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*

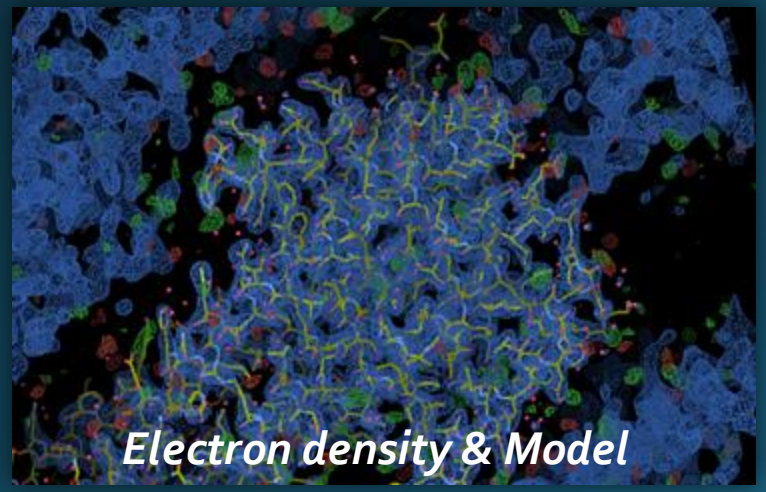
$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl}|F_o(hkl)|\cos[-2\pi(hx+ky+lz)+\varphi_{hkl}]$$

CCP4

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)



Electron density & Model

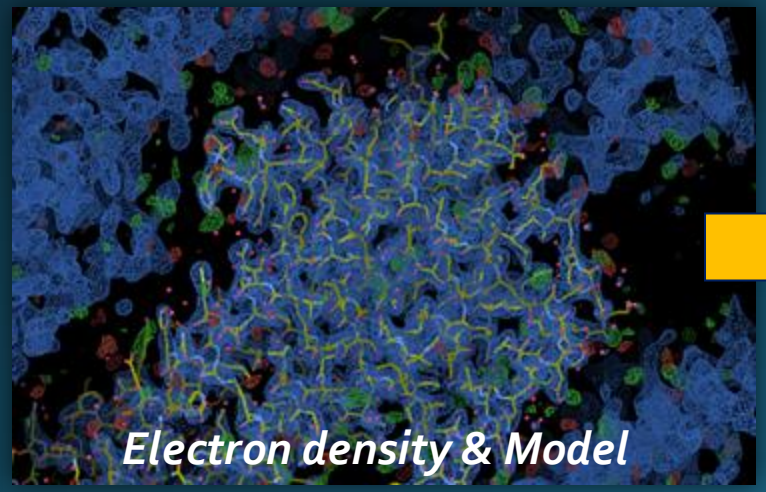
CCP4

Introduction: *Basics of Molecular Replacement*

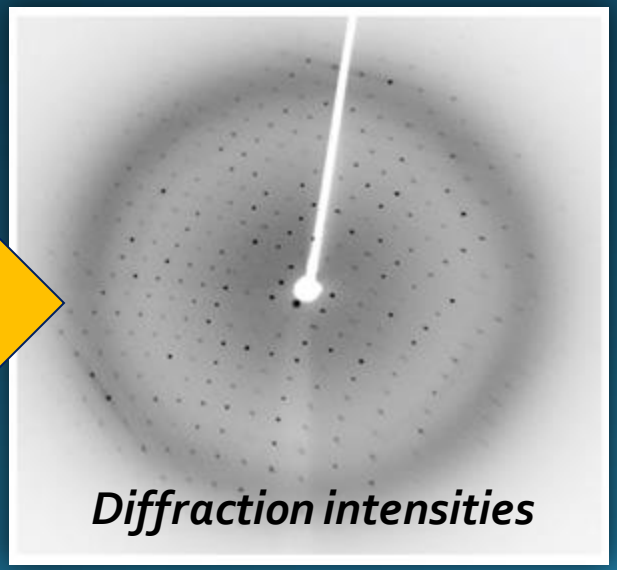
$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |F_o(hkl)| \cos[-2\pi(hx + ky + lz) + \varphi_{hkl}]$$

Electron density at (x,y,z)

