum likelihood in *Phaser*
 - Increasing availability of suitable search models
 - New: AlphaFold/RoseTTAFold

PDB Structure Determination Method (X-ray)



Using Molecular Replacement in Structure Solution

- Successfully performing MR depends on:
 1. Understanding the limitations of your data
 2. Finding suitable search models
 3. Preparing models for MR
 4. Assessing the final solution from MR



Understanding the limitations of your data

Understanding the limitations of your data

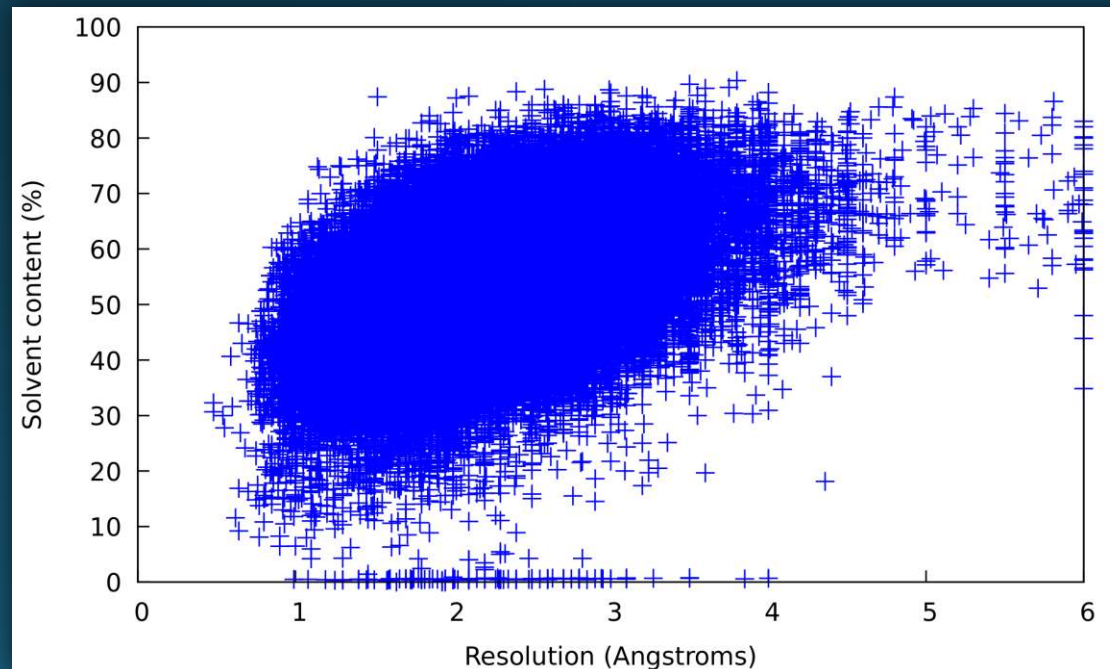
- Data issues can have an impact on how well MR will work

Examine the data

Understanding the limitations of your 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

Examine the data

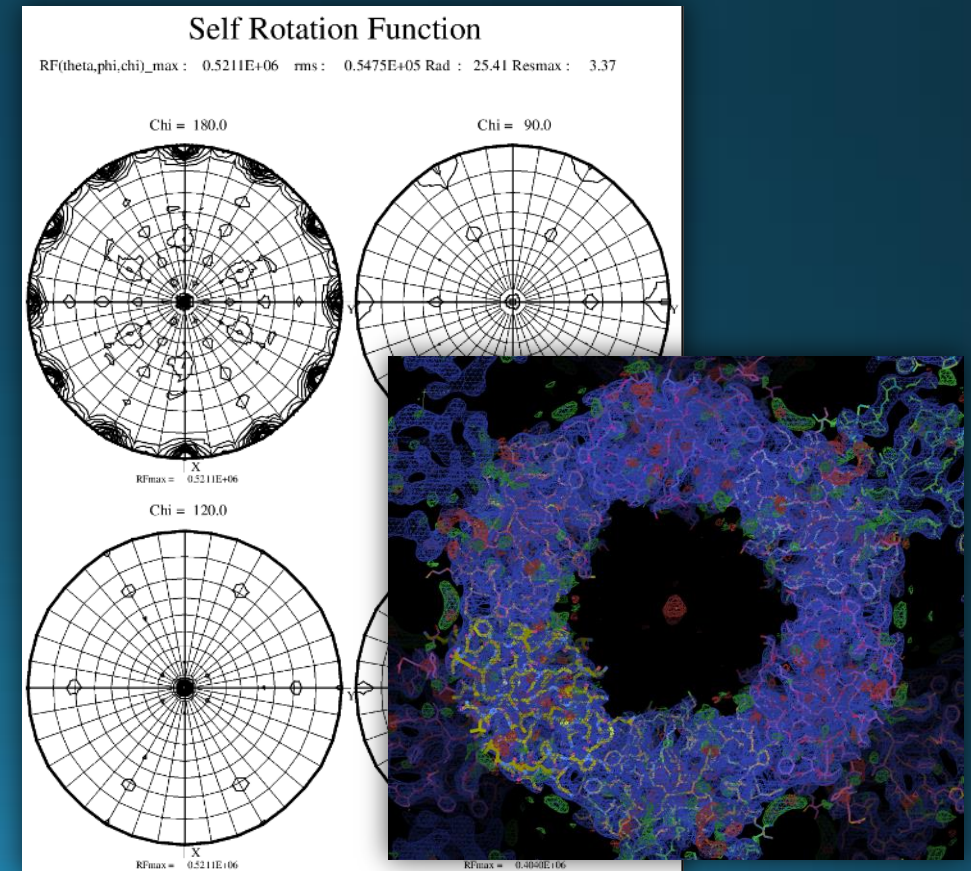


Distribution of solvent content across 156,000 X-ray datasets in the PDB

Understanding the limitations of your 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
 - Self-rotation function – signs of NCS?

Examine the data



*Self rotation function and
map/model for 5i72 – 6 fold
symmetry*

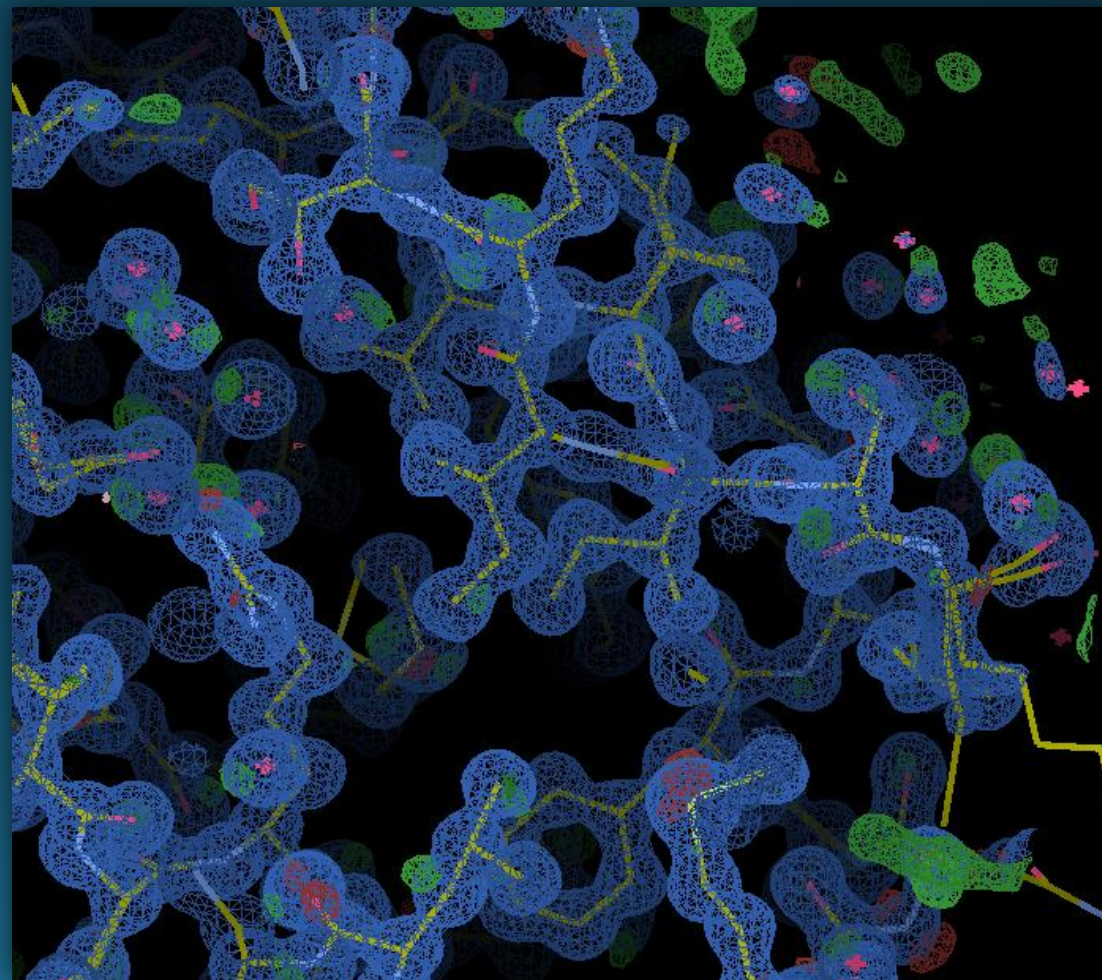
Understanding the limitations of your data

- 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
 - Self-rotation function – signs of NCS?
 - Resolution of the data

Examine the data

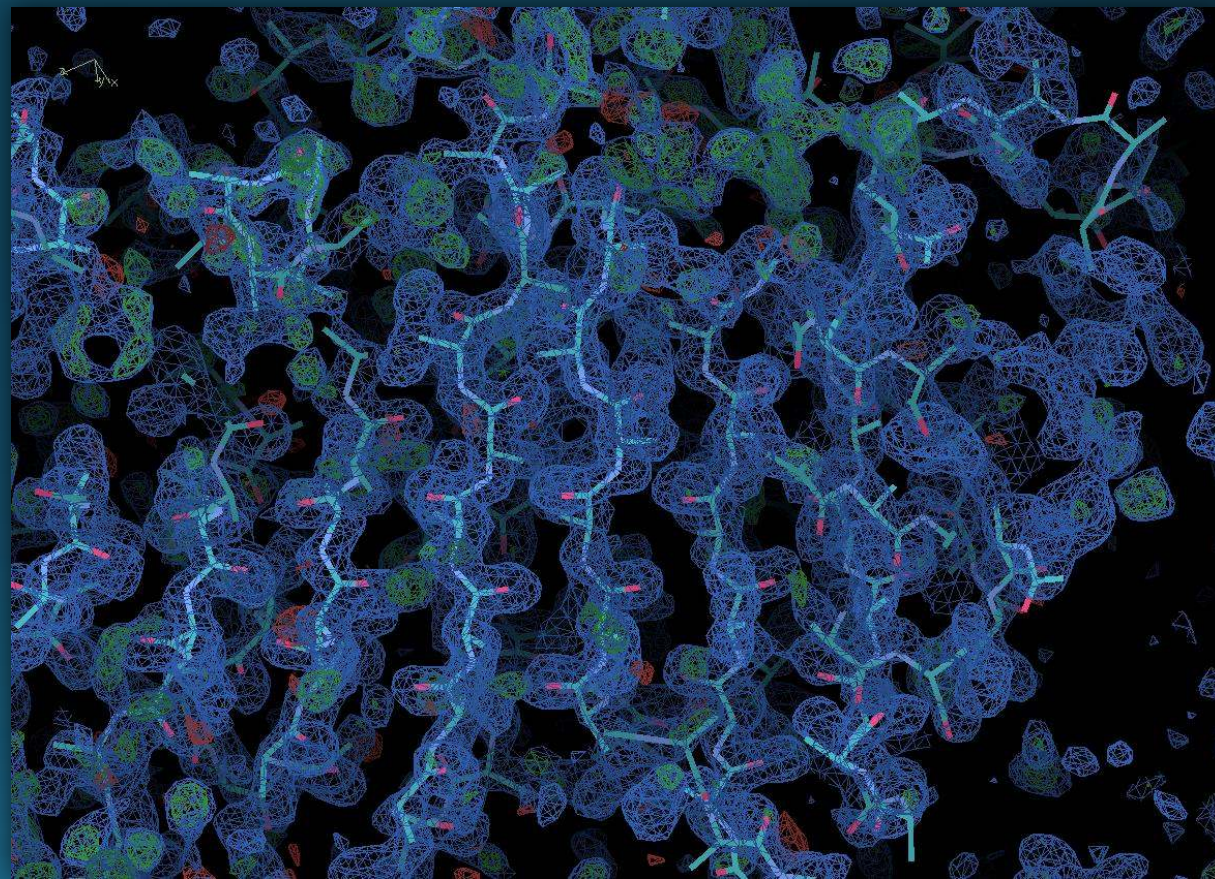
Resolution and Molecular Replacement

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



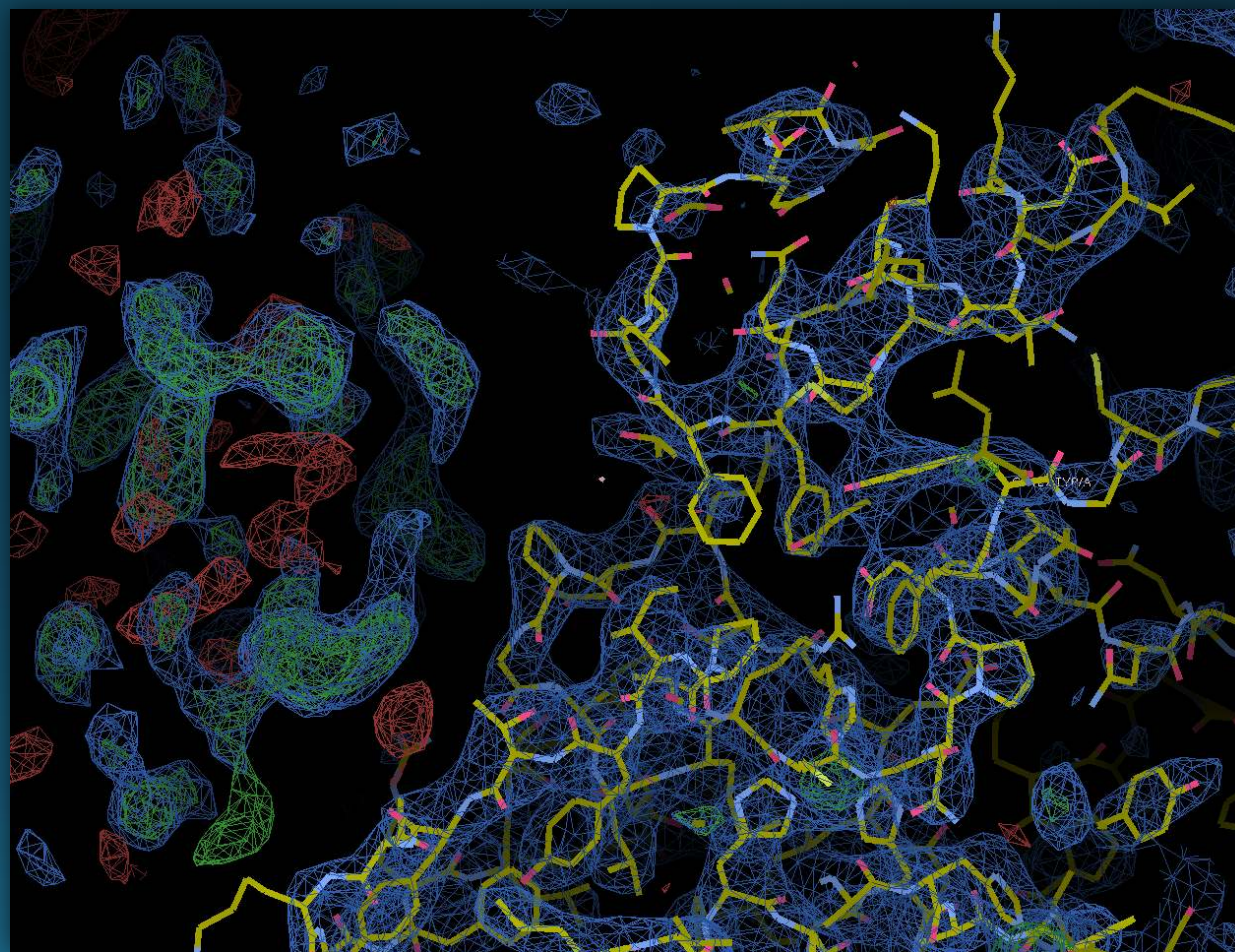
Resolution and Molecular Replacement

- At resolutions above $\sim 2.5 \text{ \AA}$
 - *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



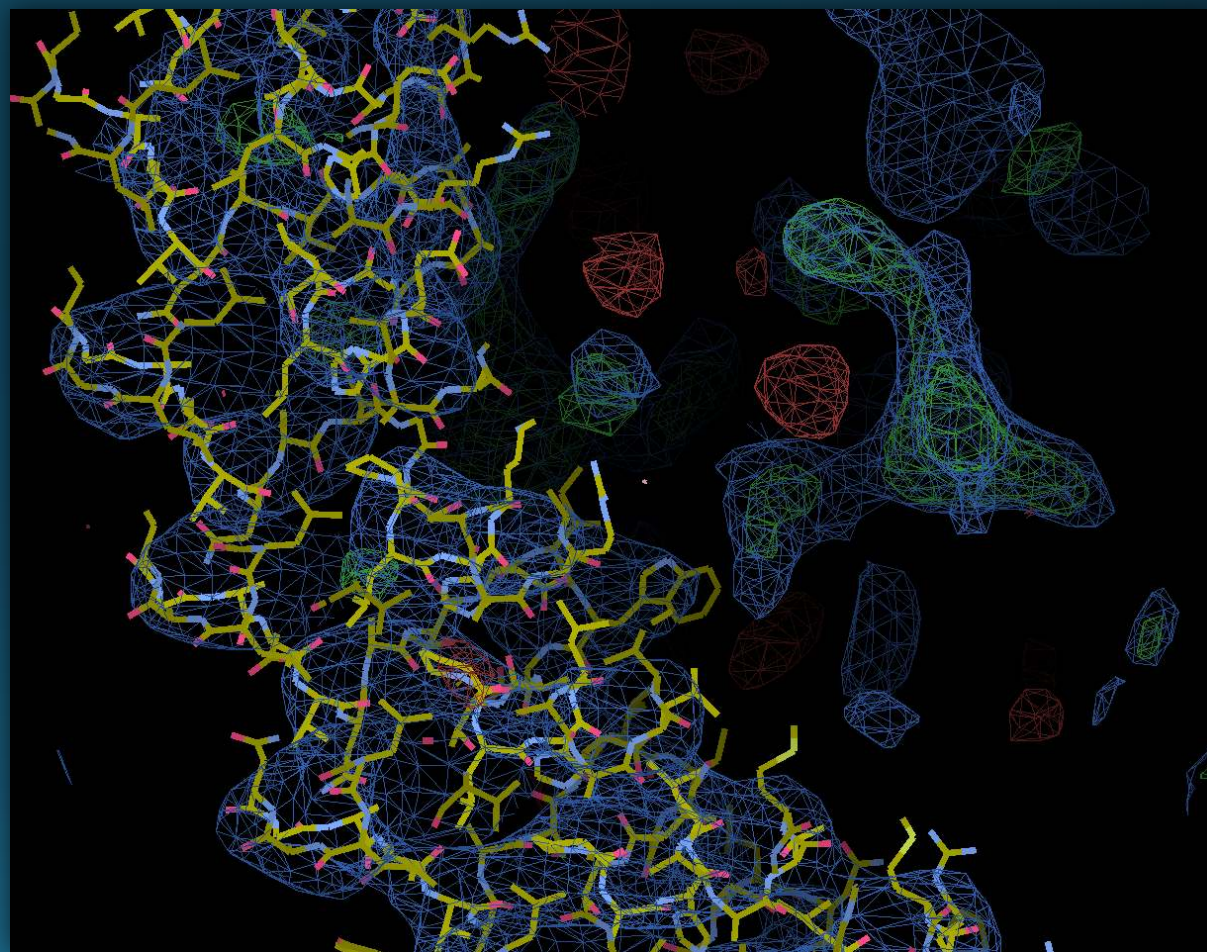
Resolution and Molecular Replacement

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



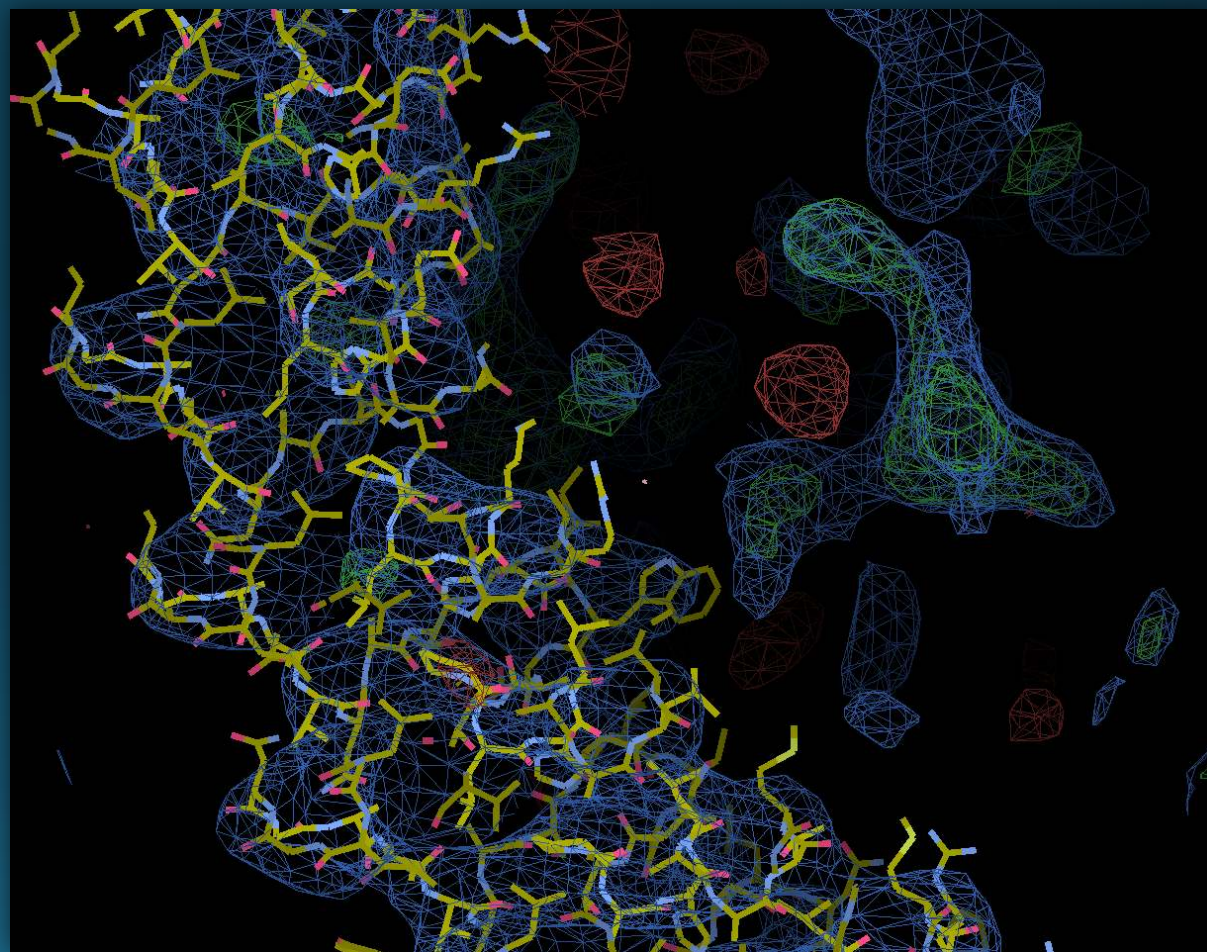
Resolution and Molecular Replacement

- Below 4Å automated model building becomes difficult



Resolution and Molecular Replacement

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



Understanding the limitations of your 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
 - Self-rotation function – signs of NCS?
 - Resolution of the data
- Potential problems:
 - Pseudo translation?

Examine the data

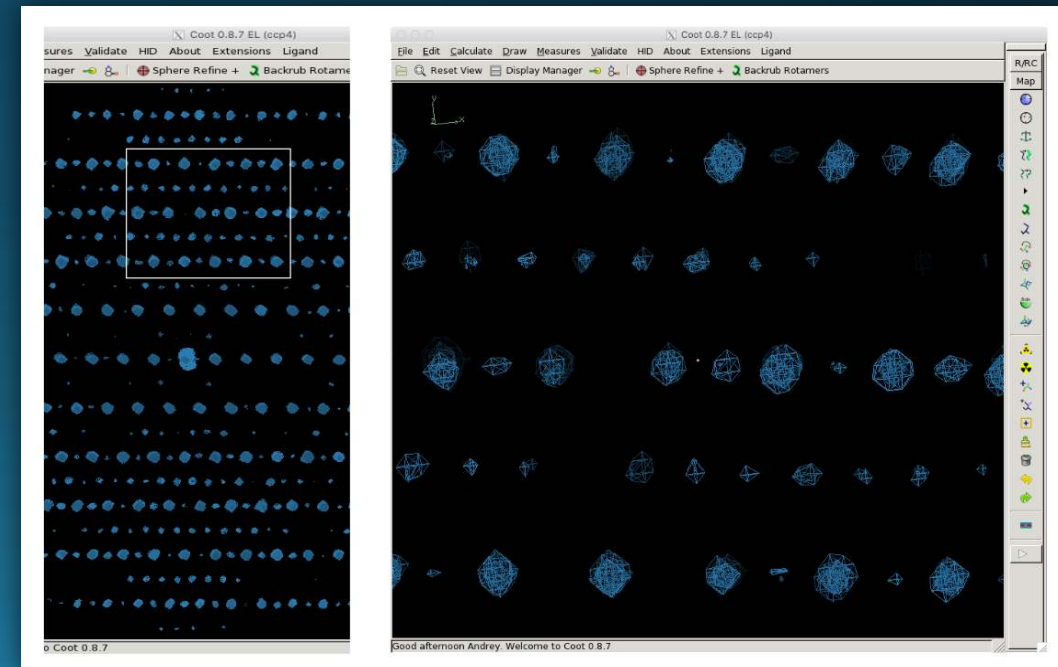
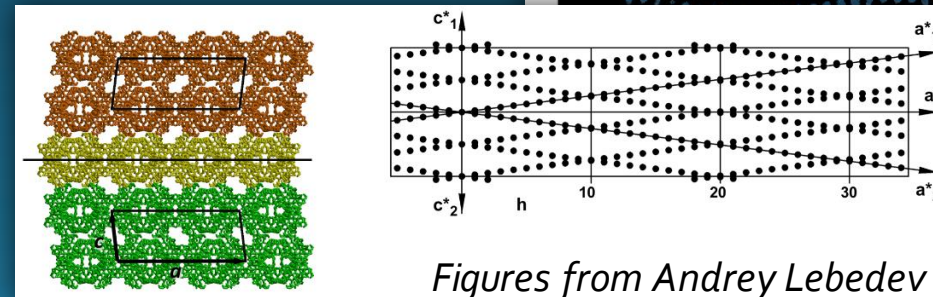
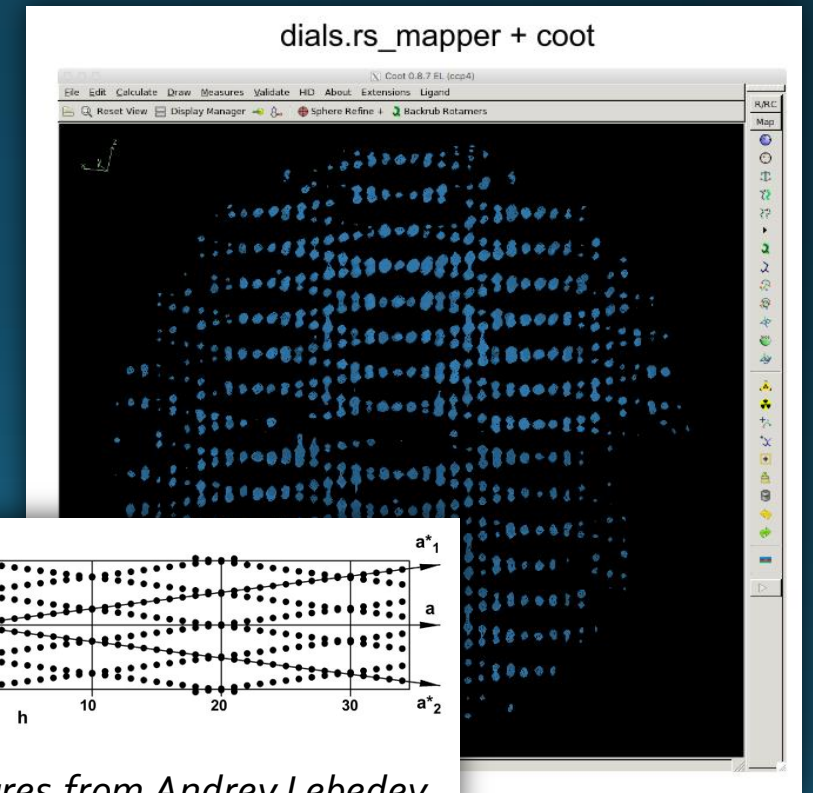


Figure from Andrey Lebedev

Understanding the limitations of your 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
 - Self-rotation function – signs of NCS?
 - Resolution of the data
- Potential problems:
 - Pseudo translation?
 - Twinned data?

Examine the data



Understanding the limitations of your data

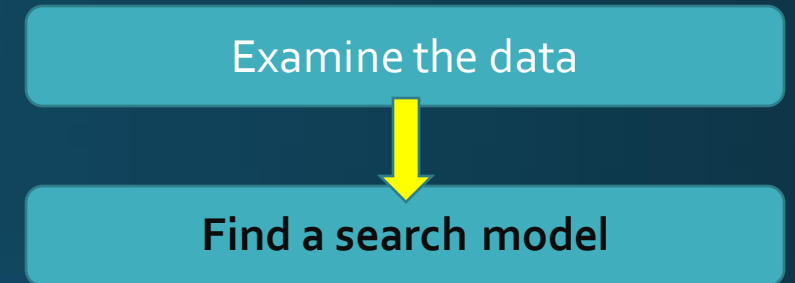
- Data issues can have an impact on how well MR will work
- Things to think about:
 - How many copies in the asymmetric unit?
 - Estimate through Matthews Coefficient
 - Self-rotation function – signs of NCS?
 - Resolution of the data
- Potential problems:
 - Pseudo translation?
 - Twinned data?
 - Space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

Examine the data

Finding suitable Search Models

Finding suitable search models

- Considerations



Finding suitable search models

- Considerations
 - Target Size:
 - The bigger the structure the more likely there will be sizeable conformational changes

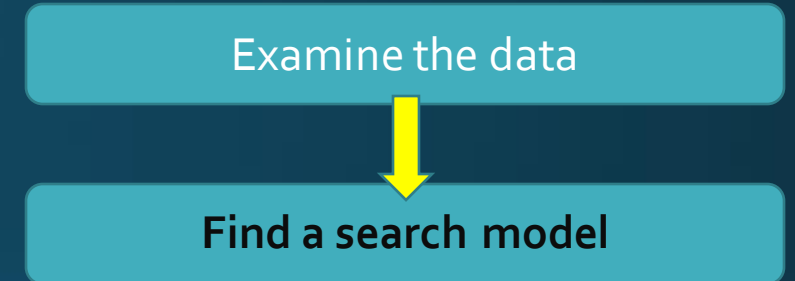
Examine the data



Find a search model

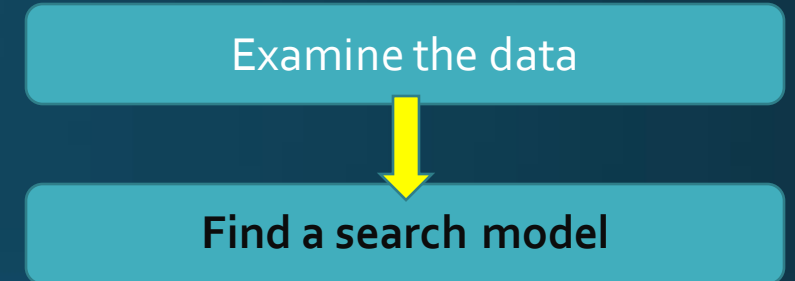
Finding suitable search models

- 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
 - Data Resolution – as discussed before



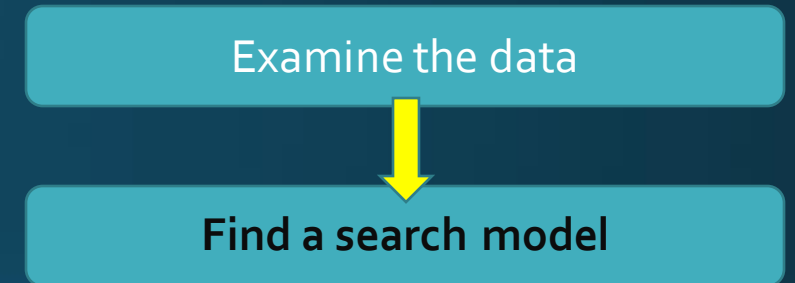
Finding suitable search models

- 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
 - Data Resolution – as discussed before
 - **eLLG** from Phaser is a good measure of whether or not a search model has a chance to be placed correctly



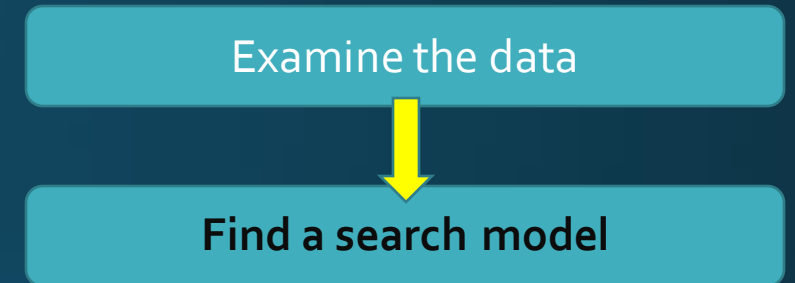
Finding suitable search models

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



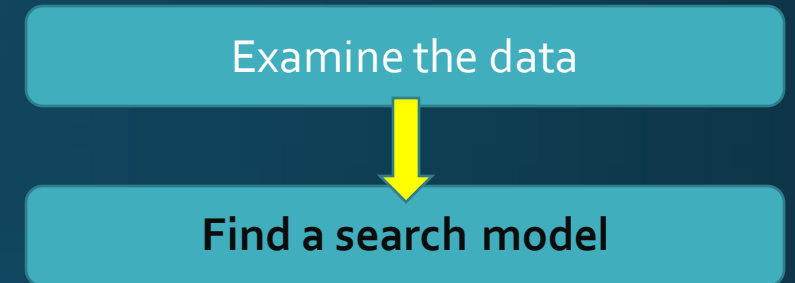
Finding suitable search models

- 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



Finding suitable search models

- 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



How to find a search model for MR?

How to find a search model for MR

- Traditional method
 - Used in the old days (pre-2021)
 - Search the PDB for homologous structures related to the target by sequence

How to find a search model for MR

- Traditional method
 - Used in the old days (pre-2021)
 - Search the PDB for homologous structures related to the target by sequence
- New method
 - Highly accurate prediction of protein structures
 - Google Colab servers
 - EBI - AlphaFold database
 - RoseTTAFold

The traditional method..

Finding suitable search models

- Amino acid sequence similarity often correlates well with structural similarity

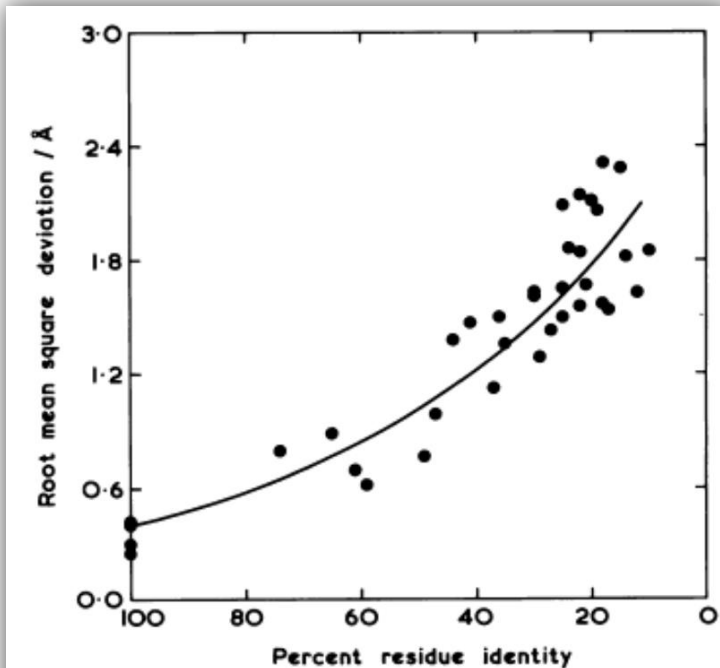


Fig. 2. The relation of residue identity and the r.m.s. deviation of the backbone atoms of the common cores of 32 pairs of homologous proteins (see Table II).

Chotia & Lesk
The EMBO J. 1986 Apr; 5(4): 823–826.

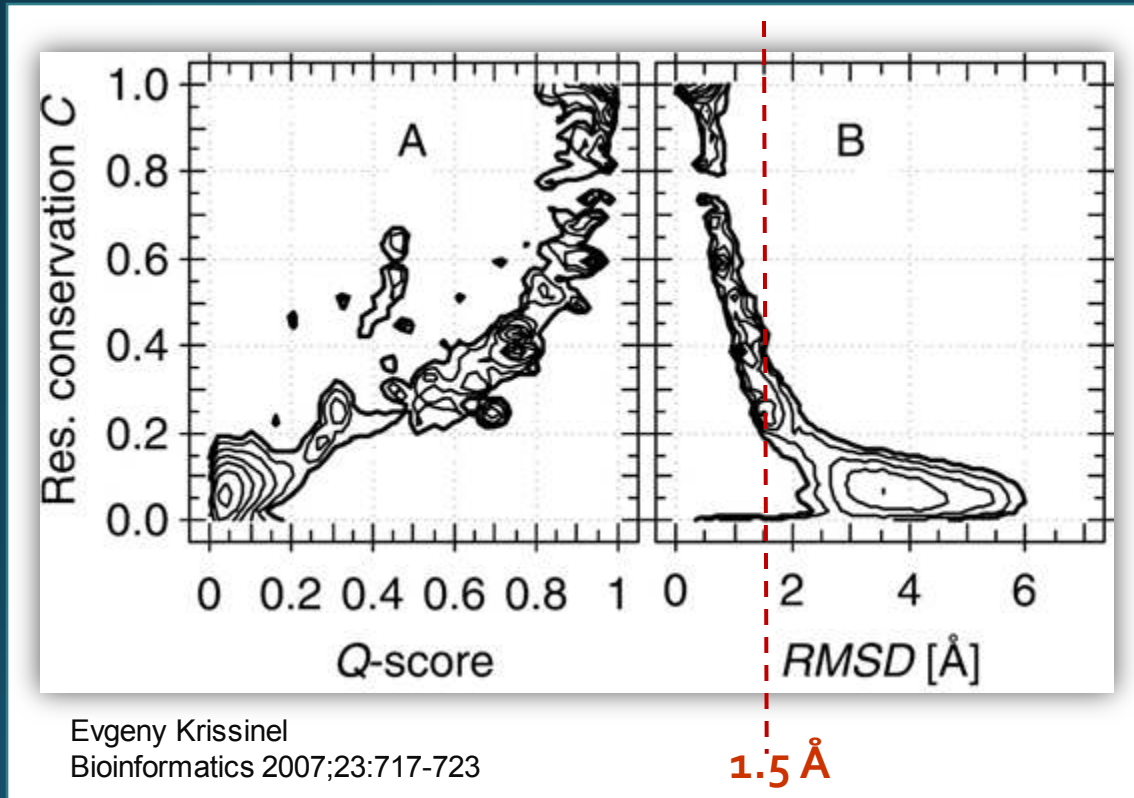
Examine the data



Find a search model

Finding suitable search models

- Amino acid sequence similarity often correlates well with structural similarity



Examine the data



Find a search model

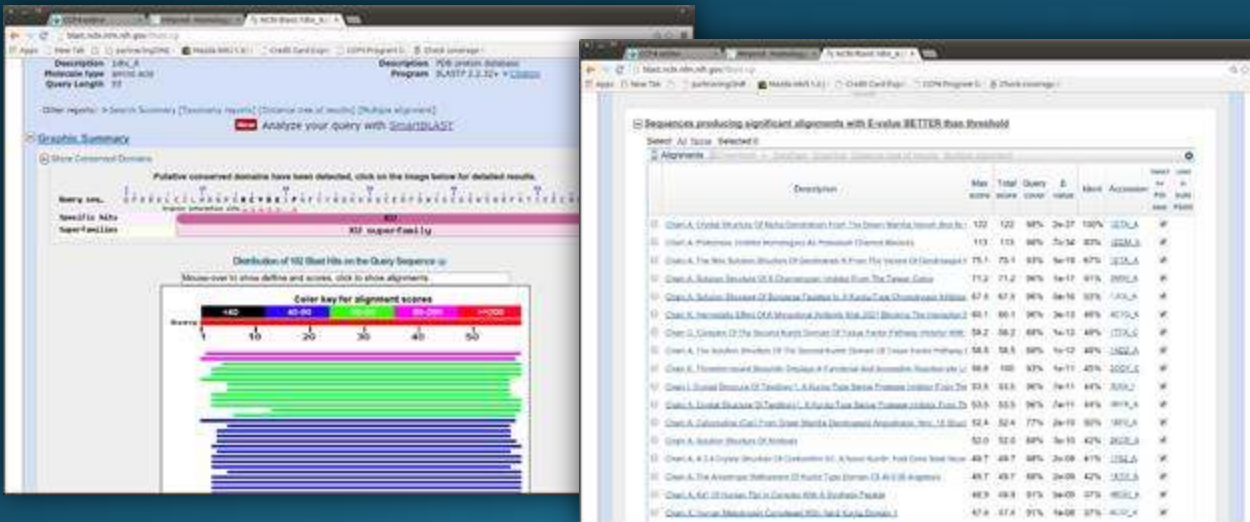
Finding suitable search models

- 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



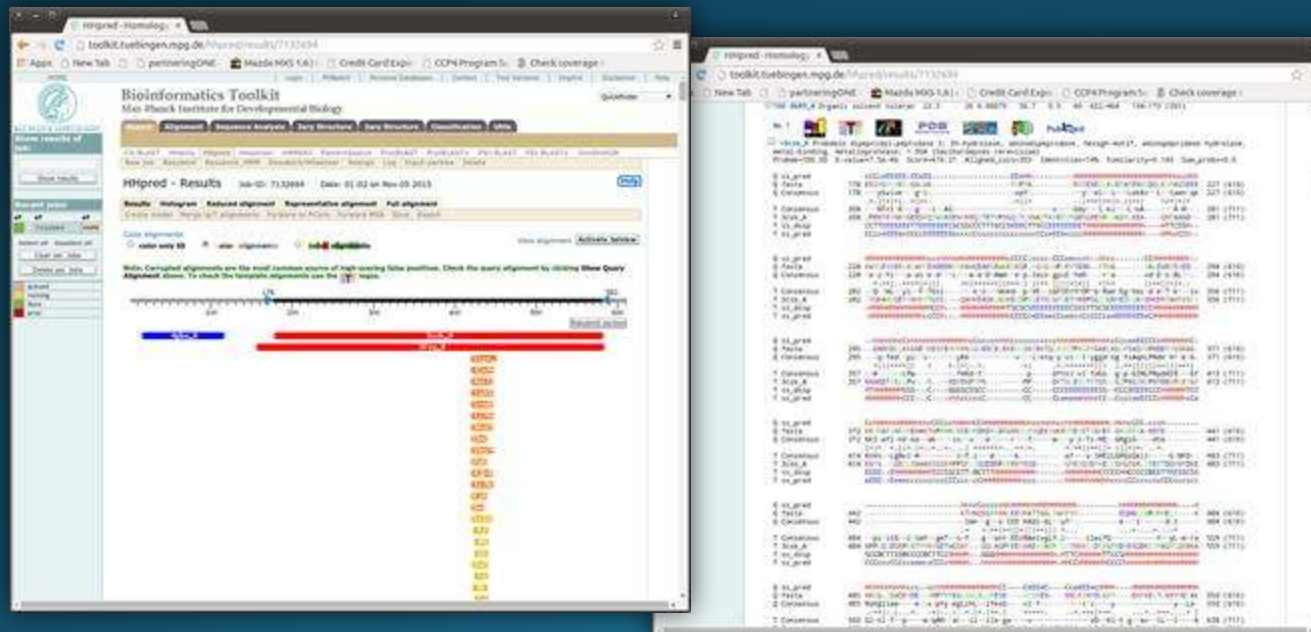
Finding suitable search models

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

Examine the data



Find a search model



The modern method..

Finding suitable search models

- CASP₁₄
 - Critical Assessment of Techniques for Protein Structure Prediction
 - Participants predict structures for unreleased data and assessed against the true structure

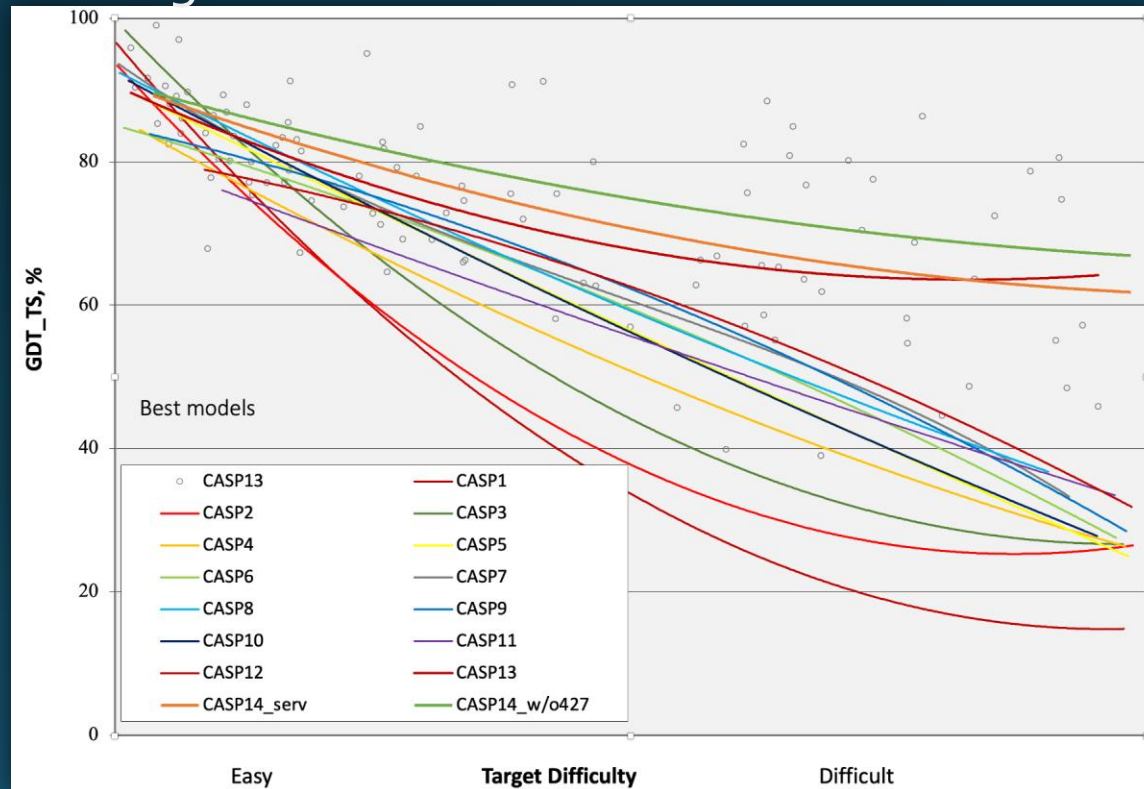
Examine the data



Find a search model

Finding suitable search models

- CASP₁₄
 - Critical Assessment of Techniques for Protein Structure Prediction
 - Participants predict structures for unreleased data and assessed against the true structure



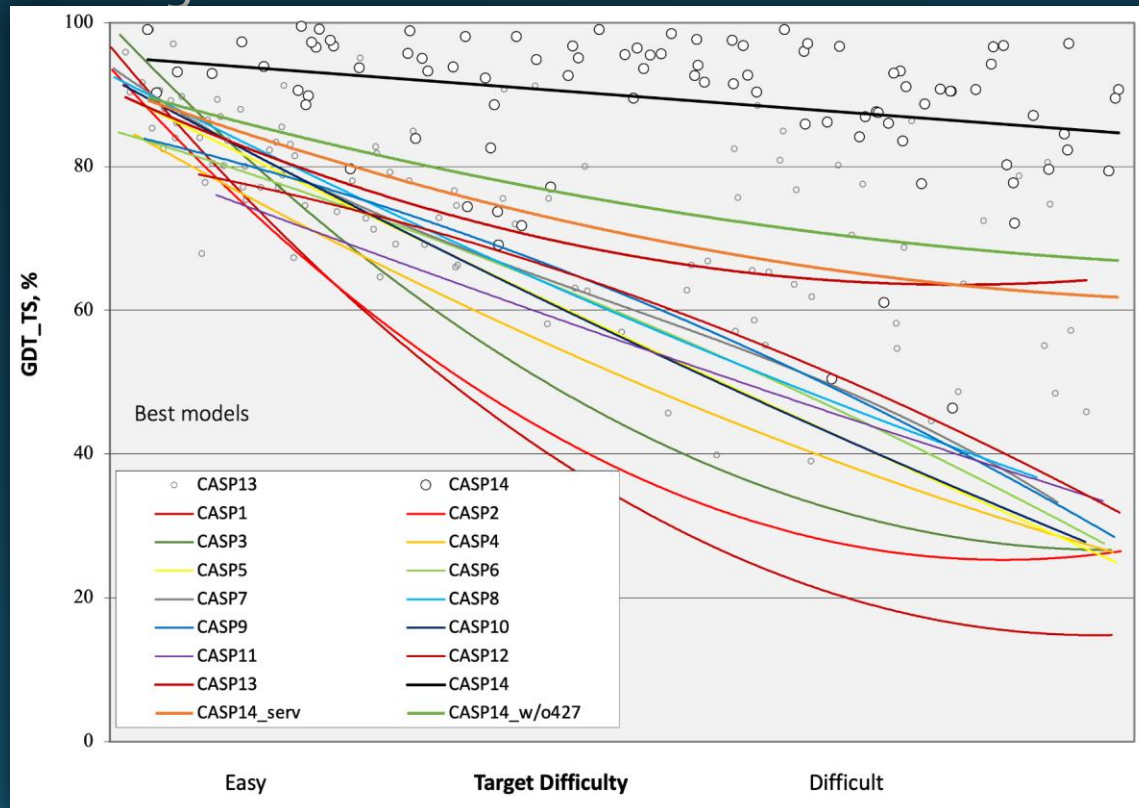
Examine the data



Find a search model

Finding suitable search models

- CASP₁₄
 - Critical Assessment of Techniques for Protein Structure Prediction
 - Participants predict structures for unreleased data and assessed against the true structure