Amplitudes



Electron density & Model



Diffraction intensities

CCP4

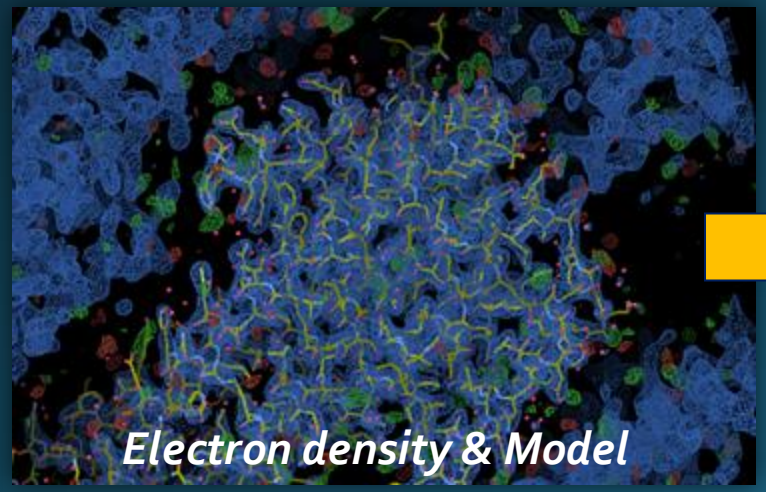
Introduction: *Basics of Molecular Replacement*

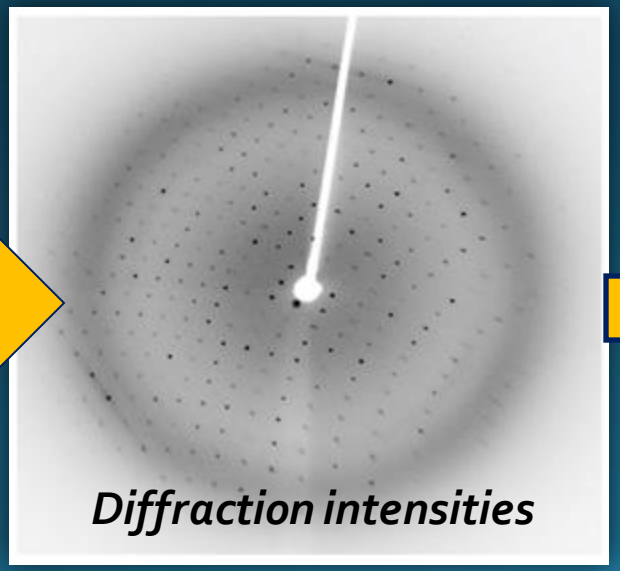
$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |F_o(hkl)| \cos[-2\pi(hx + ky + lz) + \varphi_{hkl}]$$

Electron density at (x,y,z)

Amplitudes

Phases





?

CCP4

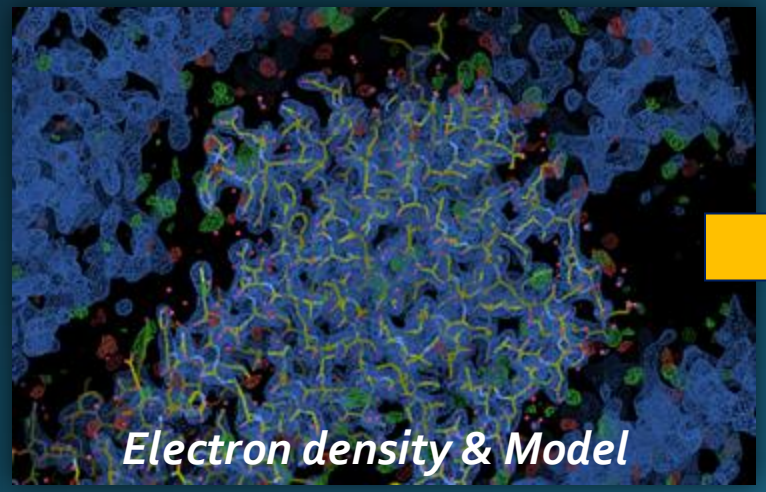
Introduction: *Basics of Molecular Replacement*

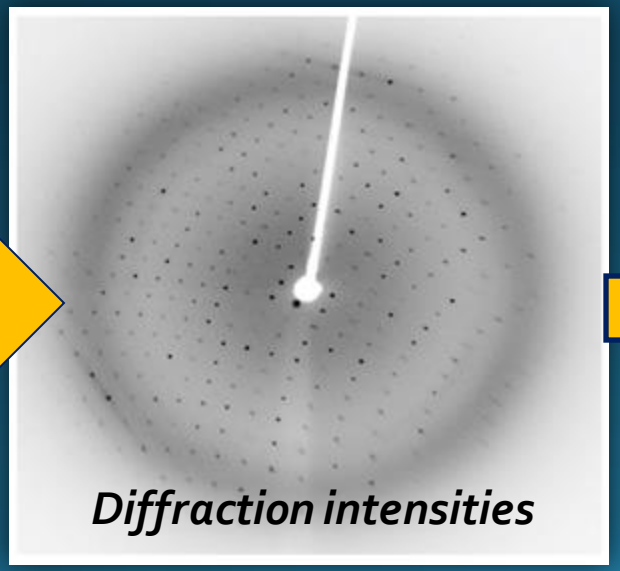
$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |F_o(hkl)| \cos[-2\pi(hx + ky + lz) + \varphi_{hkl}]$$

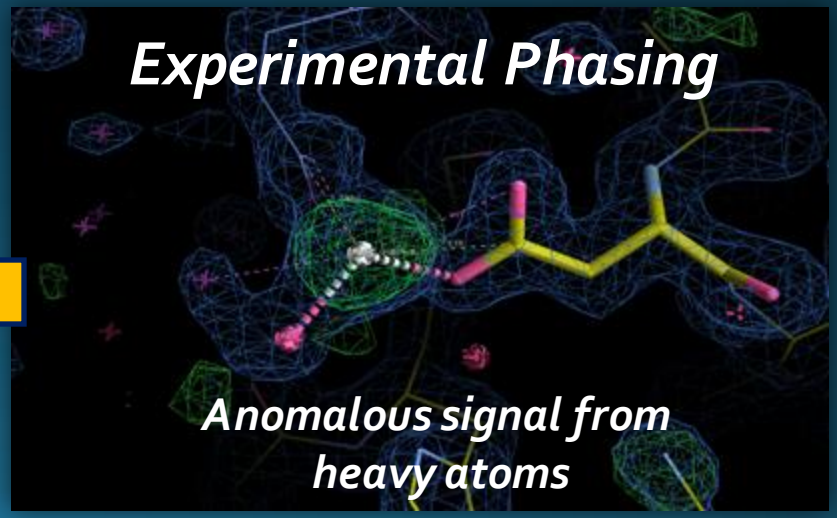
Electron density at (x,y,z)

Amplitudes

Phases







CCP4

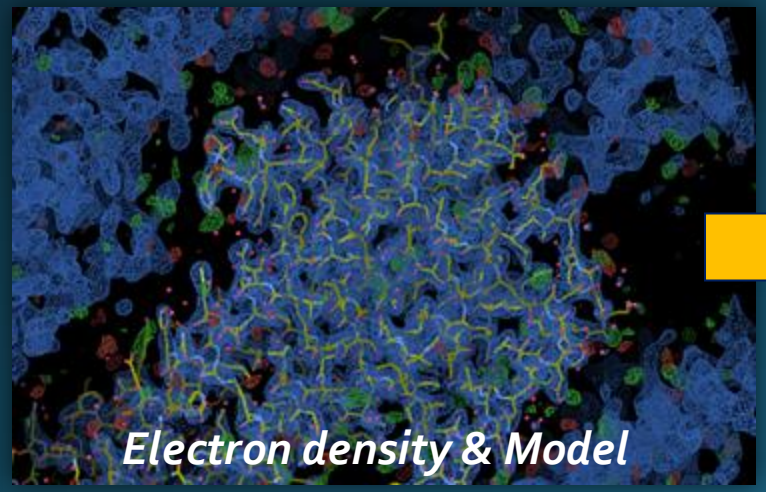
Introduction: *Basics of Molecular Replacement*