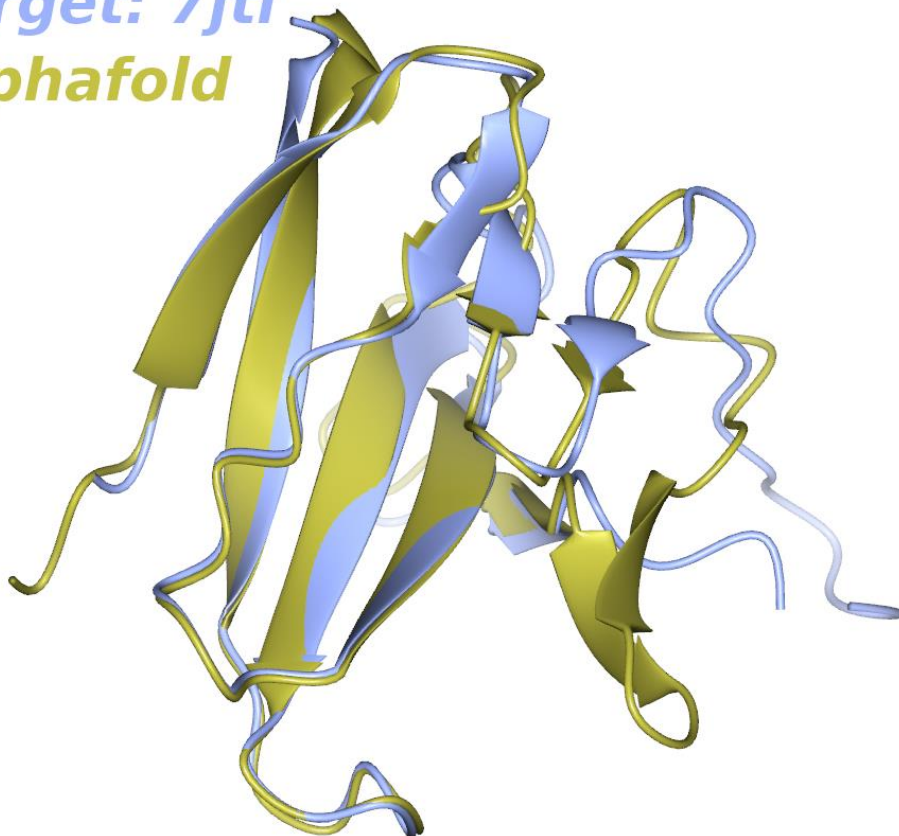
Examine the data



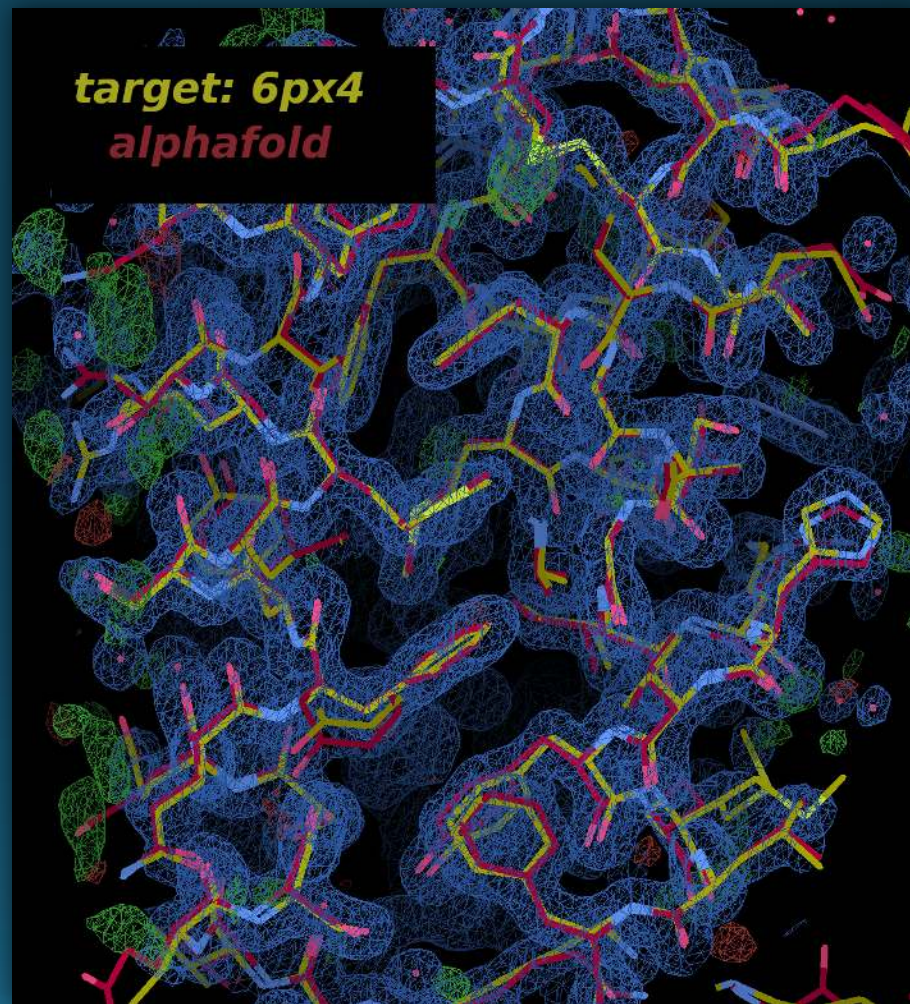
Find a search model

- *AlphaFold CASP₁₄ examples*

target: 7jtl
alphafold

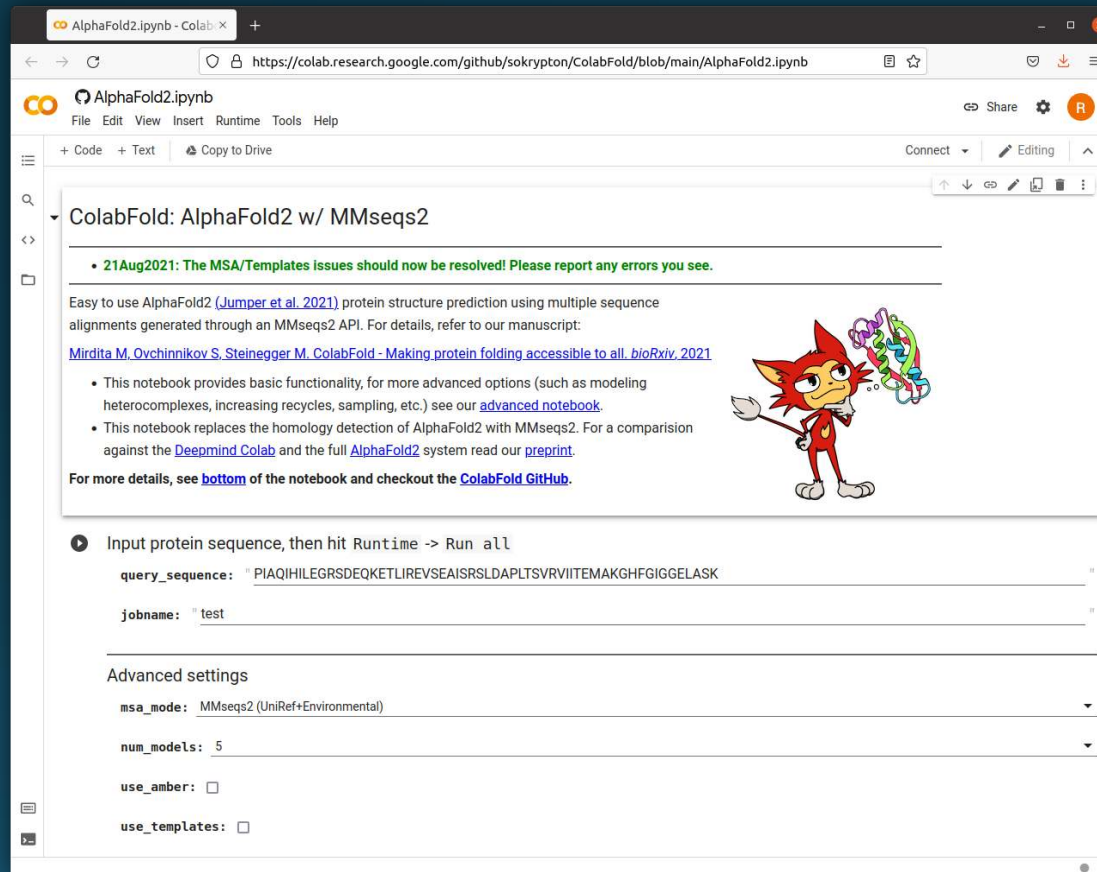


target: 6px4
alphafold



Finding suitable search models

- Accessing Alphafold
 1. Google Colab Servers (see Claudia's talk next)



Examine the data



Find a search model

Finding suitable search models

- Accessing AlphaFold
 1. Google Colab Servers (see Claudia's talk next)
 2. EBI-AlphaFold database
 1. 350000 predictions currently available
 2. 130 million in the near future (early 2022)

Examine the data



Find a search model

The screenshot shows the AlphaFold Protein Structure Database interface. The top navigation bar includes links for Home, About, FAQs, and Downloads. A search bar is present with a 'Search' button. Below the search bar, there are example search terms: 'Free fatty acid receptor 2', 'At1g58602', 'Q9VSL9', and 'E. coli'. The main content area displays the entry for 'Probable disease resistance protein At1g58602'. It includes a section for 'AlphaFold structure prediction' with download options for PDB file, mmCIF file, and Predicted aligned error. An 'Information' section provides details about the protein, gene, source organism (Arabidopsis thaliana), UniProt accession (Q9VSL9), and experimental structures. A '3D viewer' section shows a ribbon diagram of the protein structure, with a legend for model confidence (Very high, Confident, Low, Very low) and a note about the per-residue confidence score (pLDDT).

Finding suitable search models

- Accessing AlphaFold
 1. Google Colab Servers (see Claudia's talk next)
 2. EBI-AlphaFold database
 1. 350000 predictions currently available
 2. 130 million in the near future (early 2022)

Examine the data



Find a search model

The screenshot shows the AlphaFold Protein Structure Database website. The main heading is "Probable disease resistance protein At1g58602". Below it, there's a section for "AlphaFold structure prediction" with download links for PDB File, mmCIF File, and Predicted aligned error. The "Information" section lists details about the protein, including its name, gene, source organism (Arabidopsis thaliana), UniProt ID (Q9W3K0), and biological function. A "3D viewer" section shows a ribbon diagram of the protein structure with a confidence scale from Very high (pLDDT > 90) to Very low (pLDDT < 50). The structure is shown in blue and yellow, with a legend indicating the confidence levels.

The screenshot shows the CCP4 Project Viewer interface. The "Job list" tab is active, displaying a list of jobs. The job "6 Automated structure solution - MrBump" is highlighted, showing its status as "Pending". Other jobs in the list include "5 CCP4mg MrBUMP", "4 Calculate self rotation function", "3 Define AU contents", "2 Import sequence(s)", and "1 Import merged". The "Input" tab for the selected job shows settings for the search, including the non-redundancy level for the homologue search (set to 100) and the search for additional models from the EBI-AlphaFoldDB.

The screenshot shows the Auto-MR with MrBump web interface. The "job description" section shows the job name "mr bump" and the output ID "mr bump". The "Parameters" section includes options to "Check alternative space groups" and "Include structures from AFDB", with a "Maximum no. of models to test" set to 20. The interface also shows a "Structure revision" section with a dropdown menu for selecting a specific model.

Finding suitable search models

- Accessing Alphafold
 1. Google Colab Servers (see Claudia's talk next)
 2. EBI-Alphafold database
 1. 350000 predictions currently available
 2. 130 million in the near future (early 2022)
- RobeTTAFold Server

Examine the data



Find a search model

The screenshot shows the Robetta website interface. At the top, there's a navigation bar with 'Robetta', 'Project', 'Structure Prediction', and links for 'Register' and 'Login'. Below the navigation bar, a description states: 'Robetta is a protein structure prediction service that is continually evaluated through CAMEO'. It lists features like deep learning-based methods (RoseTTAFold and TrRosetta), homology modeling, and constraints. A section titled 'Recent alerts and bug fixes:' contains a list of updates and issues, such as browser compatibility issues, server backlogs, and new modeling options. On the right side of the page, there's a 3D protein structure model and a sequence alignment viewer showing a sequence: MERLDI...SVPIRVTMIQAVD...

-
- MrParse**
- Version 0.2.1
- MrParse is a program to find and analyse search models for crystallographic Molecular Replacement. The program is being developed by [Daisuke Nakagawa](#) at the University of Liverpool. MrParse is currently under development and we are keen to make it as useful to the community as possible. If you have any suggestions for its development, or think we have made a mistake, please [contact us](#).
- Experimental structures from the PDB**
- | Name | PDB | Resolution | Region | Range | Length | eLIG | Mol. Wt. | eRMG | Seq. Ident. |
|----------|------|------------|--------|--------|--------|-------|----------|------|-------------|
| 3a0k_C.1 | 3a0k | 2.50 | 1 | 1-100 | 98 | 10702 | | | 0.31 |
| 1a1v_B.2 | 1a1v | 0.00 | 1 | 1-107 | 105 | 10444 | | | 0.24 |
| 6a55_L.3 | 6a55 | 0.00 | 2 | 12-76 | 63 | 7196 | | | 0.28 |
| 1a1v_B.2 | 1a1v | 0.00 | 2 | 31-75 | 43 | 2956 | | | 0.35 |
| 6a55_L.2 | 6a55 | 0.00 | 3 | 31-98 | 66 | 7601 | | | 0.34 |
| 3a02_D.1 | 3a02 | 2.89 | 3 | 54-99 | 44 | 5227 | | | 0.26 |
| 3a01_D.2 | 3a01 | 0.00 | 3 | 54-99 | 44 | 5227 | | | 0.26 |
| 3a01_D.3 | 3a01 | 0.00 | 3 | 54-99 | 44 | 5227 | | | 0.26 |
| 3a01_D.4 | 3a01 | 0.00 | 3 | 54-98 | 43 | 5114 | | | 0.27 |
| 3a01_B.1 | 3a01 | 2.91 | 3 | 32-105 | 71 | 8015 | | | 0.28 |
| 3a11_A.1 | 3a11 | 1.67 | 3 | 32-105 | 72 | 8045 | | | 0.27 |
| 3a07_A.1 | 3a07 | 2.59 | 3 | 54-99 | 44 | 5227 | | | 0.26 |
| 3a07_A.2 | 3a07 | 0.00 | 3 | 54-99 | 44 | 5206 | | | 0.26 |
| 3a07_A.3 | 3a07 | 0.00 | 3 | 54-97 | 43 | 5173 | | | 0.27 |
| 6a5_C.1 | 6a5 | 1.99 | 3 | 82-101 | 19 | 1723 | | | 0.37 |
| 8a0_A.2 | 8a0 | 0.00 | 3 | | | | | | |
| 3a0_A.1 | 3a0 | 1.51 | 3 | | | | | | |
| 3a0_A.2 | 3a0 | 0.00 | 3 | | | | | | |
| 6a5_L.1 | 6a5 | 1.80 | 4 | | | | | | |
| 1a1v_B.1 | 1a1v | 2.80 | 4 | | | | | | |
| 8a0_A.1 | 8a0 | 1.83 | 4 | | | | | | |
- Visualisation of Regions**
- Visualisation of the protein structure showing the regions of interest. The structure is a ribbon diagram, colored by domain. The regions are labeled with their corresponding PDB IDs and sequence ranges.
- Structure predictions from the EBI AlphaFold database**
- | Name | model | Date Made |
|--------|--------|-----------|
| P06027 | P06027 | 01-JAN-20 |
| P02042 | P02042 | 01-JAN-20 |
| P06079 | P06079 | 01-JAN-20 |
| P06088 | P06088 | 01-JAN-20 |
| P06088 | P06088 | 01-JAN-20 |

- correspond to sequence identity (red for high, green for low). For the EBI-APDB hits, the colours indicate the predicted model quality (pLDDT). They range from blue to indicate high confidence to orange for low confidence residues.



- The Interactive model preparation CCP4mg-MrBUMP task can be found in the Bioinformatics menu. Provide the sequence loaded above and use default settings. By default, the task will identify homologues from the PDB and the EBI-AFDB and display them in CCP4mg (takes a couple of minutes). You can search only the PDB or EBI-AFDB by unticking the relevant box in the input menu. For AFDB models, you can adjust the pLDDT cut-off. Higher values mean only the most confidently predicted residues are retained in the search models (**Note:** the CCP4mg-MrBUMP task will automatically convert pLDDT scores to B-factors for EBI-AFDB models).

job 2: Interactive model preparation - CCPing and MrT. The job is Pending

topA Results Comments

Input data

job id: CCPLing.mrT.msp

 the data from job 1 ☐ Define all contents ☐ as input below

Sequences from all contents

☒ job contents ☐ All contents from both further to contents

☒ If a variable is in a not included item, you can press the cross, it then further to clearly check one

Save job

name: ginseng

Model databases

☒ search MrT for possible new compounds

☒ non-redundancy limit for development search

120

☒ search standards for possible new compounds

☒ use MRTP job2007 model to compare with

12

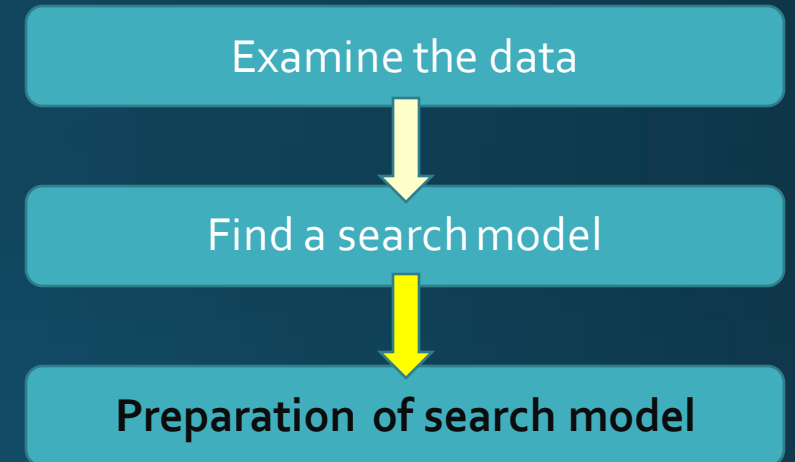
Maximum no. of search results to create

10

Preparing Search Models for Molecular Replacement

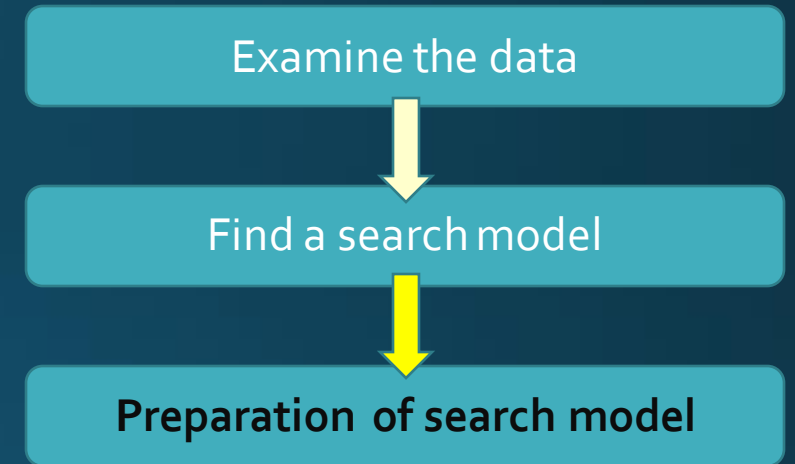
Preparing models for Molecular Replacement

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



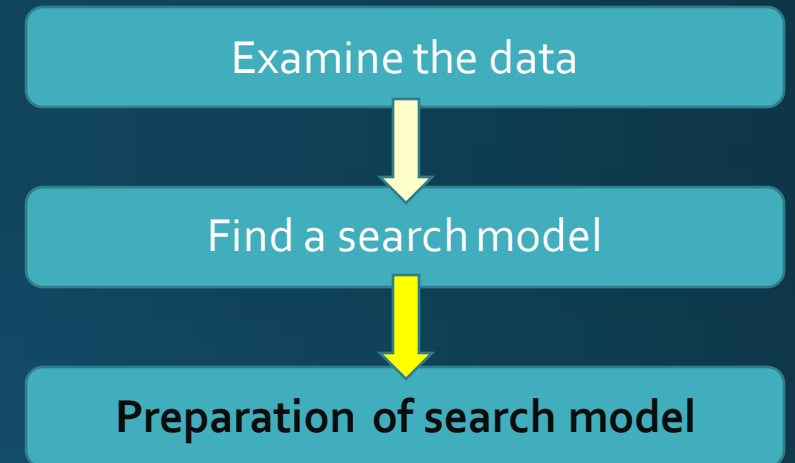
Preparing models for Molecular Replacement

- 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



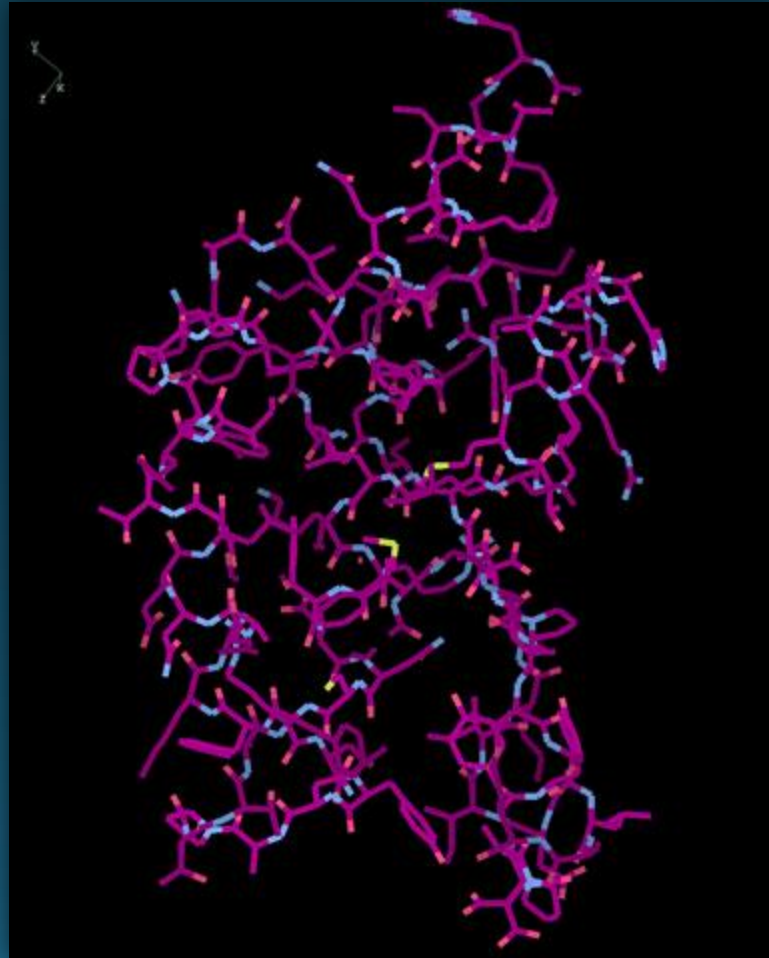
Preparing models for Molecular Replacement

- 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



Preparing models for Molecular Replacement

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



Examine the data



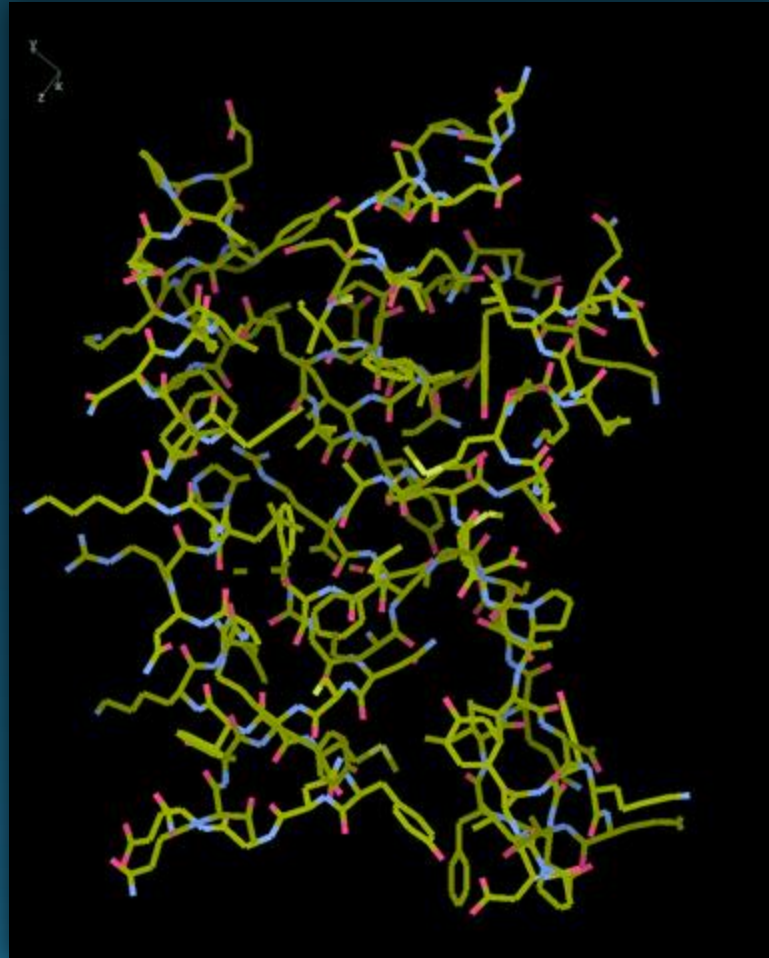
Find a search model



Preparation of search model

Preparing models for Molecular Replacement

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



Examine the data



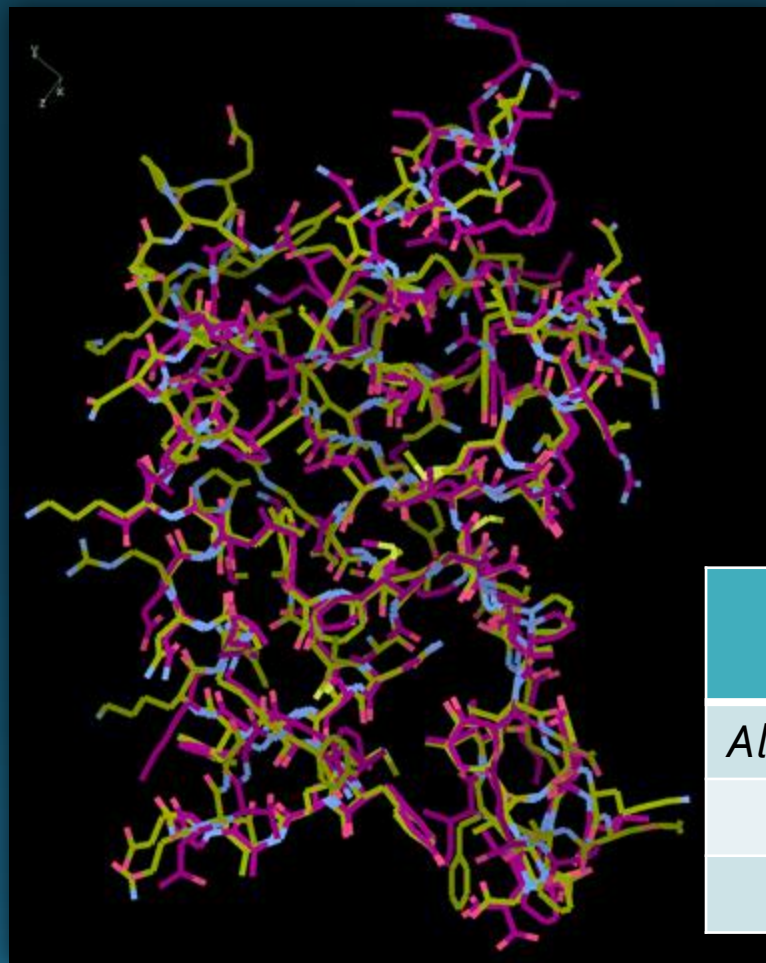
Find a search model



Preparation of search model

Preparing models for Molecular Replacement

- Search Model preparation and MR:
 - All atoms?