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

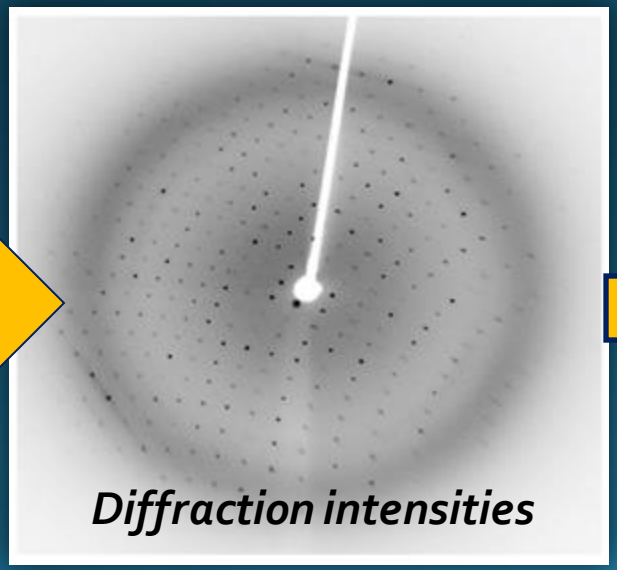
Electron density at (x,y,z)

Amplitudes

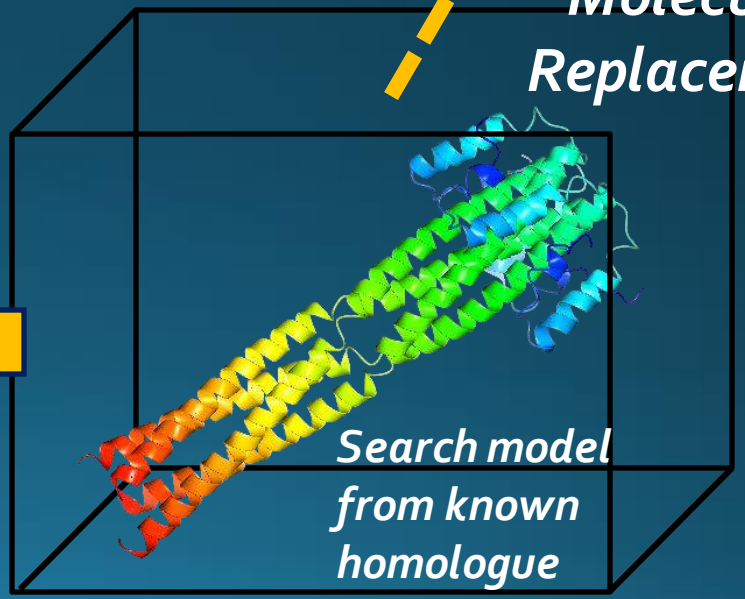
Molecular Replacement



Electron density & Model



Diffraction intensities



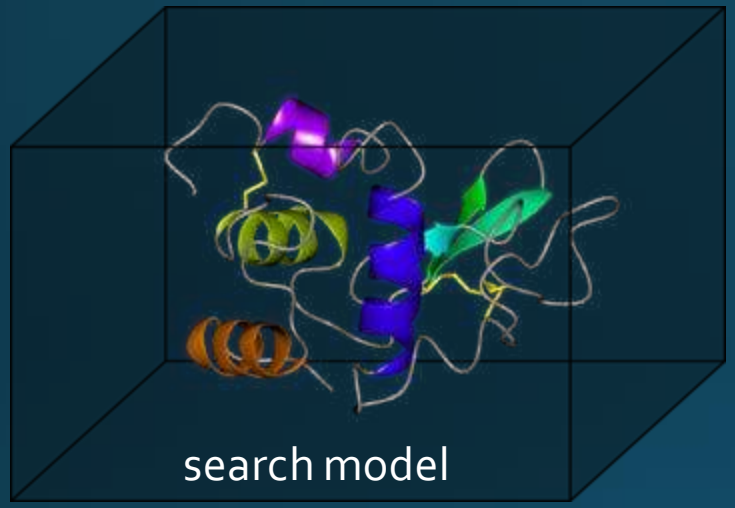
Search model from known homologue

CCP4

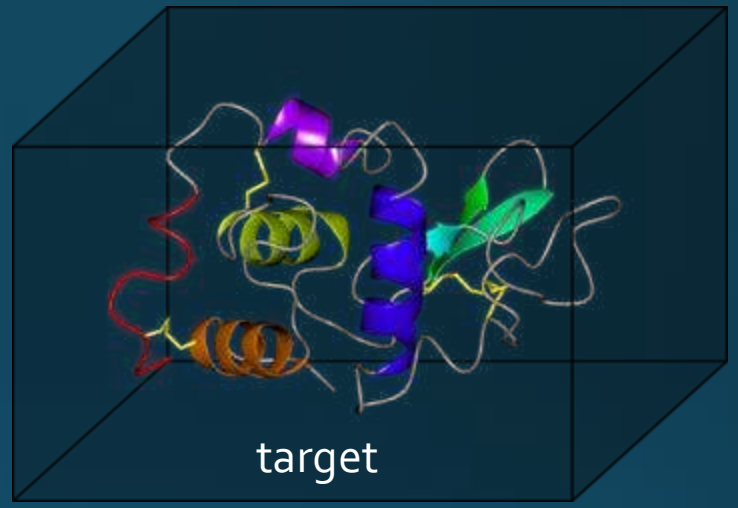
What is Molecular Replacement?

- 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*



search model



target

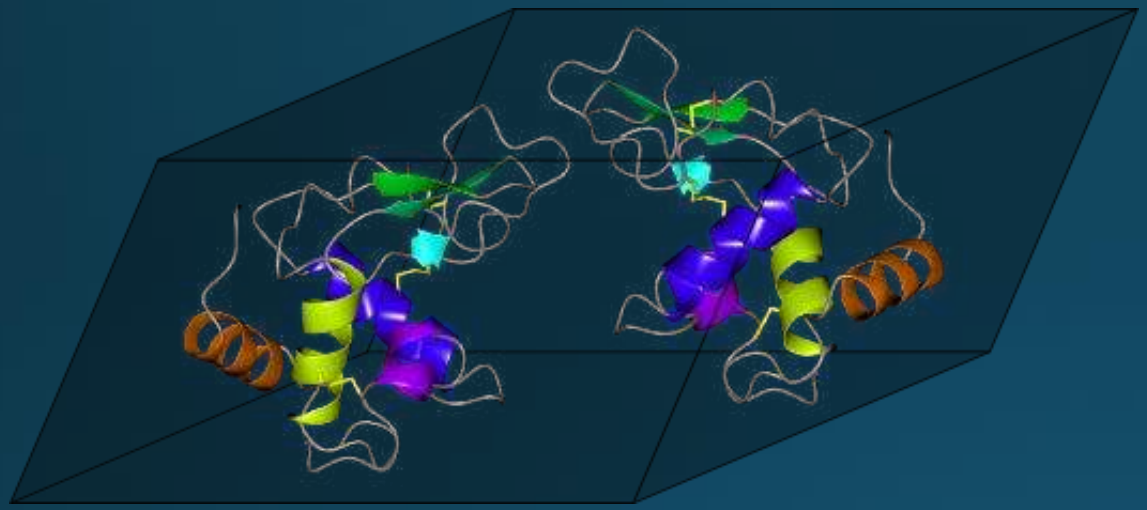
h	k	l	F	ϕ
0	0	1	12.6	123
0	0	2	2.1	12
0	0	3	69.9	287

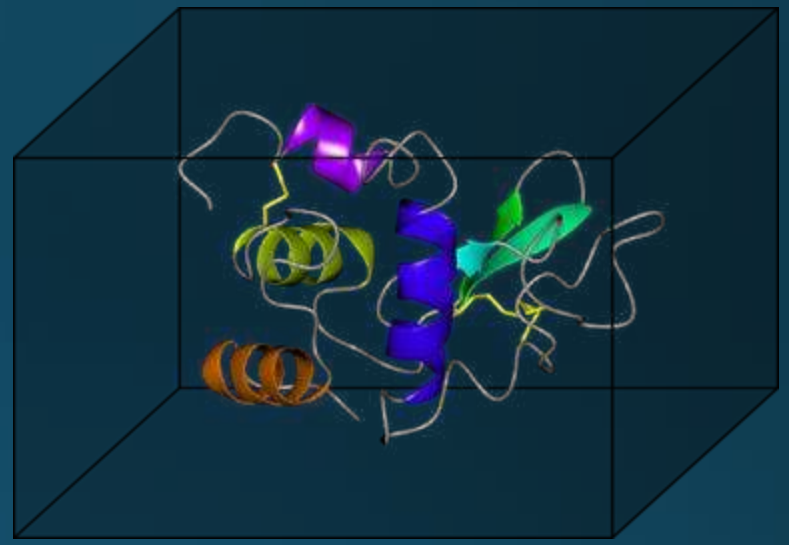
h	k	l	F	ϕ
0	0	1	10.4	113
0	0	2	3.5	18
0	0	3	57.2	265