Examine the data



Find a search model

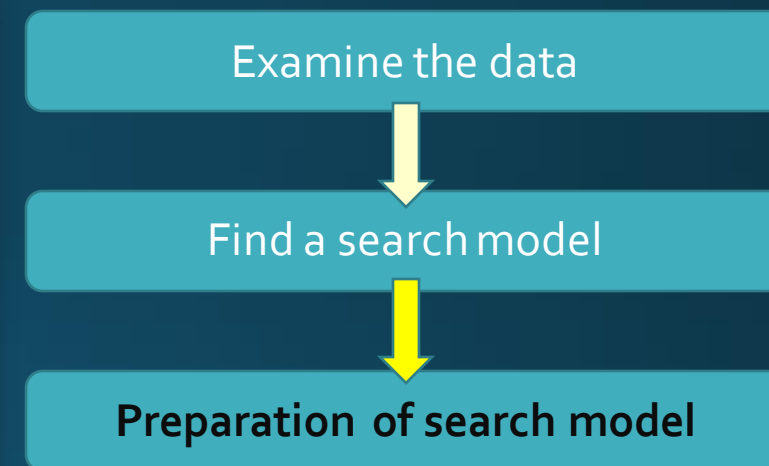
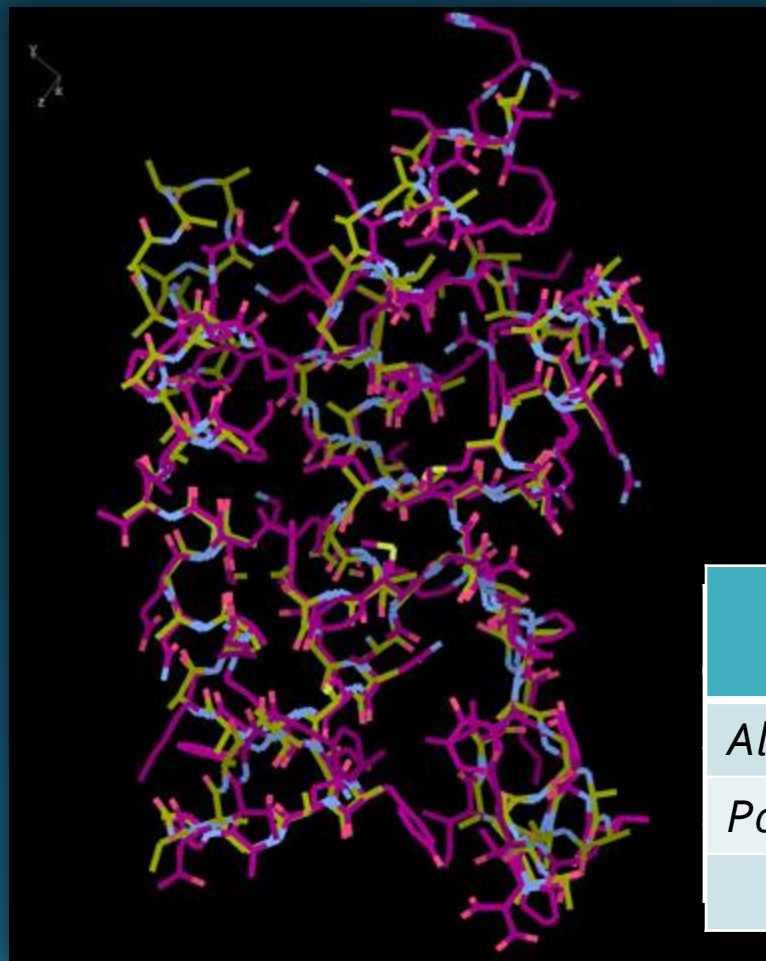


Preparation of search model

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

Preparing models for Molecular Replacement

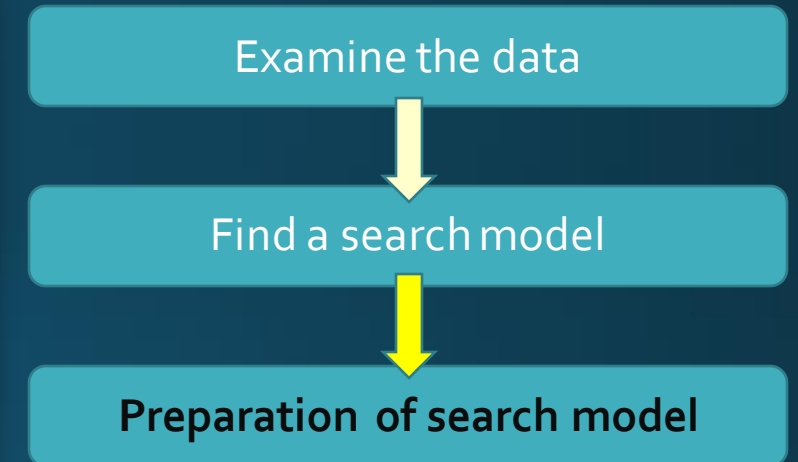
- Search Model preparation and MR:
 - All atoms?
 - Polyalanine?



	<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

Preparing models for Molecular Replacement

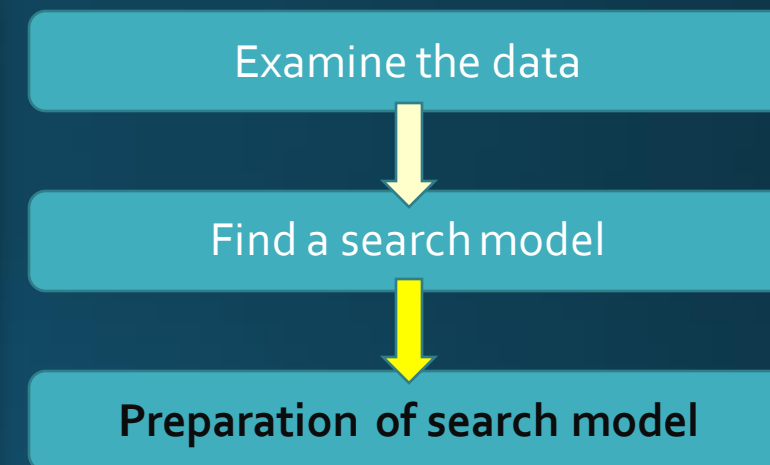
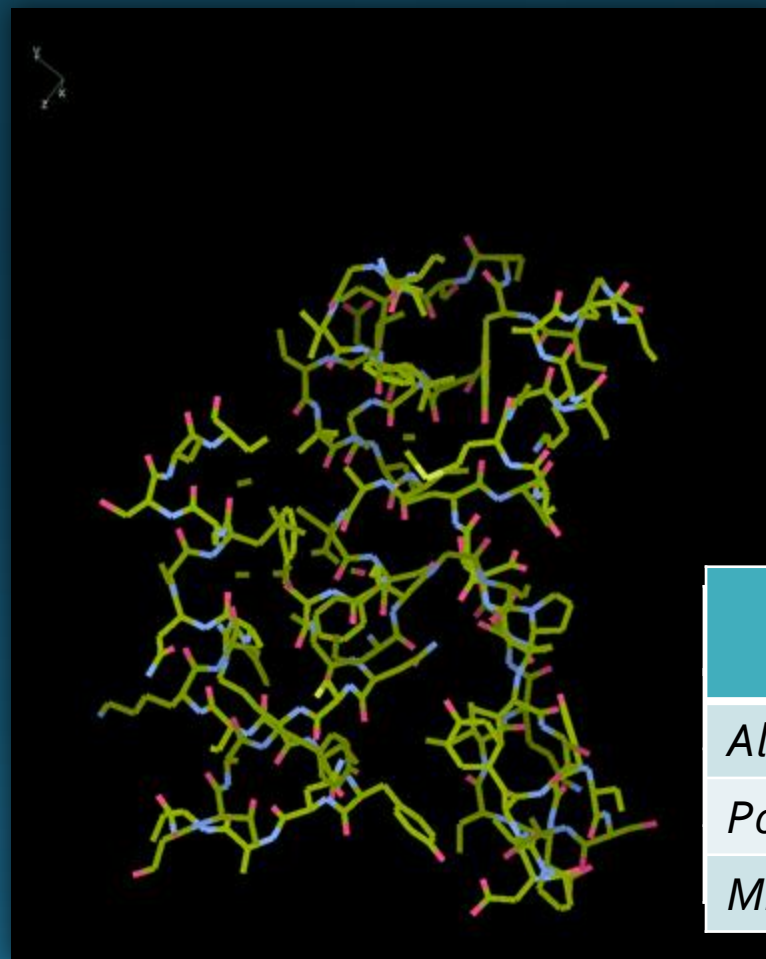
- Search Model preparation and MR:
 - All atoms?
 - Polyalanine?
 - Mixed model based on sequence alignment
 - Keep common residues
 - Truncate to common atoms on aligned residues



	<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

Preparing models for Molecular Replacement

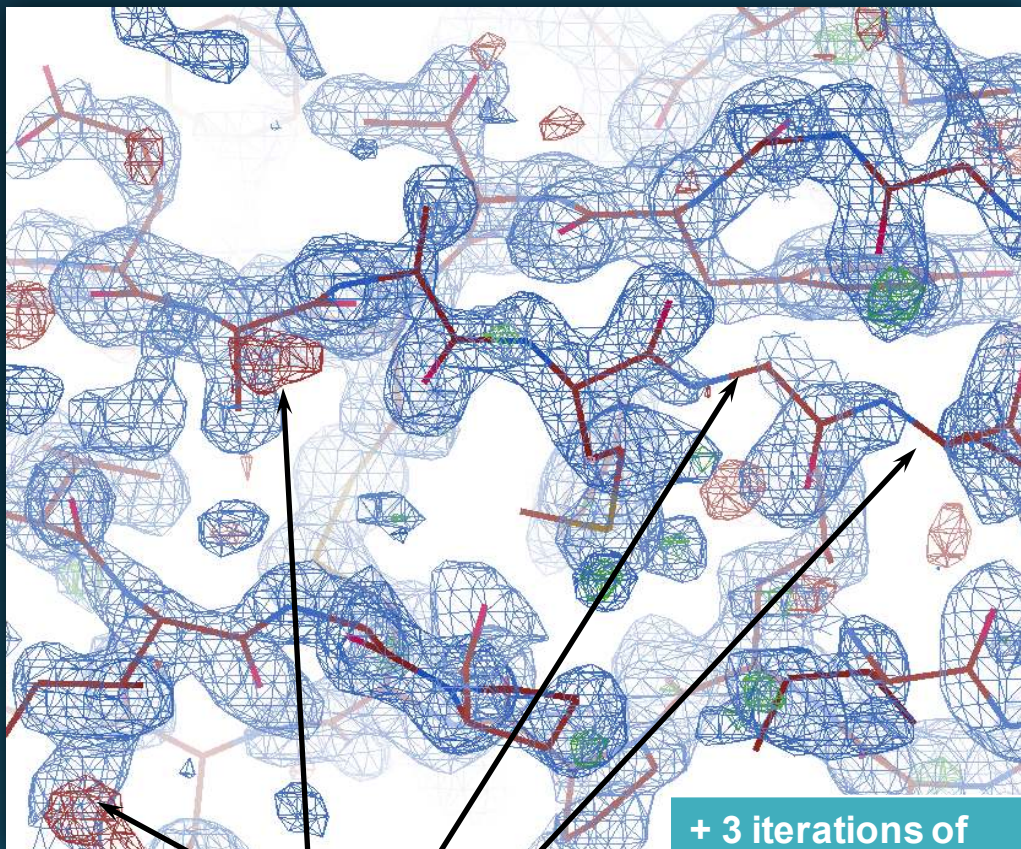
- Search Model preparation and MR:
 - All atoms?
 - Polyalanine?
 - Mixed model based on sequence alignment
 - Keep common residues
 - Truncate to common atoms on aligned residues



	<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

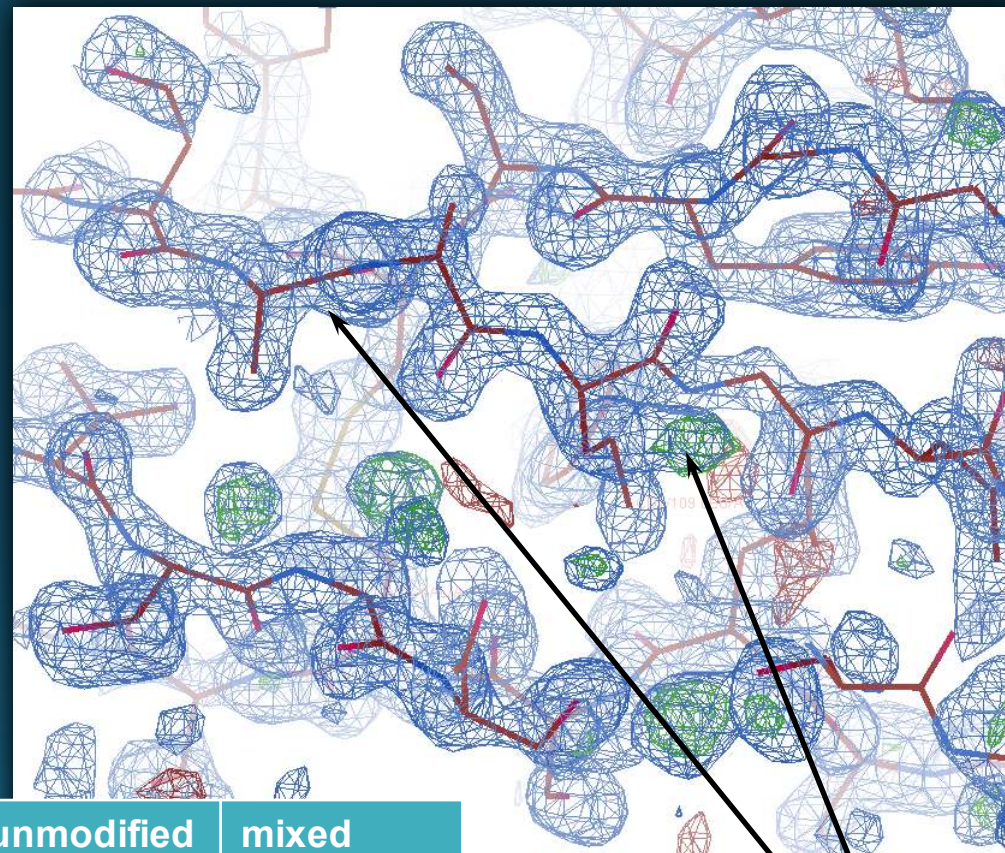
Preparing models for Molecular Replacement

Unmodified model:



bad density features

Mixed model:

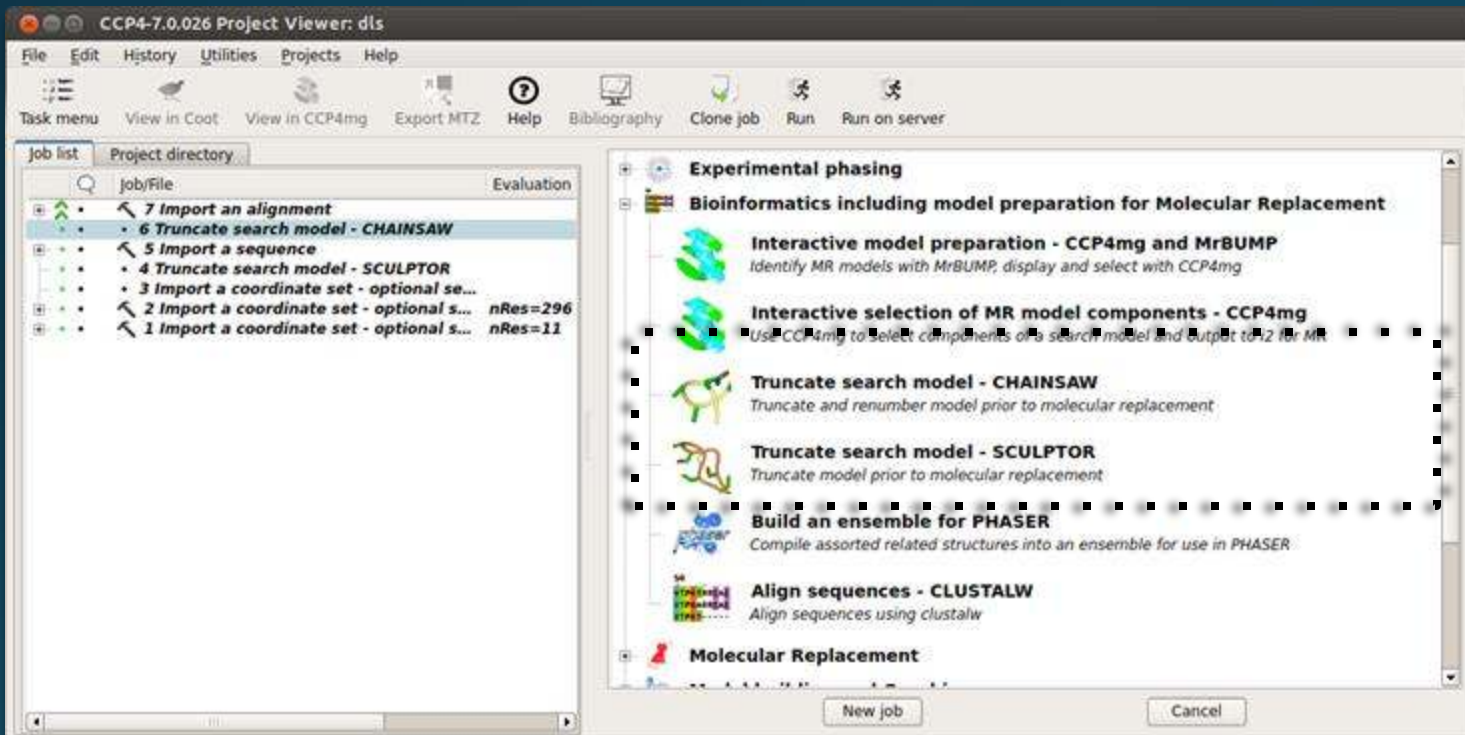


good density features

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

Preparing models for Molecular Replacement

- Mixed-model generation in CCP4
 - Chainsaw & Sculptor



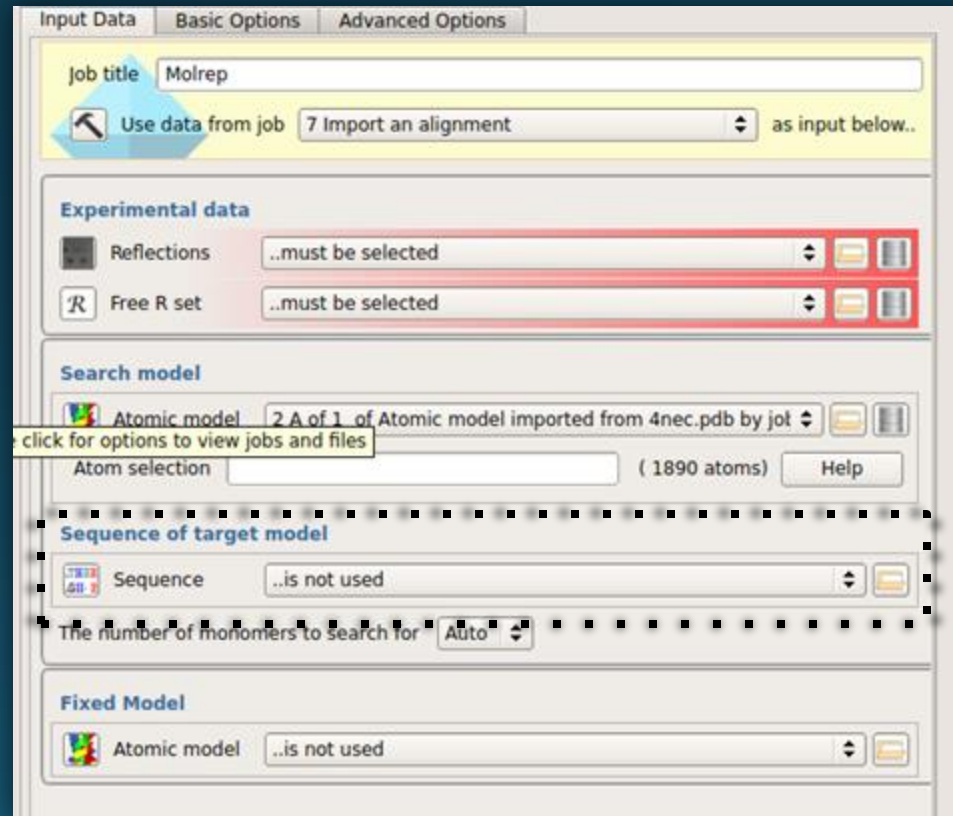
Examine the data

Find a search model

Preparation of search model

Preparing models for Molecular Replacement

- Mixed-model generation in CCP4
 - Molrep



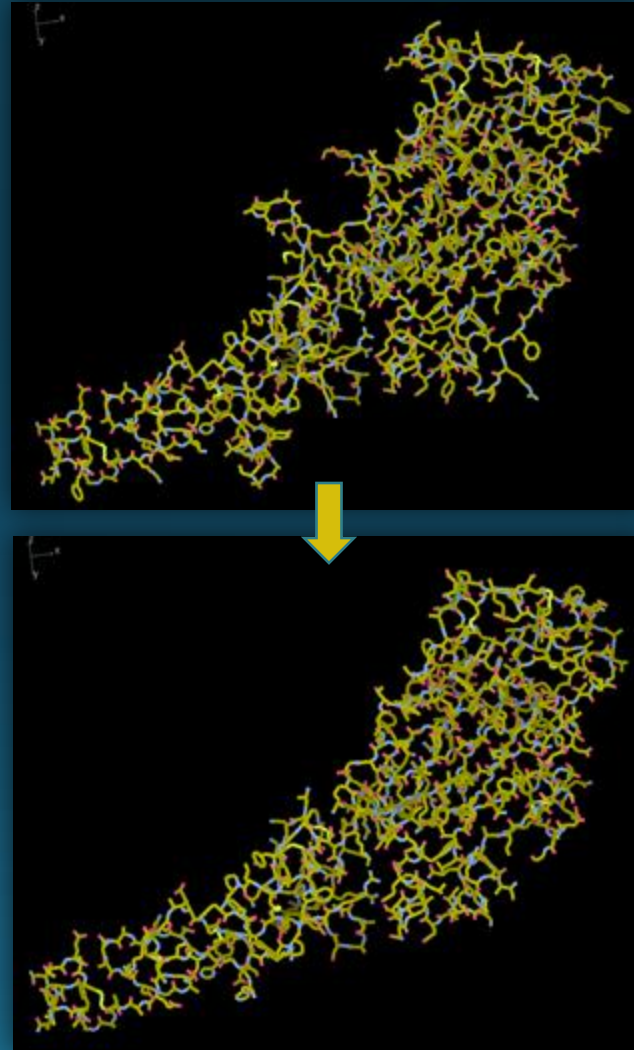
Examine the data

Find a search model

Preparation of search model

Preparing models for Molecular Replacement

- 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”



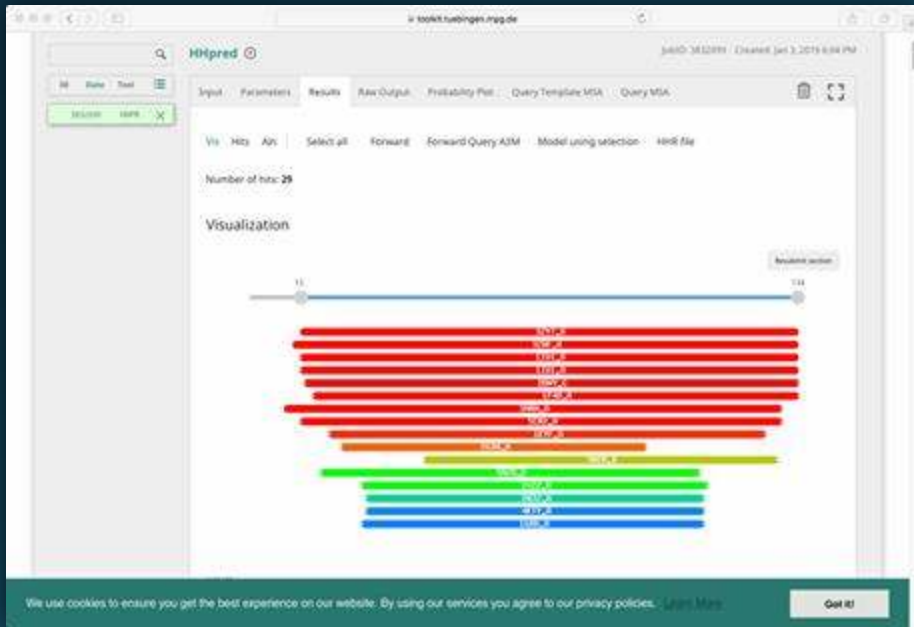
Examine the data

Find a search model

Preparation of search model

Preparing models for Molecular Replacement

- Ensembles: alignment of search models : several homologues available



Examine the data



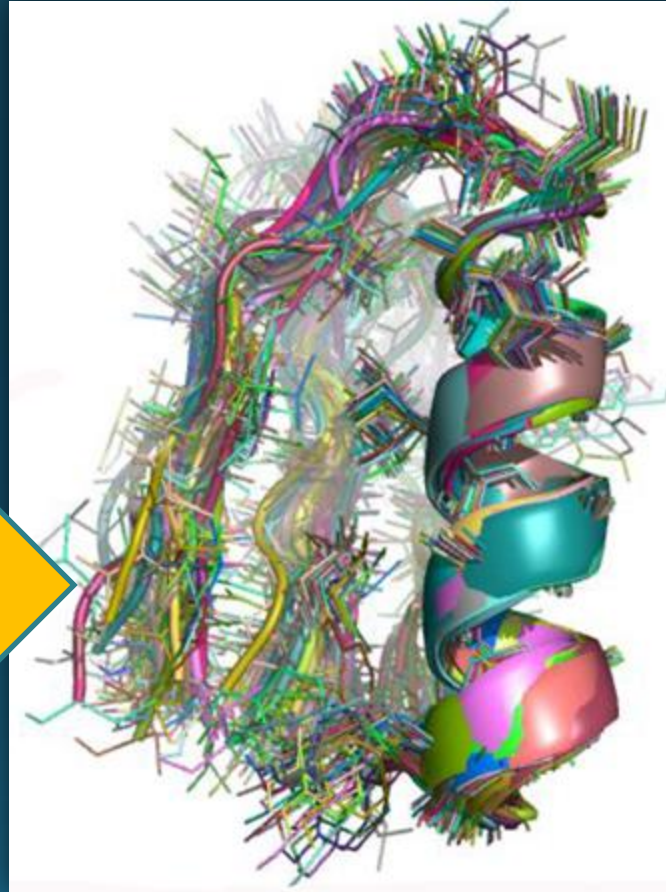
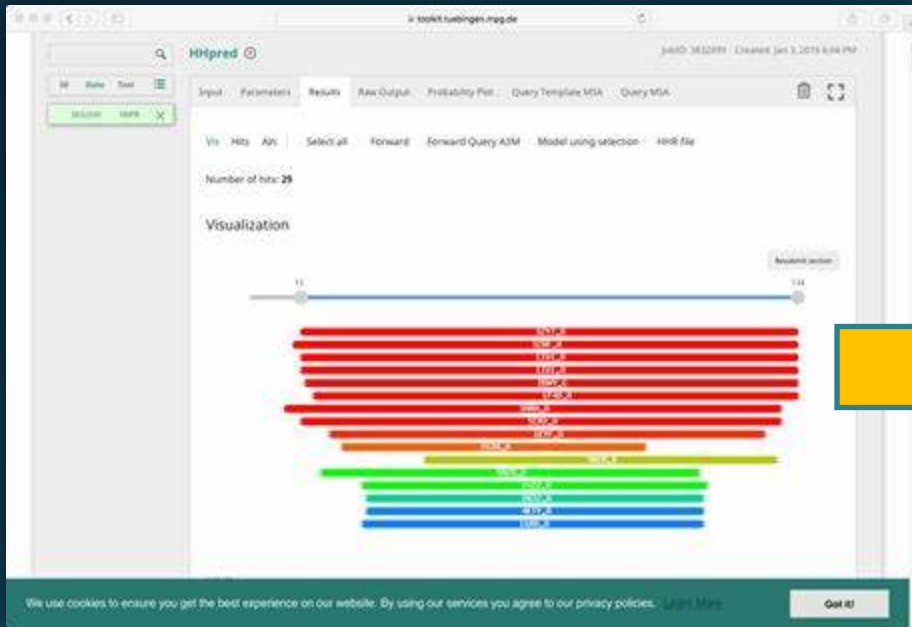
Find a search model



Preparation of search model

Preparing models for Molecular Replacement

- Ensembles: alignment of search models



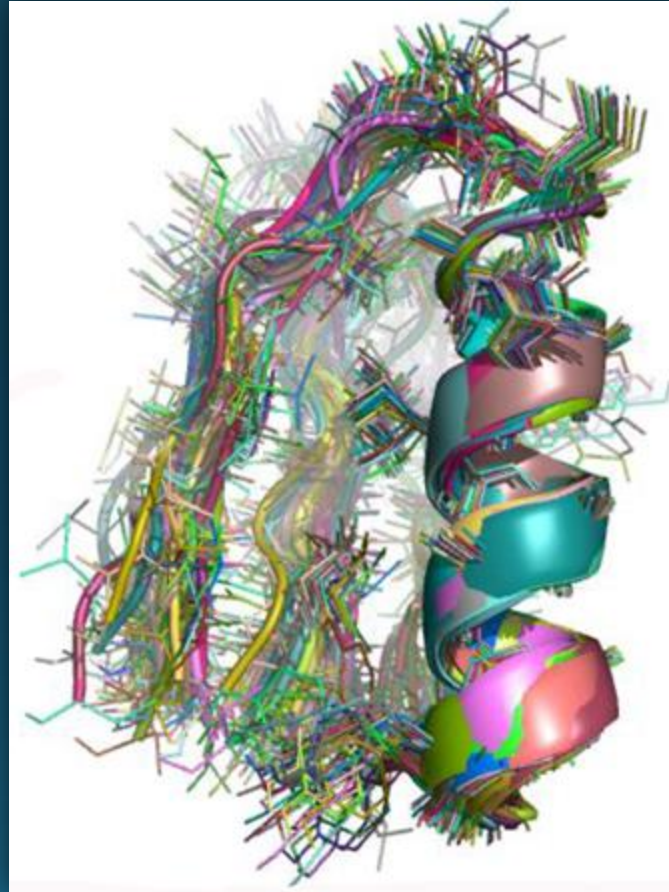
Examine the data

Find a search model

Preparation of search model

Preparing models for Molecular Replacement

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



Examine the data



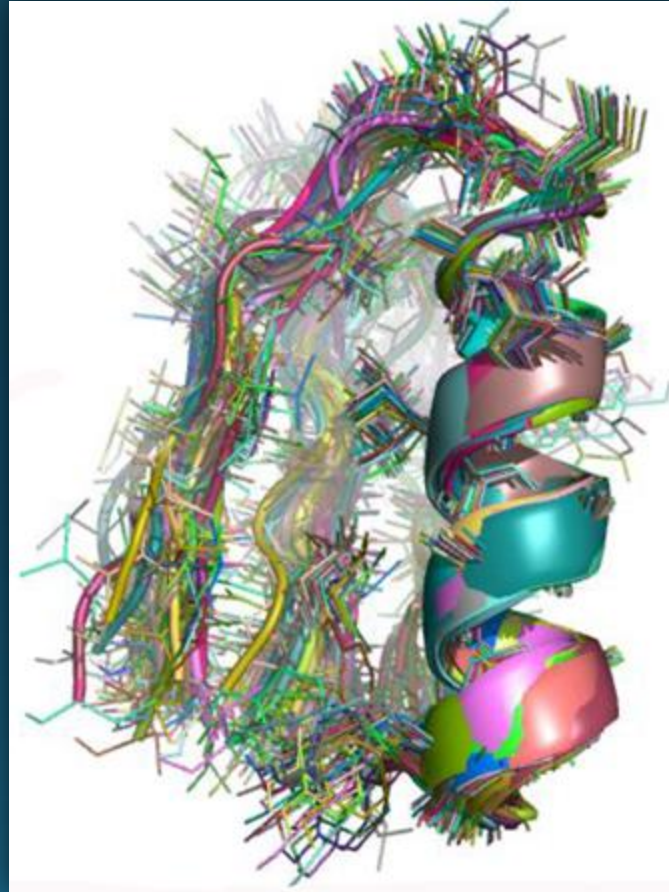
Find a search model



Preparation of search model

Preparing models for Molecular Replacement

- 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



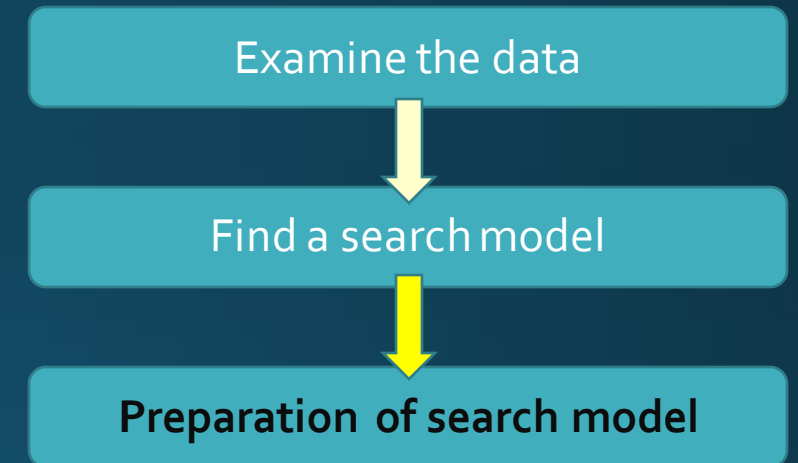
Find a search model



Preparation of search model

Preparing models for Molecular Replacement

- 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



Preparing models for Molecular Replacement

The screenshot displays the CCP4mg version 2.10.10 interface. The main window shows a 3D ribbon model of a protein structure, labeled 'A/35(ASP)/CA', with a scale bar indicating 15.3 Å. The text 'Ensemble truncation "slider"' is overlaid on the model. The 'Display Table' on the right lists objects like 'sculpt_1gxt_A1' and 'sculpt_1gxu_A1'. A 'Finished MrBUMP search' dialog box is open, showing a list of models (1gxt_A, 1gxu_A, 4g9i_G, etc.) and a 'Domain 1' slider. A blue arrow points from the 'Domain 1' slider to the 'Display models for' section, which is highlighted with a red dashed box. The 'Display models for' section shows 'Show residues with gesamt variance' and 'Show MrBUMP models' checked. The 'Sequence Viewer' at the bottom displays the protein sequence in PIR format.

Ensemble generation with CCP4mg-MrBUMP

Ensemble truncation "slider"

15.3 Å

Finished MrBUMP search

Domain 2 1

1gxt_A

1gxu_A

4g9i_G

4g9i_B

4g9i_E

4g9i_F

4g9i_D

4g9i_A

3vth_A

3vti_A

Display models for

Domain 1

Show residues with gesamt variance

110

✓ Show MrBUMP models

□ Show unscripted PDB models

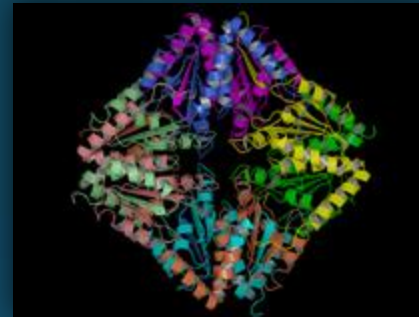
Close MrBUMP models Close Cancel

mp_1\logs\phmmmerALN.log from 3 mrbump_basic

12:50 24/05/2018

Preparing models for Molecular Replacement

- 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



Examine the data



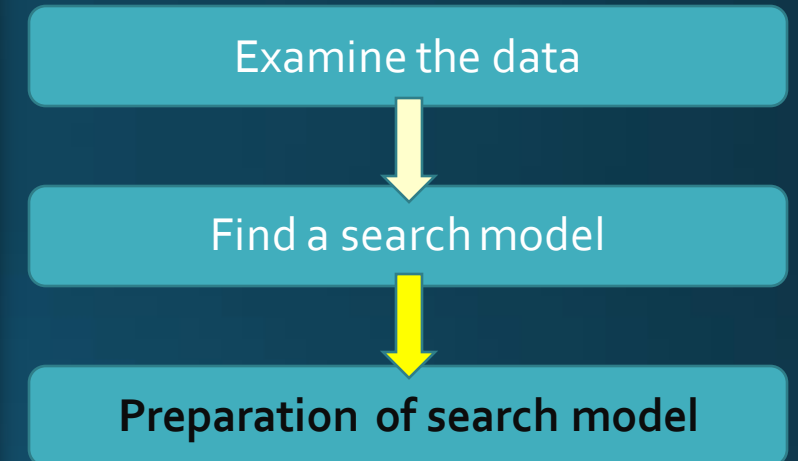
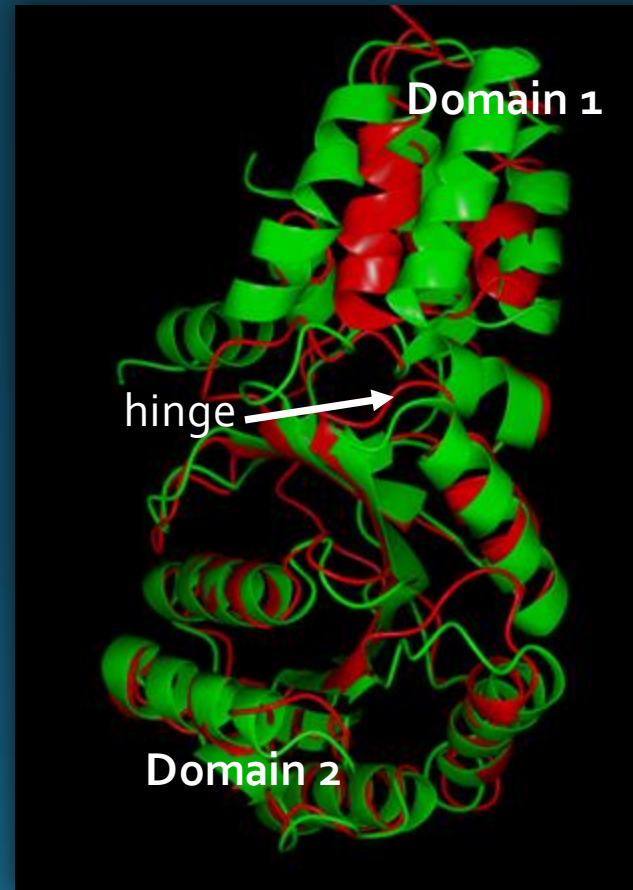
Find a search model



Preparation of search model

Preparing models for Molecular Replacement

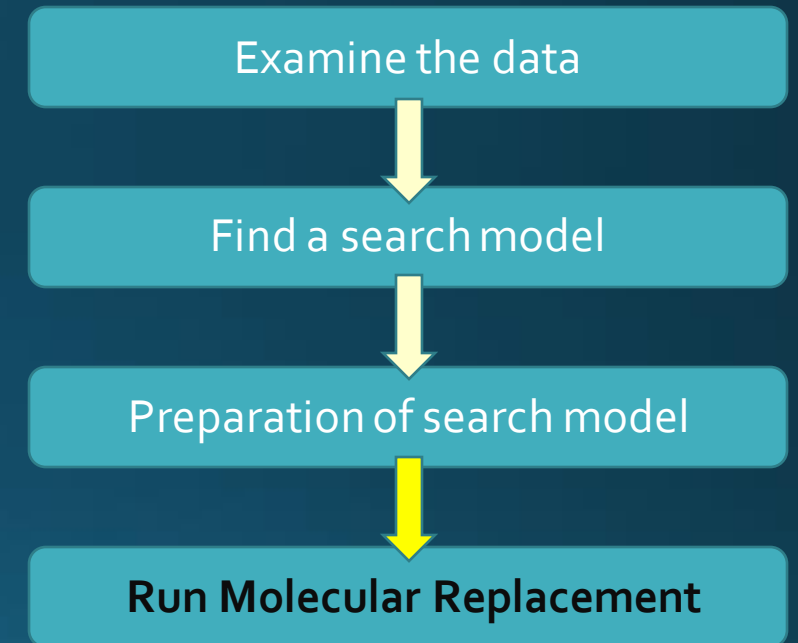
- Domains as search models
 - A hinge motion may alter the relative orientation of domains within a chain
 - Domain models should be isolated from the parent search model and used separately in MR



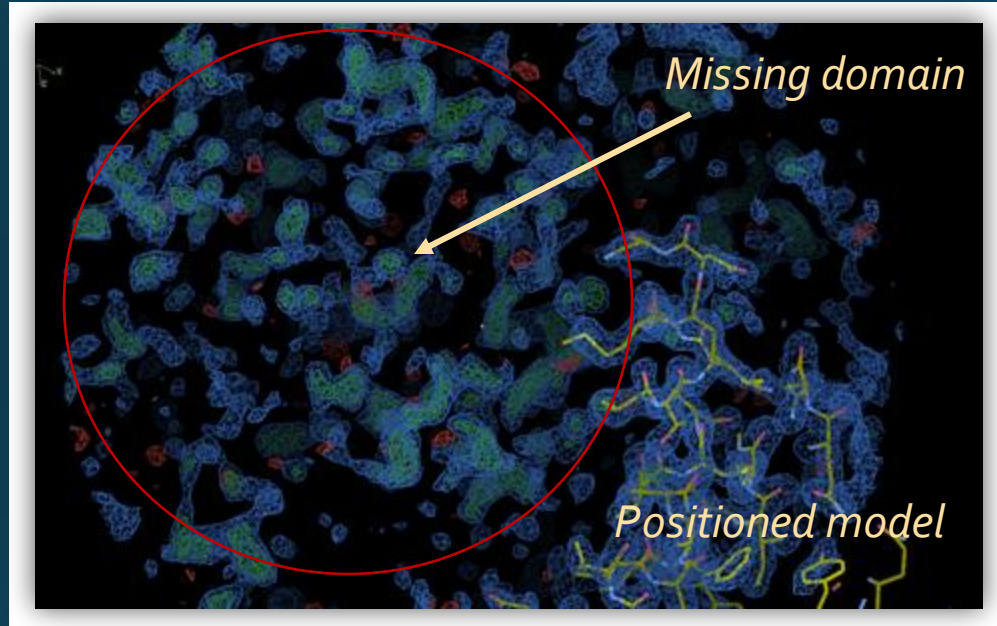
Running Molecular Replacement

Molecular Replacement

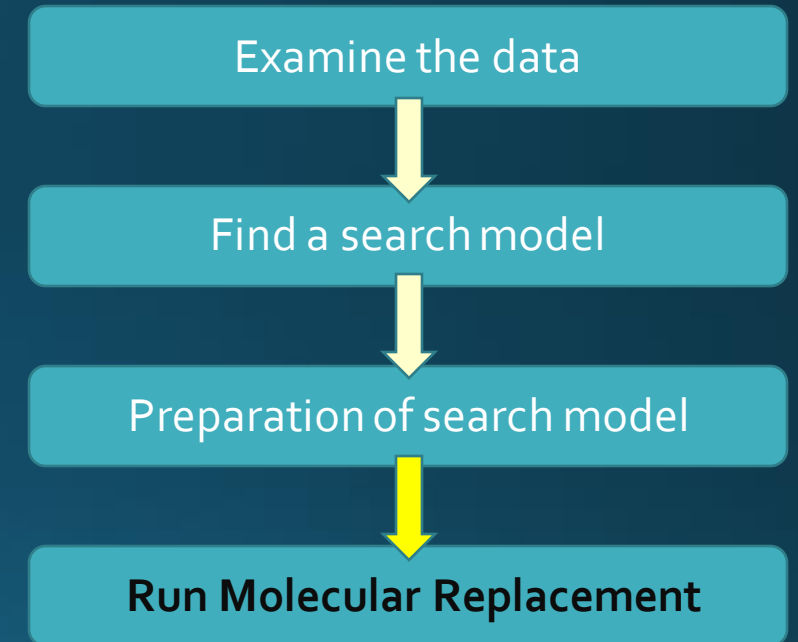
- CCP₄ has several programs for doing Molecular Replacement
 - Amore
 - Manual steps but very fast
 - Molrep
 - Automated MR
 - Several useful features e.g. searching a map
 - Phaser
 - Maximum likelihood approach
 - Accounts for potential model errors
 - Best for difficult cases and for correctly positioning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



Molecular Replacement



- Phased translation search
 - Useful when searching for several copies or dealing with a complex
 - Available through MOLREP (3 protocols) and Phaser
 - Can help in some cases, particularly when looking for small domains

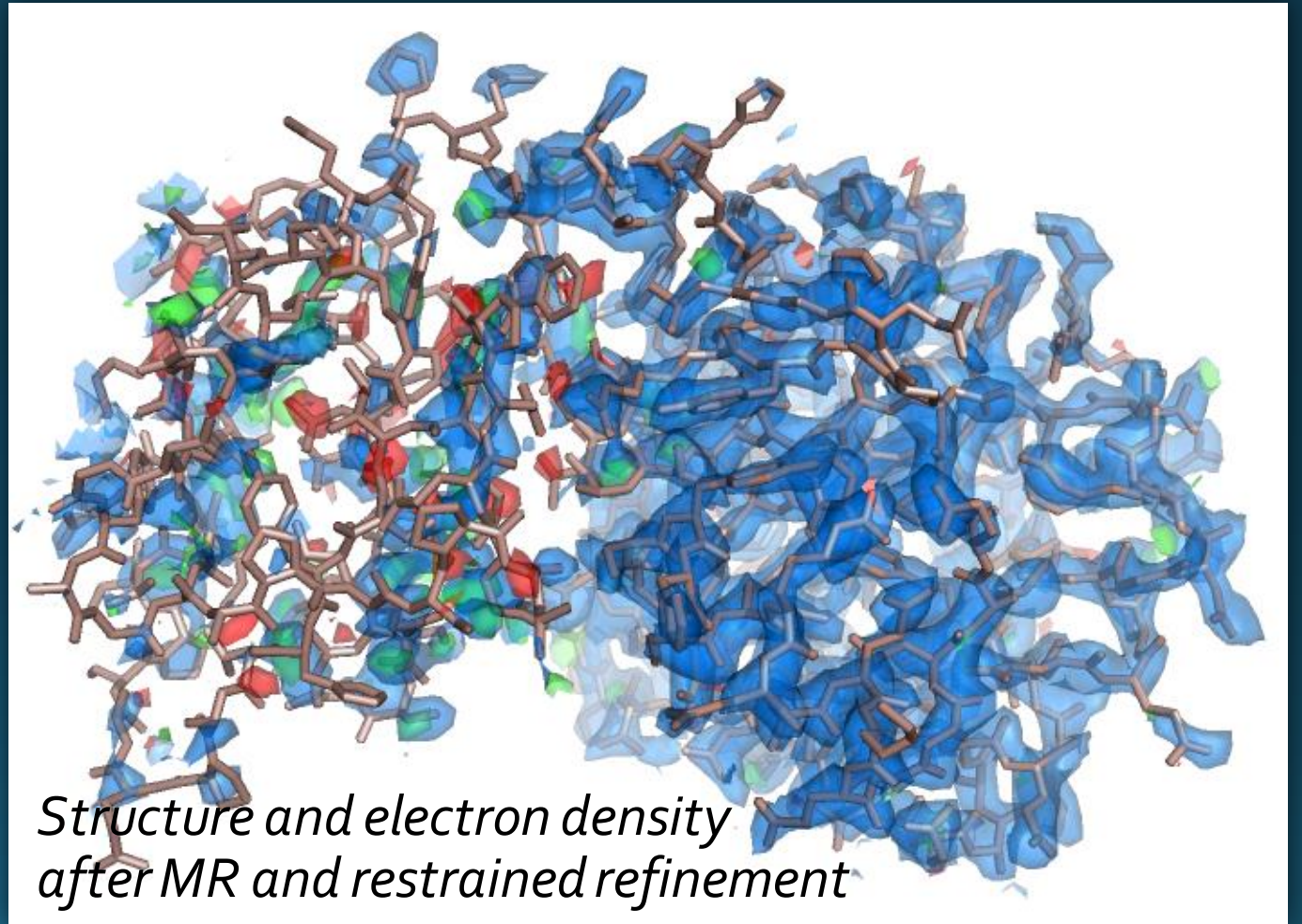


Molecular Replacement

Phased translation search:

Example: 1tj3

Search model: 1s20, chain A



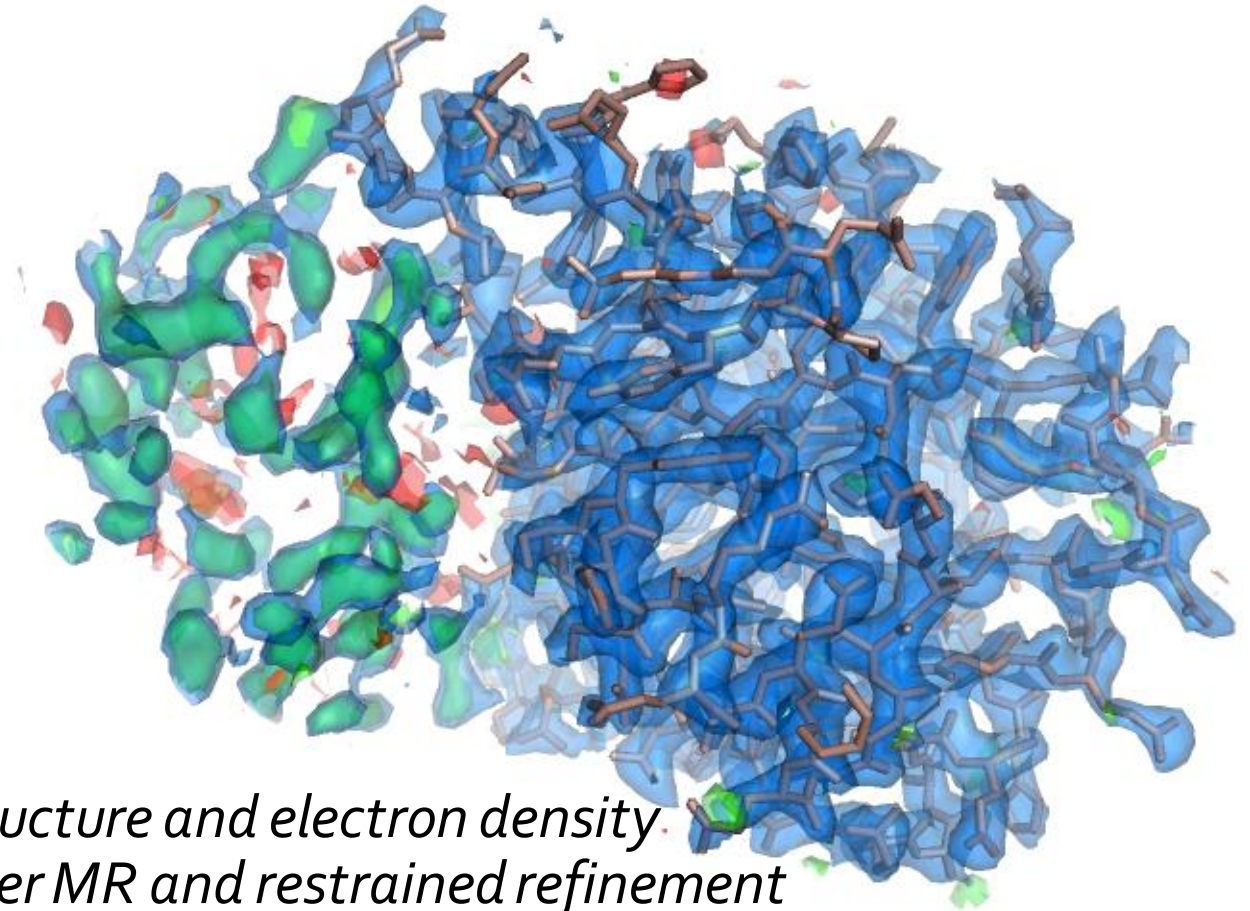
Molecular Replacement

Phased translation search:

Example: 1tj3

Search model: 1s20, chain A

Large Domain only



*Structure and electron density
after MR and restrained refinement*

Molecular Replacement

Phased translation search:

Example: 1tj3

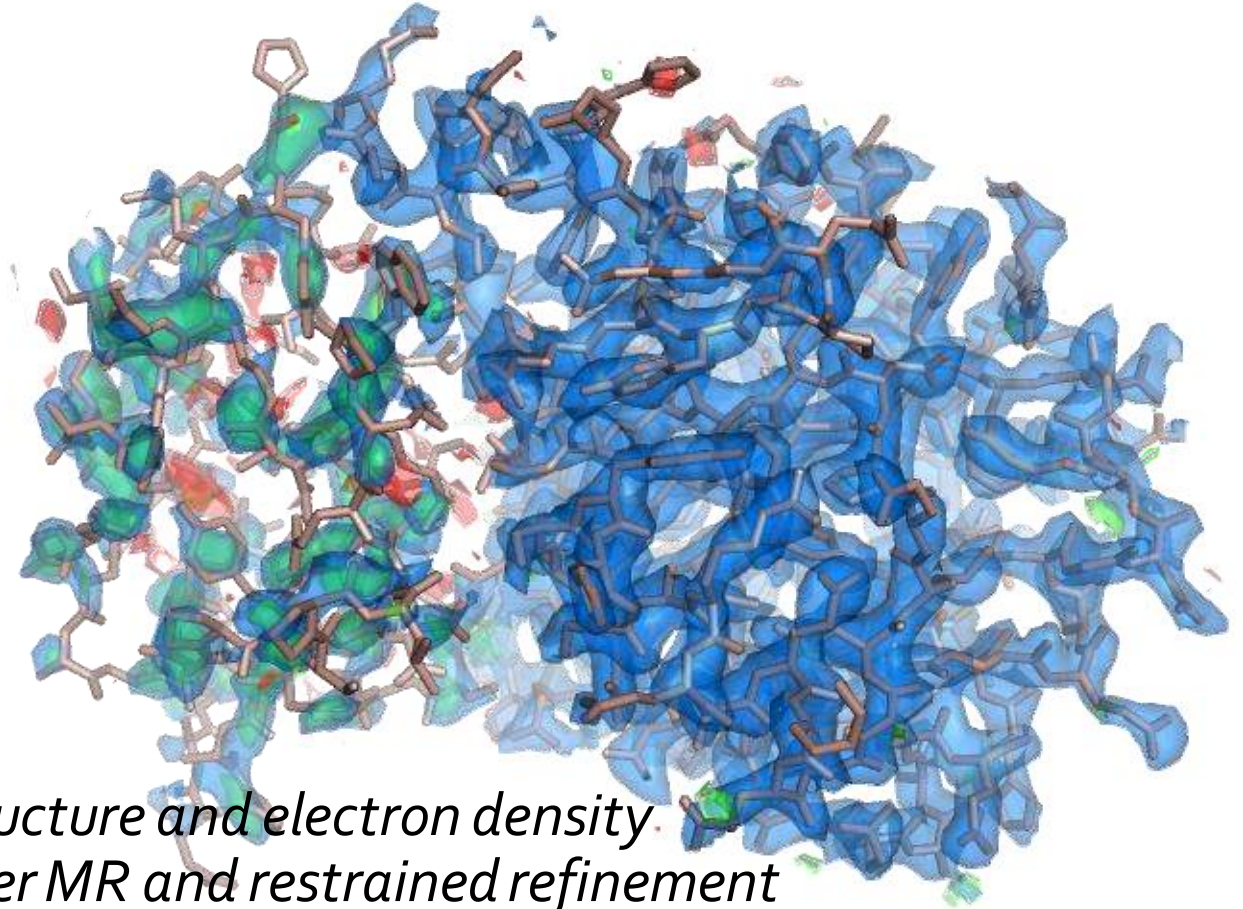
Search model: 1s20, chain A

Large Domain fixed

Small Domain search

Fit into density

(Molrep & Phaser)



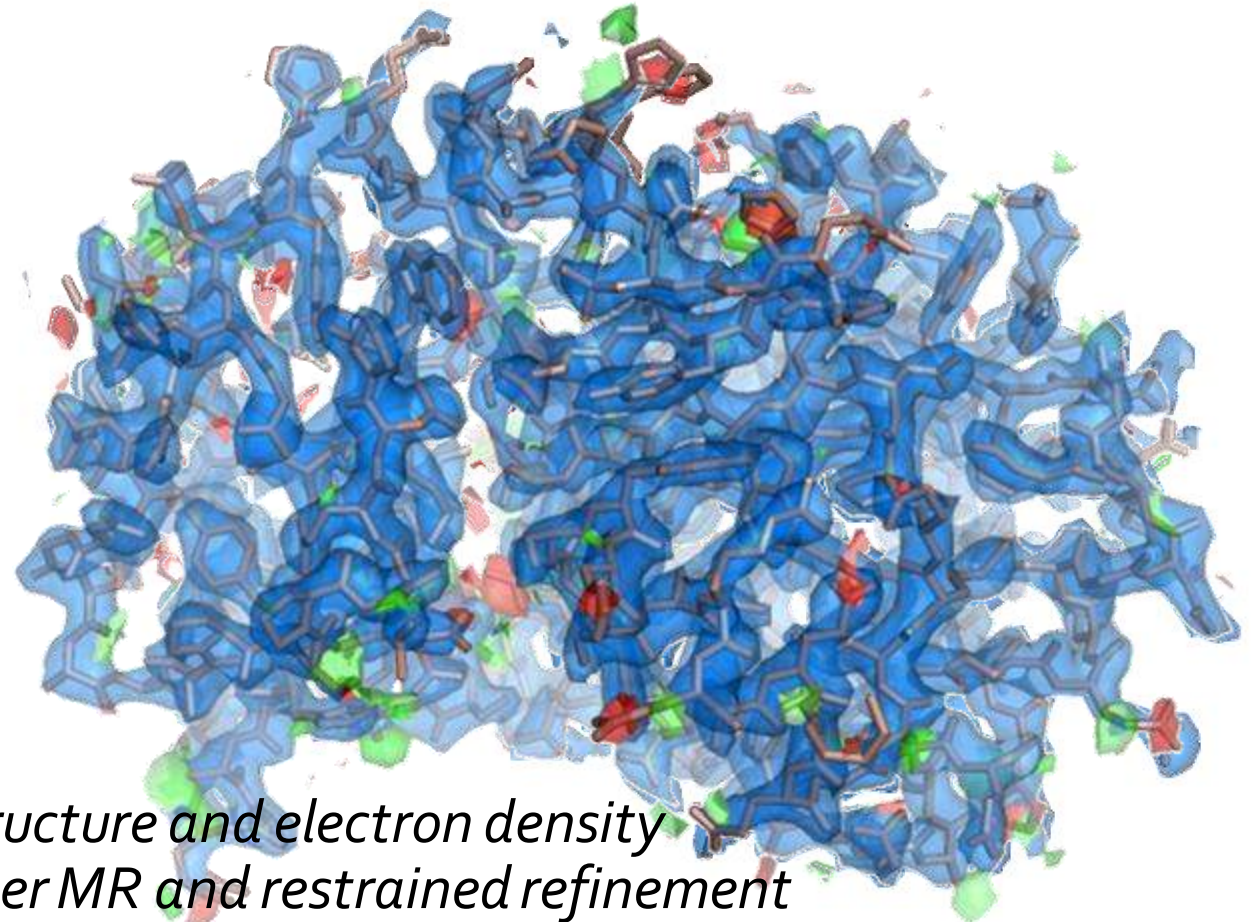
*Structure and electron density
after MR and restrained refinement*

Molecular Replacement

Phased translation search:

Example: 1tj3

Complete structure and
electron density after
restrained refinement

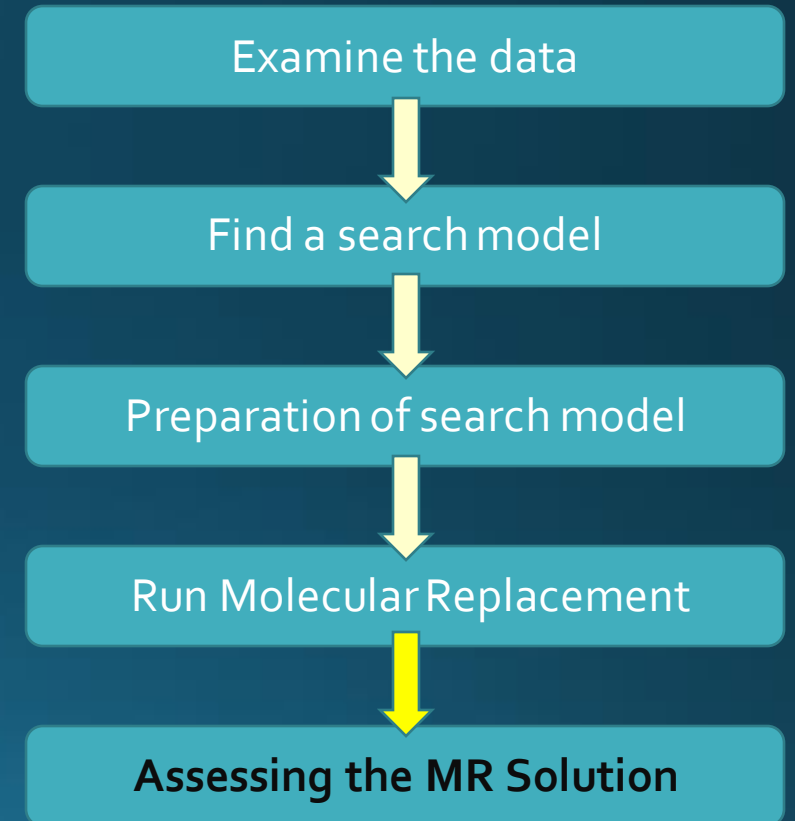


*Structure and electron density
after MR and restrained refinement*

How do I know my MR solution is successful?

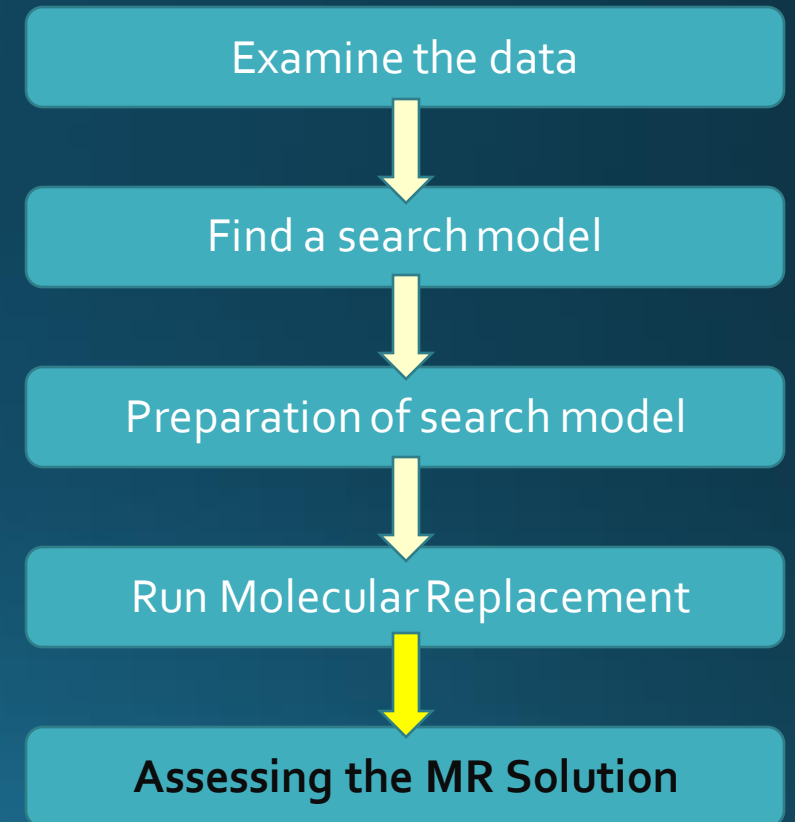
Assessing the solution

- 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

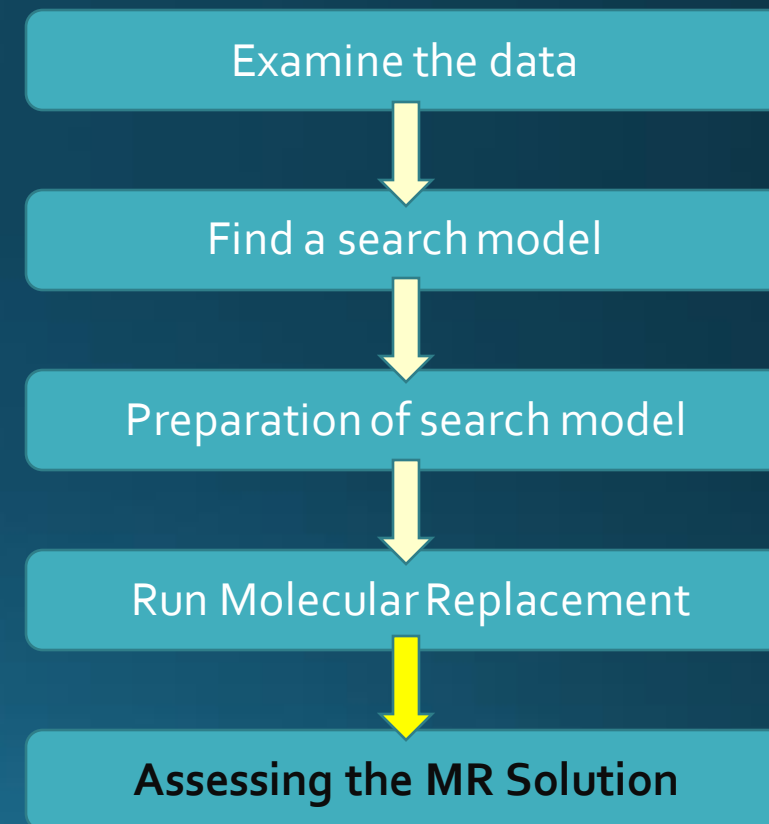
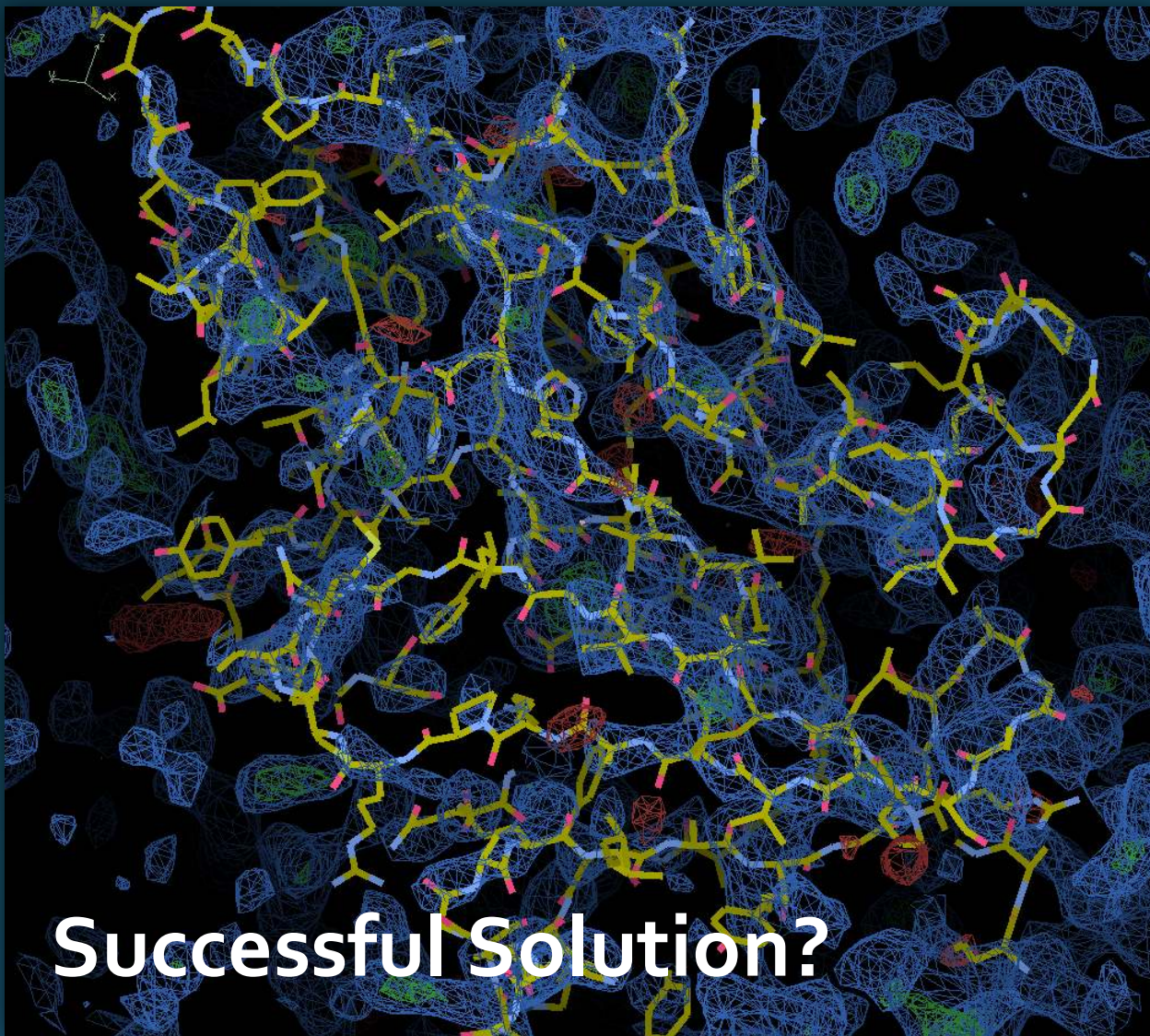


Assessing the solution

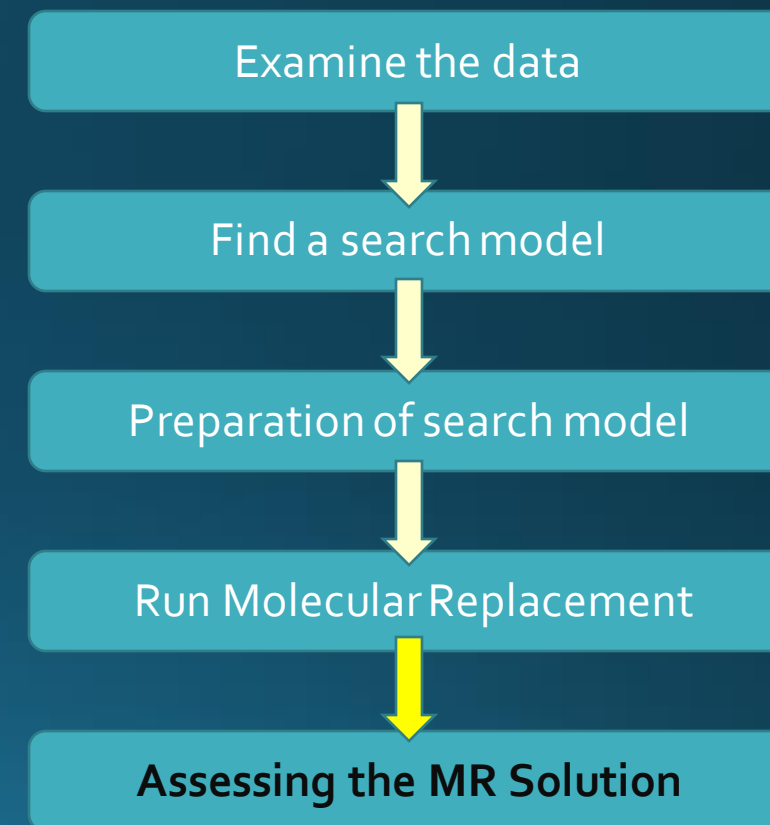
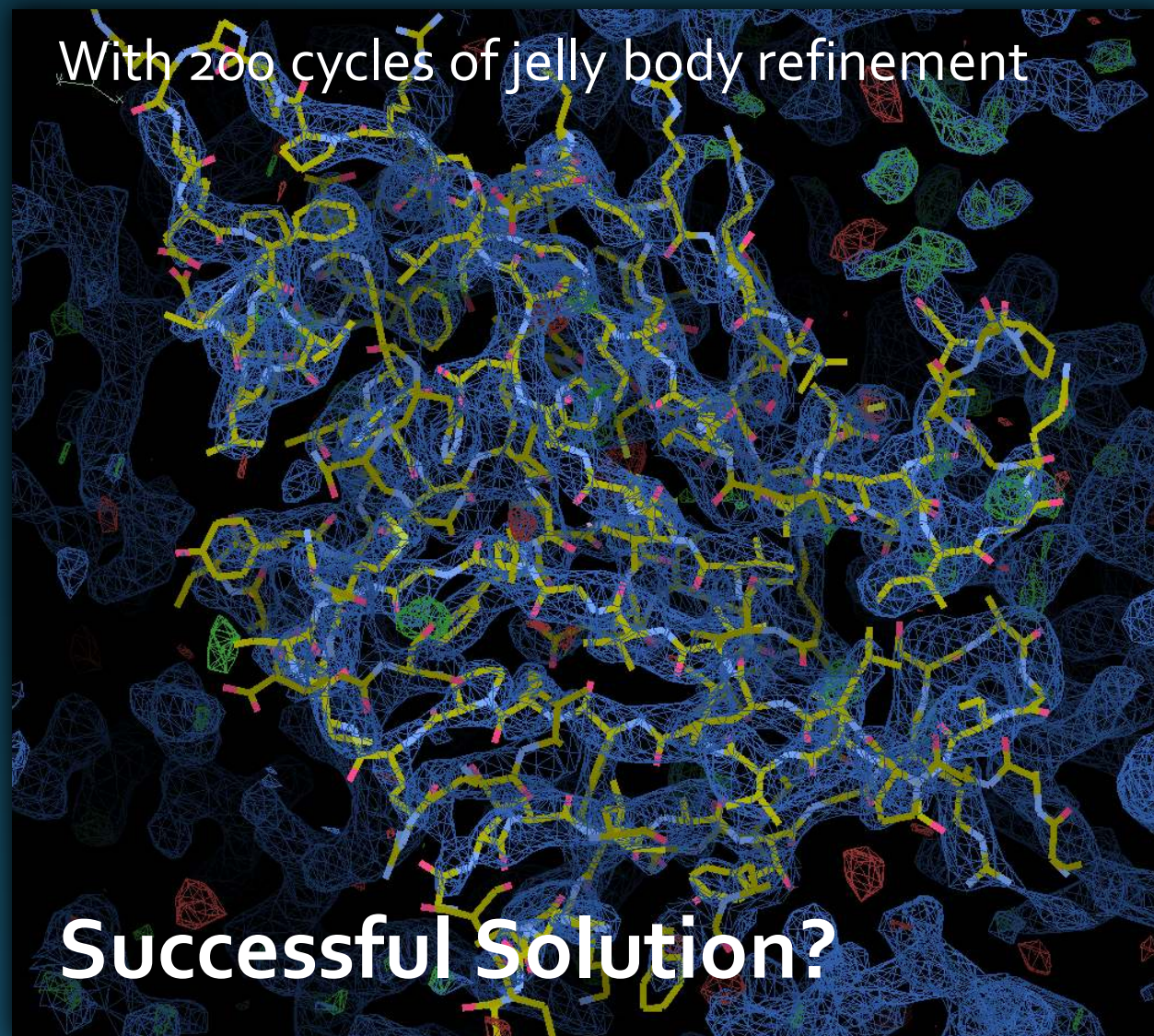
- In difficult cases the position may be correct but getting from the MR solution to a complete model may not be straight forward
- Assessment often involves performing additional structure solutions steps such as refinement and density modification



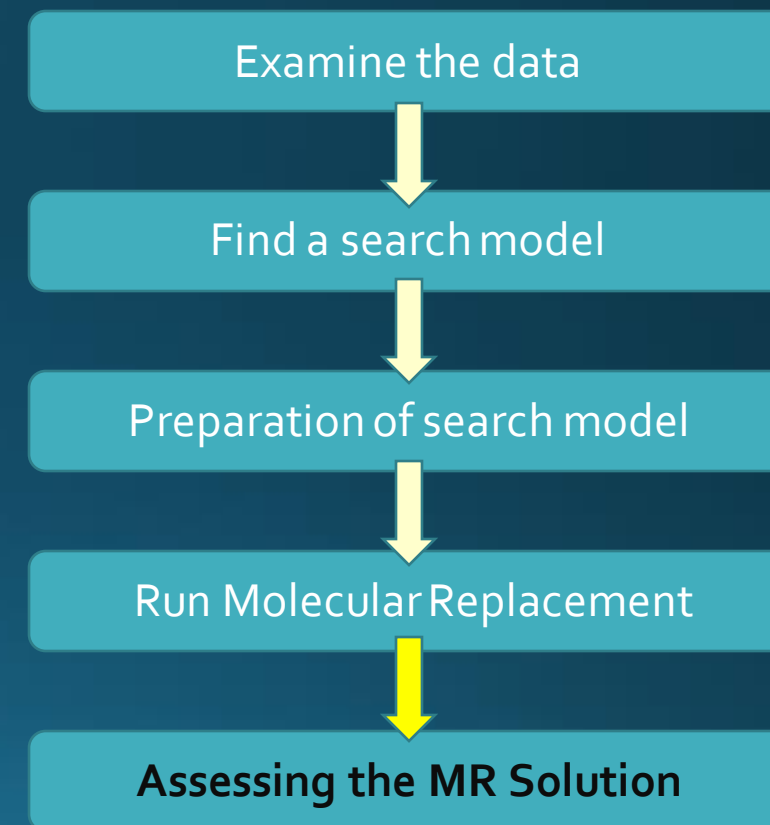
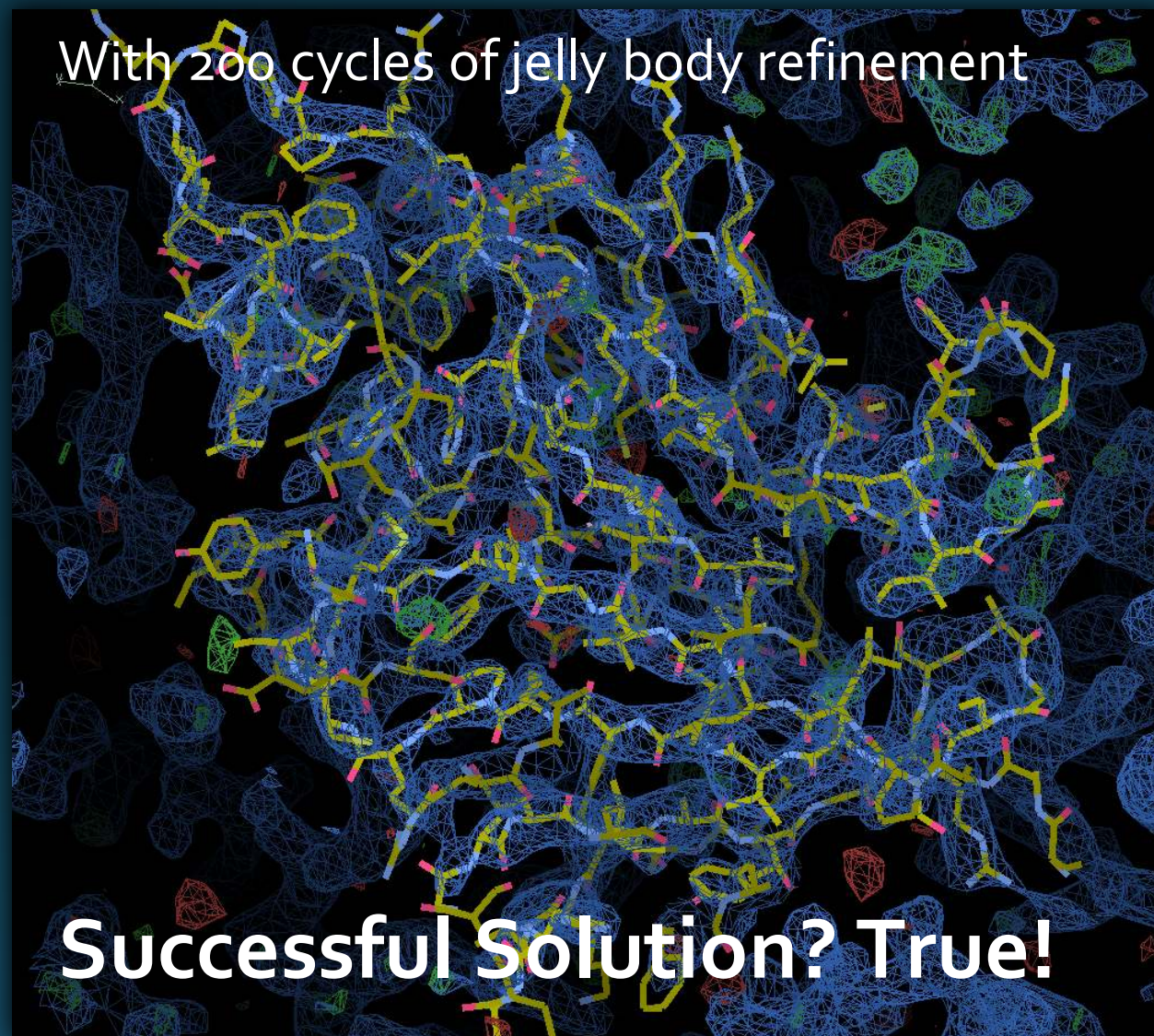
Assessing the solution



Assessing the solution

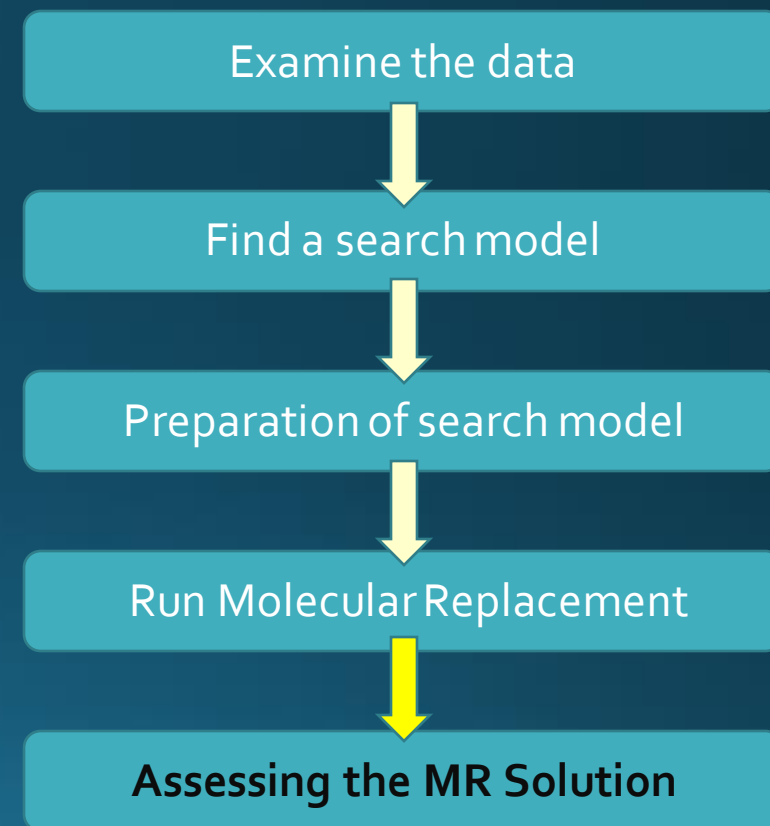


Assessing the solution



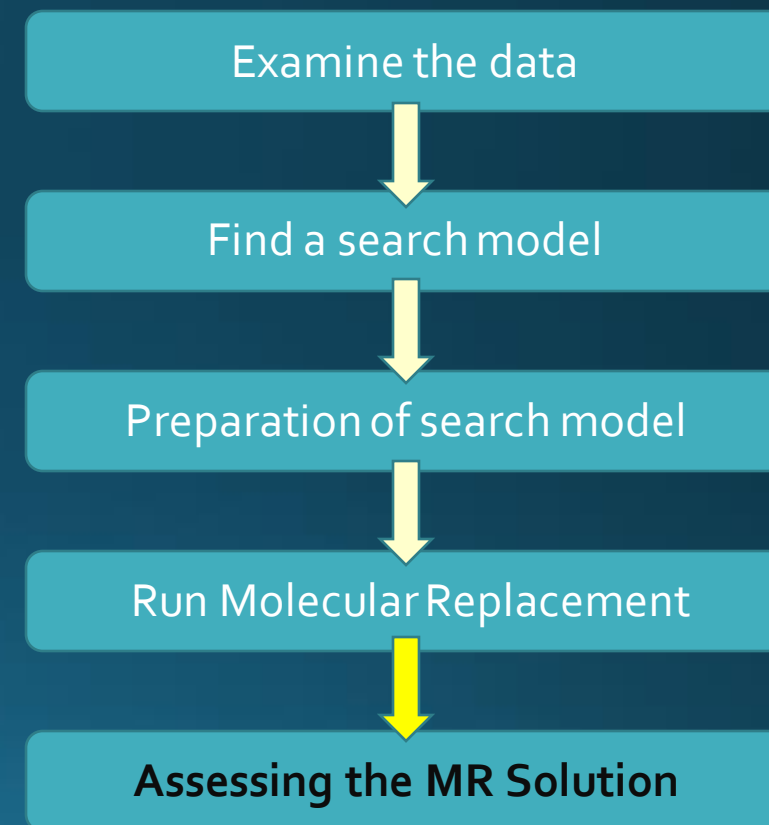
Assessing the solution

- Rough guide to MR program scoring



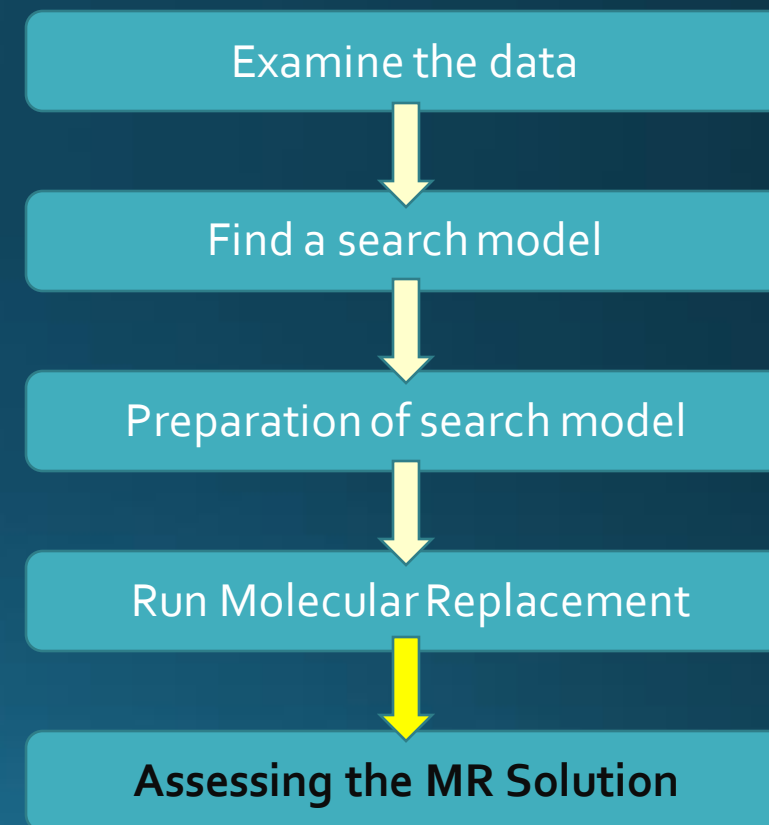
Assessing the solution

- 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



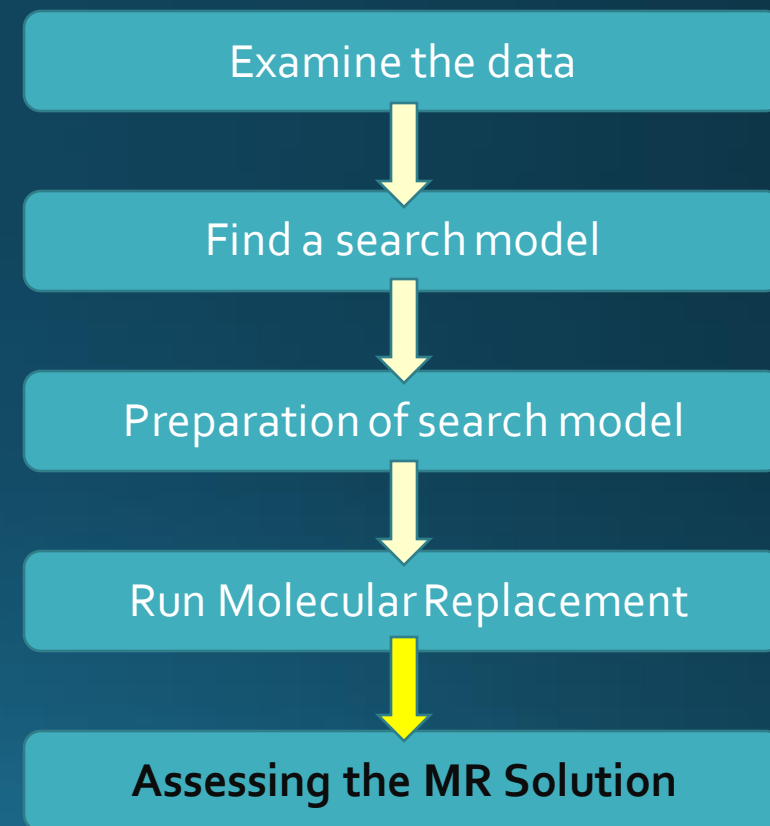
Assessing the solution

- 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?



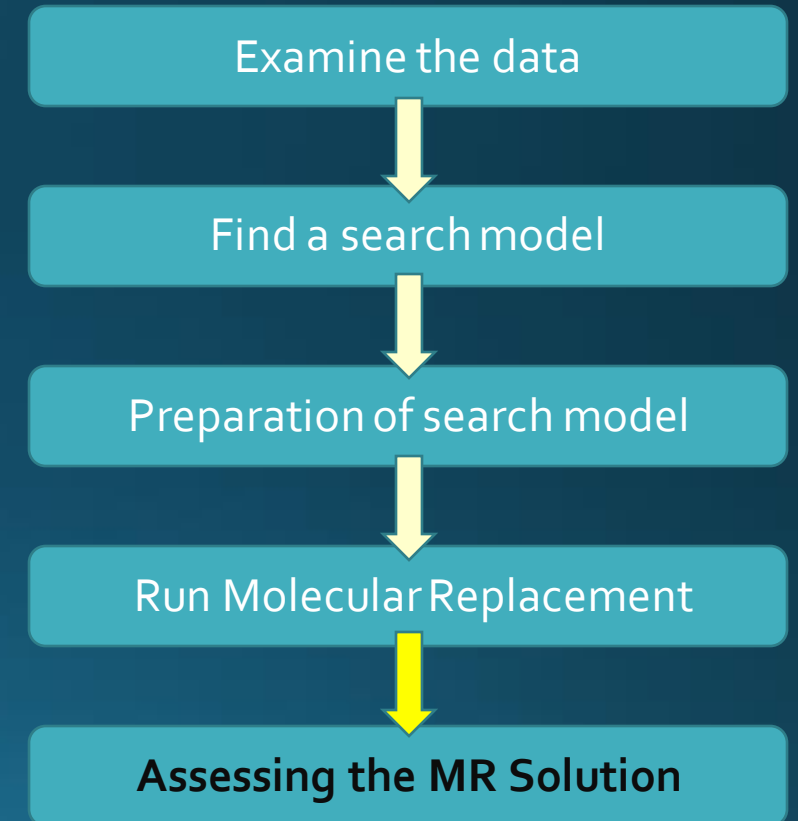
Assessing the solution

- Refinement



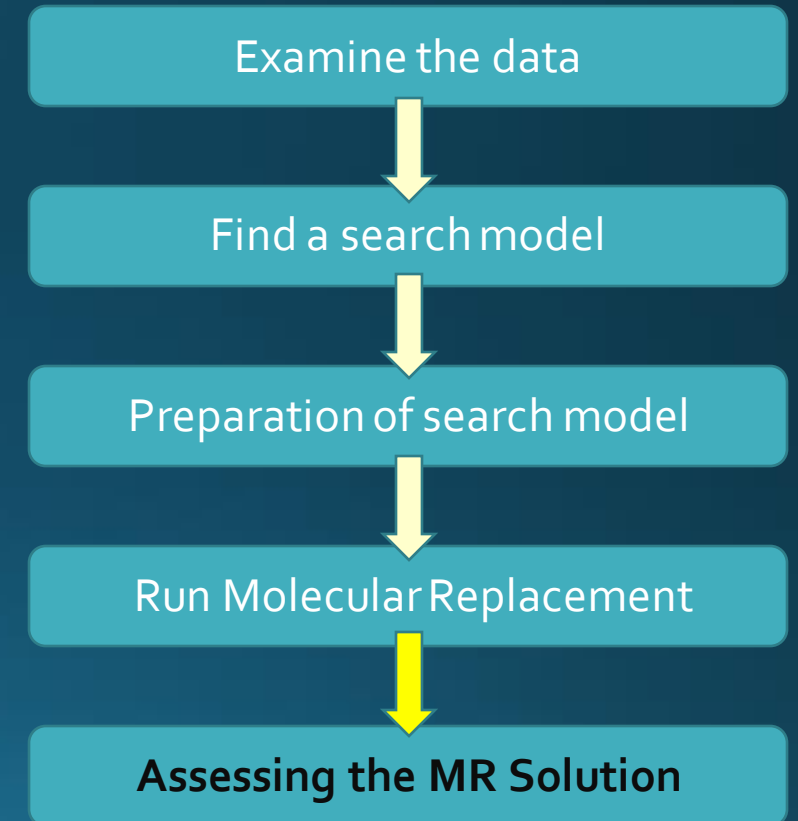
Assessing the solution

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



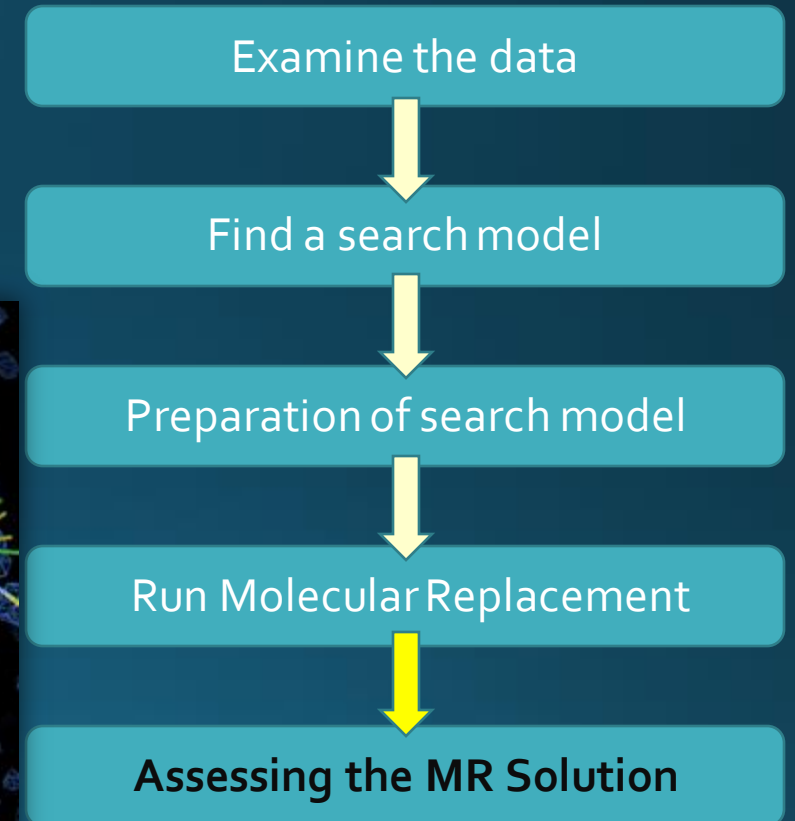
Assessing the solution

- 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



Assessing the 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



Phaser positioned model

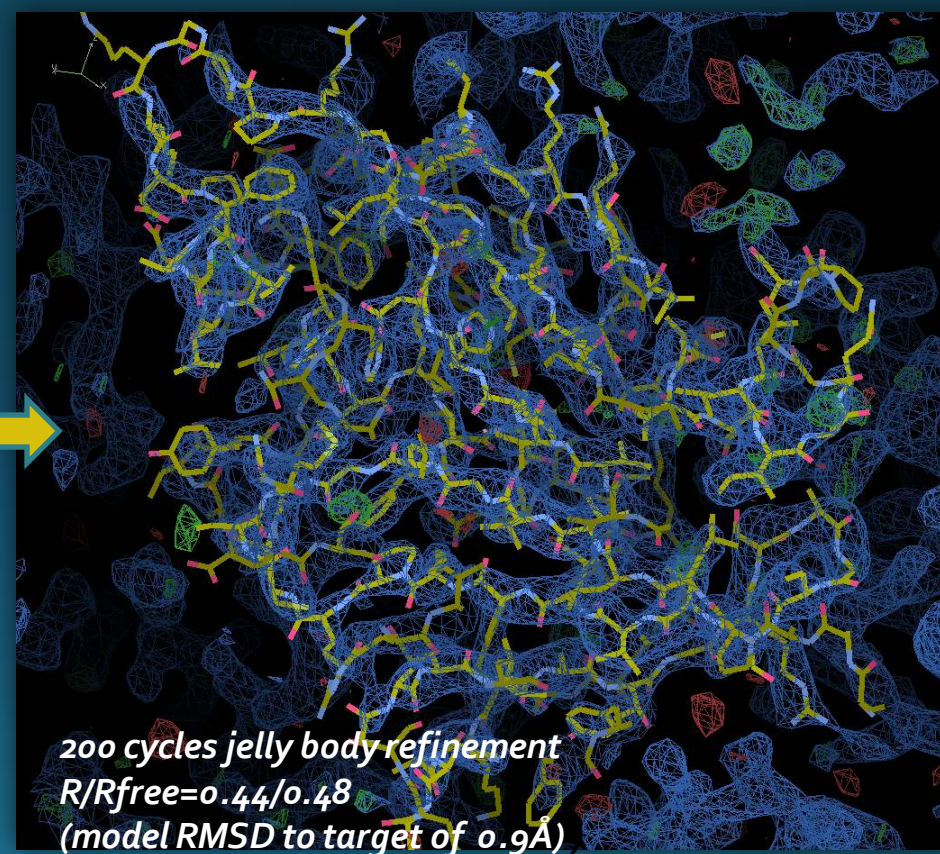
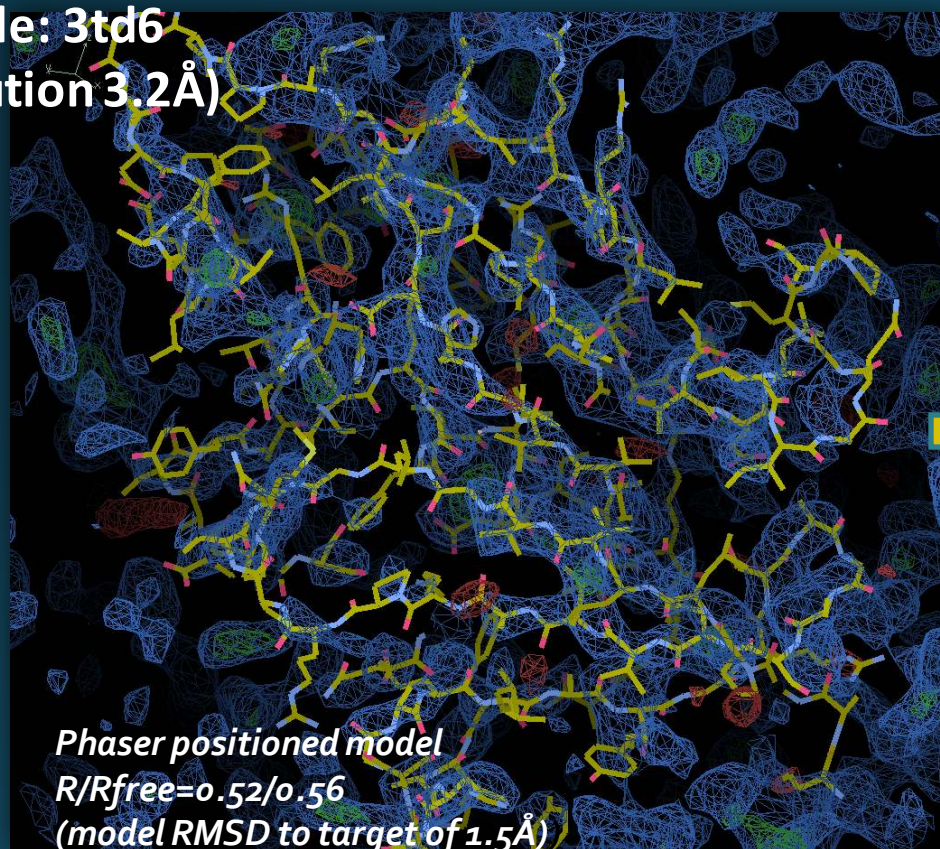
200 cycles jelly body refinement



Assessing the solution

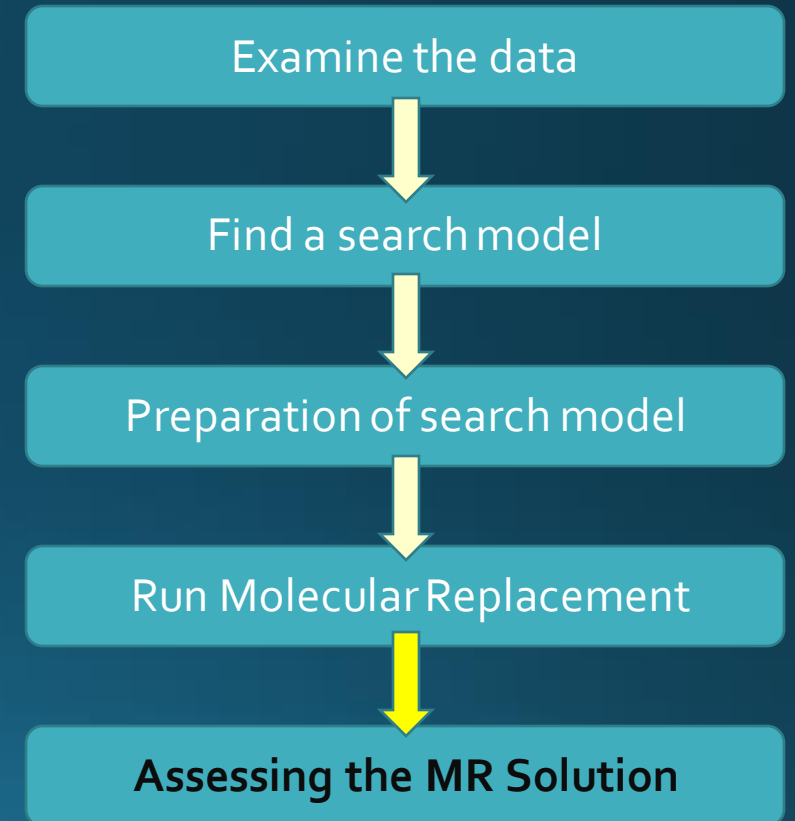
- Refinement
 - Use 200 cycles of jelly body refinement option in Refmac post MR

Example: 3td6
(resolution 3.2Å)

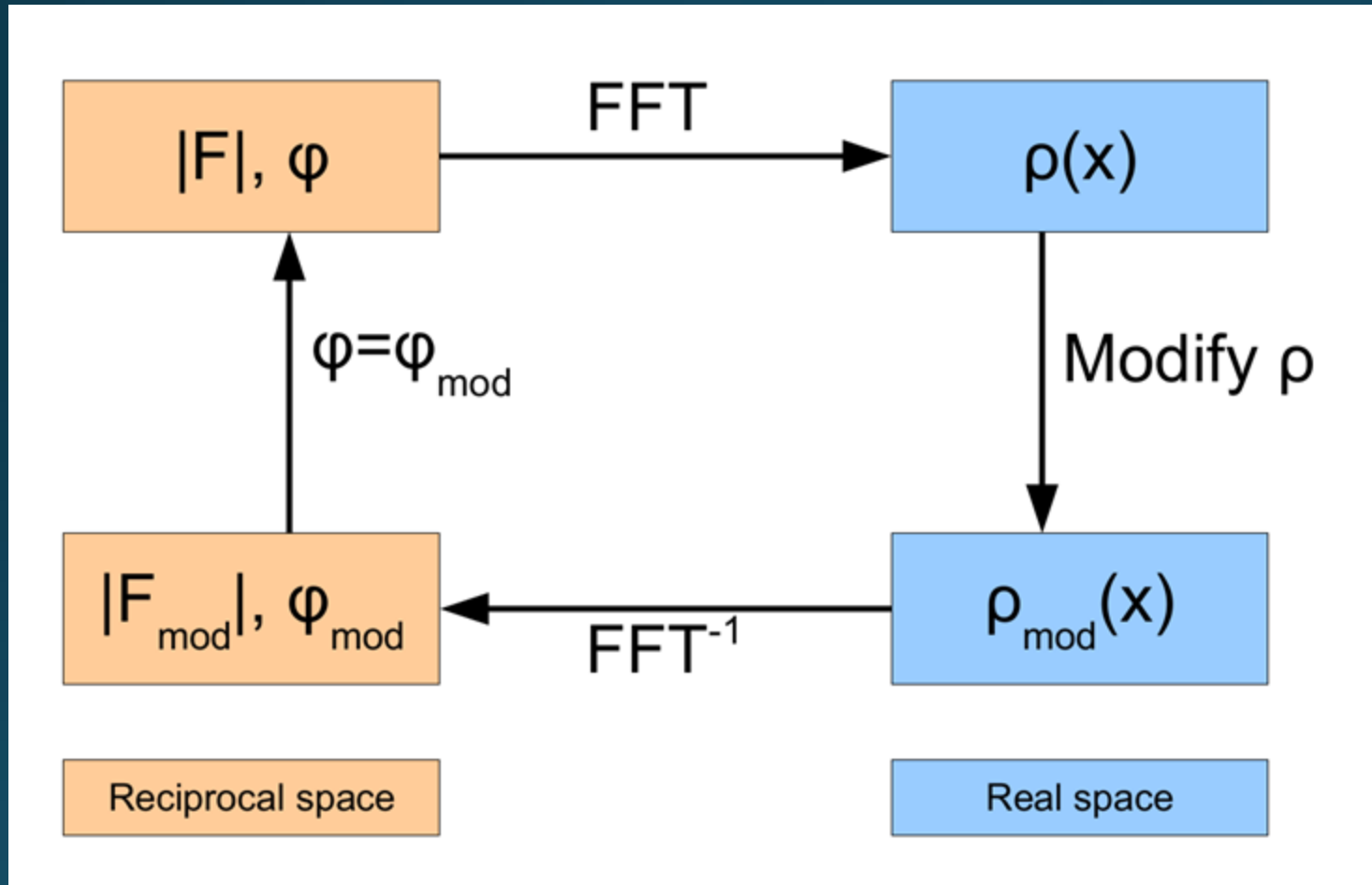


Assessing the solution

- Density Modification & C-alpha tracing



Density Modification

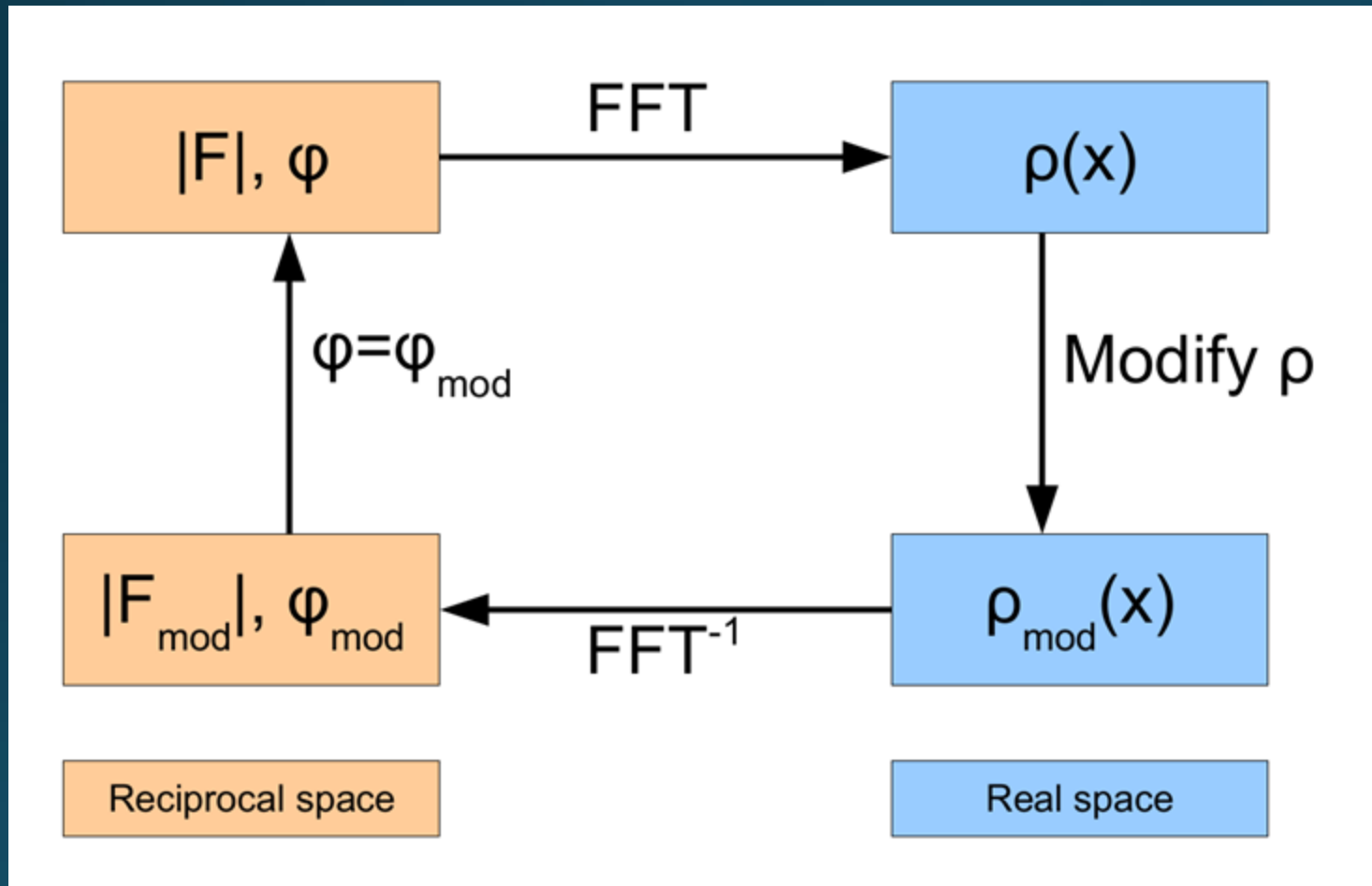


Main techniques:

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

(slide from Kevin Cowtan)

Density Modification



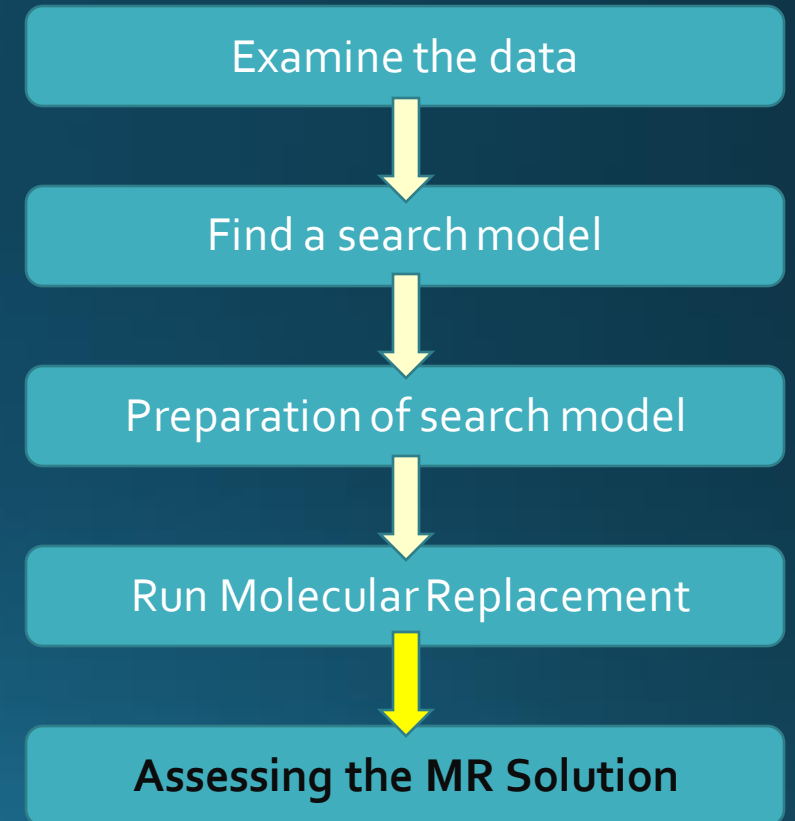
CCP4 Applications:

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

(slide from Kevin Cowtan)

Assessing the 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

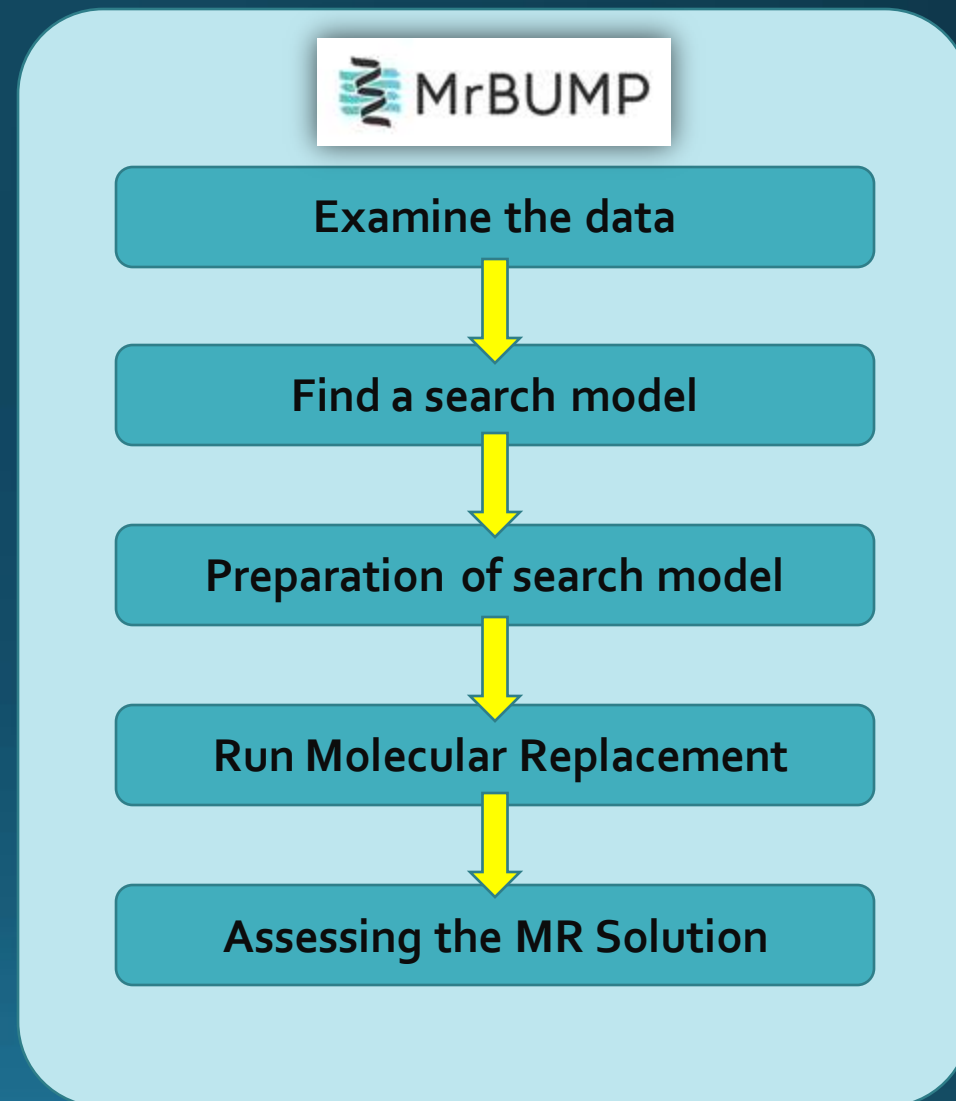


Automated Molecular Replacement in CCP4

- 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, 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



Examine the data

Find a search model

Preparation of search model

Run Molecular Replacement

Assessing the MR Solution

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

The screenshot displays the CCP4 Online web interface. The main window shows the 'New MrBUMP Run' page, which includes a header with the CCP4 logo and navigation links. Below the header, there is a section for 'Programs' with a table listing various tools like Balbes, MrBUMP, Zanuda, and CRANK2. The 'New MrBUMP Run' form includes fields for 'Structure Factors' and 'Sequence Target', both with 'Choose File' buttons. A note indicates that users can paste their FASTA sequence instead of uploading a file. A 'Submit' button is at the bottom of the form. To the right, a table displays results for 'Monomer 1' and 'Monomer 2', including protein names, X,Y,Z coordinates, and various interface parameters. An 'Interaction radar' chart is also shown, along with a list of 'Interface parameters' such as IA, DG, BE, PV, HB, SB, and DS.

Monomer 1		Monomer 2	
	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

Interface parameters

Parameter	Value
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

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



MR on the CCP4 Cloud

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

The screenshot displays the CCP4 Cloud web interface. The main panel shows a project titled 'venom2019' with a list of processing steps:

- [adamVenom2019] venom2019
- [0001] imported: HKL (1) Sequence (1) XYZ (1) -- done.
- [0002] asymmetric unit contents -- Solv=54.7%
- [0003] ensemble preparation (xyz) -- done.
- [0004] phaser MR -- $N_{sol}=2$ LLG=1689 TFZ=3.9 R=0.4618 $R_{free}=0.462$
- [0005] phaser MR -- $N_{sol}=1$ LLG=2157 TFZ=38.8 R=0.4595
- [0006] molrep -- R=0.4945 $R_{free}=0.5254$
- [0007] change space group (ASU) -- SpG=P 21 21 2
- [0008] asymmetric unit contents -- Solv=54.7%
- [0009] ensemble preparation (xyz) -- done.
- [0010] molrep -- R=0.4850 $R_{free}=0.4855$
- [0012] refmac5 -- R=0.4069 $R_{free}=0.4186$
- [0013] shelxe -- CC=25.85% FOM=0.65
- [0037] refmac5 -- R=0.4581 $R_{free}=0.48$
- [0015] zanuda -- done.
- [0017] ensemble preparation (xyz) -- d
- [0018] ensemble preparation (xyz) -- d
- [0011] phaser MR -- $N_{sol}=2$ LLG=1606 TFZ=1.5 R=0
- [0016] refmac5 -- R=0.4215 $R_{free}=0.4372$
- [0014] phaser MR -- $N_{sol}=1$ LLG=2157 TFZ=34.2 R=
- [0020] +(0017) phaser MR -- $N_{sol}=1$ LLG=2228 TFZ=41.
- [0022] refmac5 -- R=0.4515 $R_{free}=0.4632$

On the right, a panel titled '[0005] phaser MR' shows 'Job completed' and buttons for 'Input', 'Output', 'Export', 'Ref.', and 'Help'. Below it, a panel titled '0005-02_phaser-mr Structure and electron density' displays a 3D molecular model with a map level of 0.3609 e/Å³ (1.56 rmsd). The bottom of the interface includes logos for CCP4 on-line, Science & Technology Facilities Council, and DBSRC, along with the version 'CCP4 Cloud v.1.4.001 [07.10.2019]'.

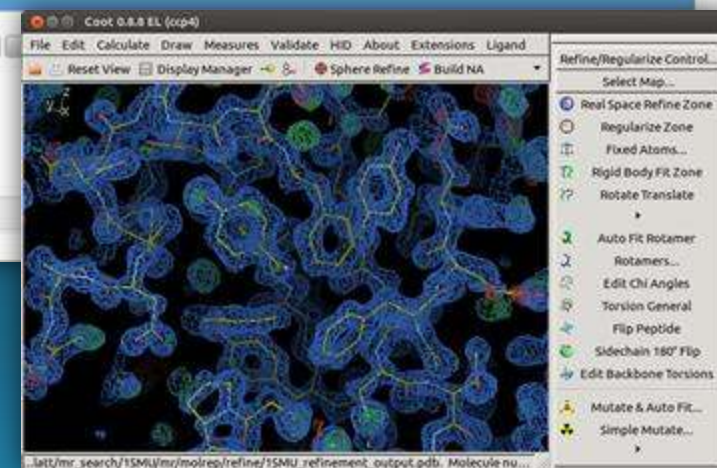
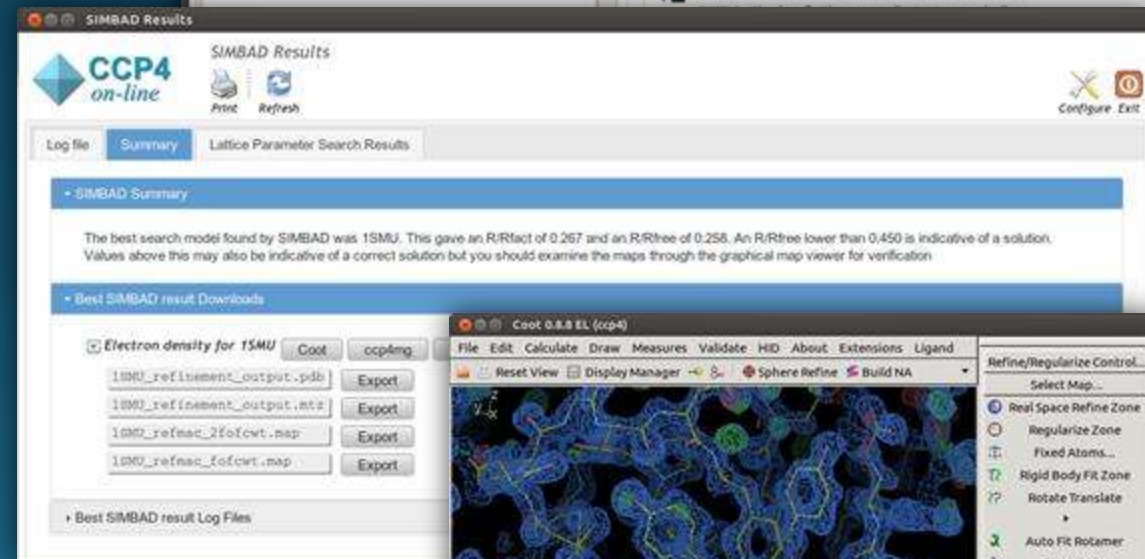
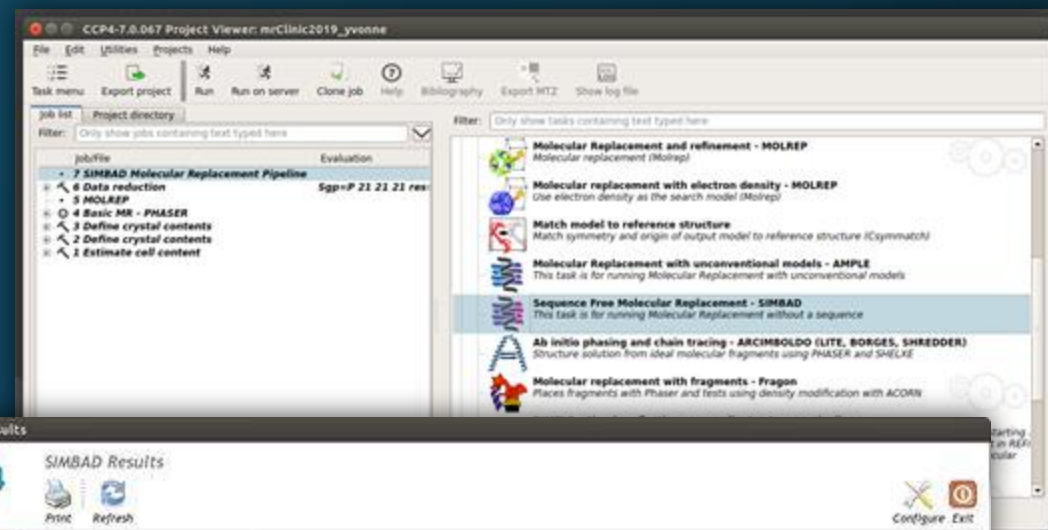
Crystal Contamination

- **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



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***
- 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