(slide from Gabor Bunkoczi)

CCP4

Introduction: *Basics of Molecular Replacement*





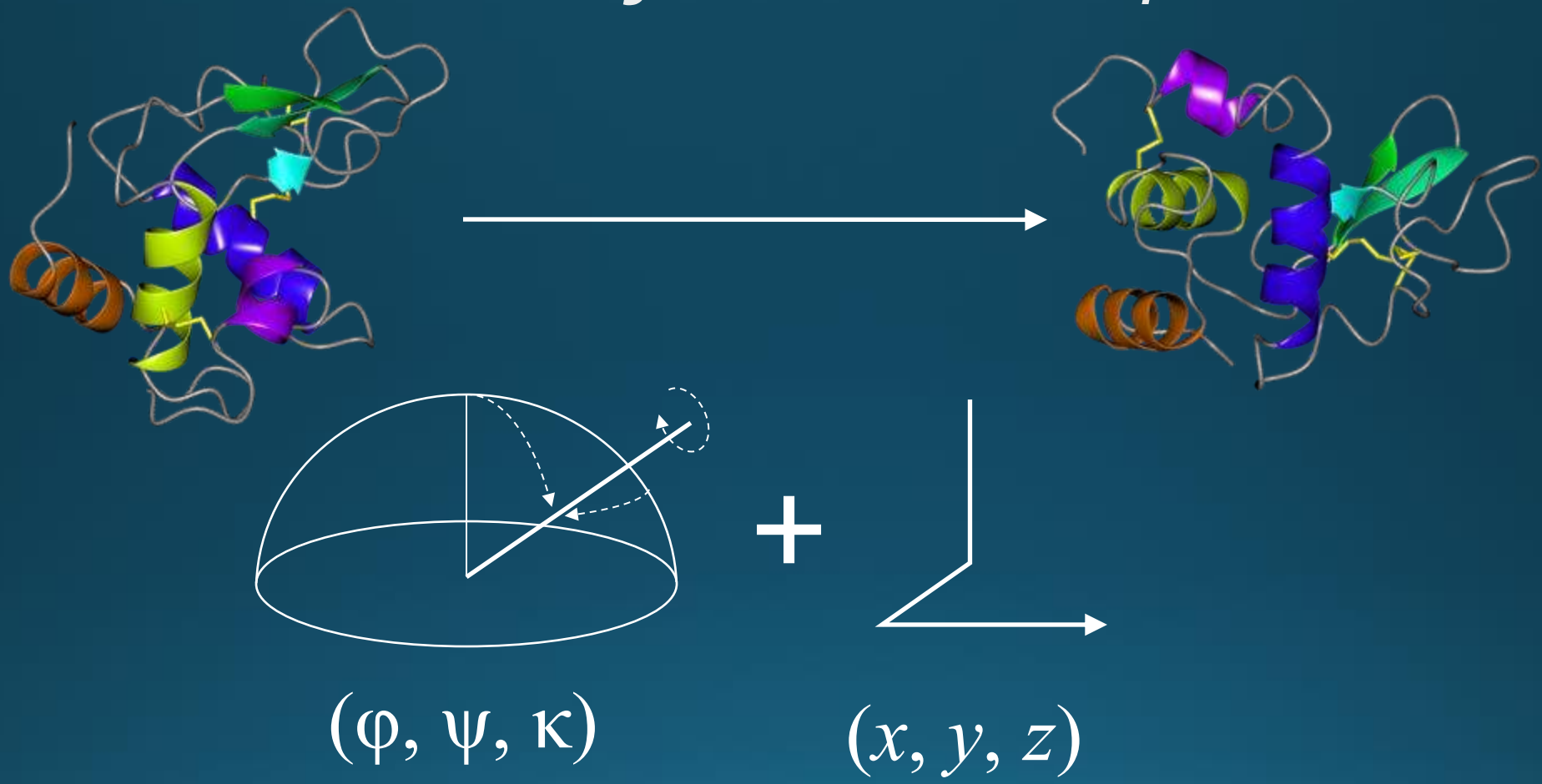
Search model crystal form

Target crystal form

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

(slide from Gabor Bunkoczi)

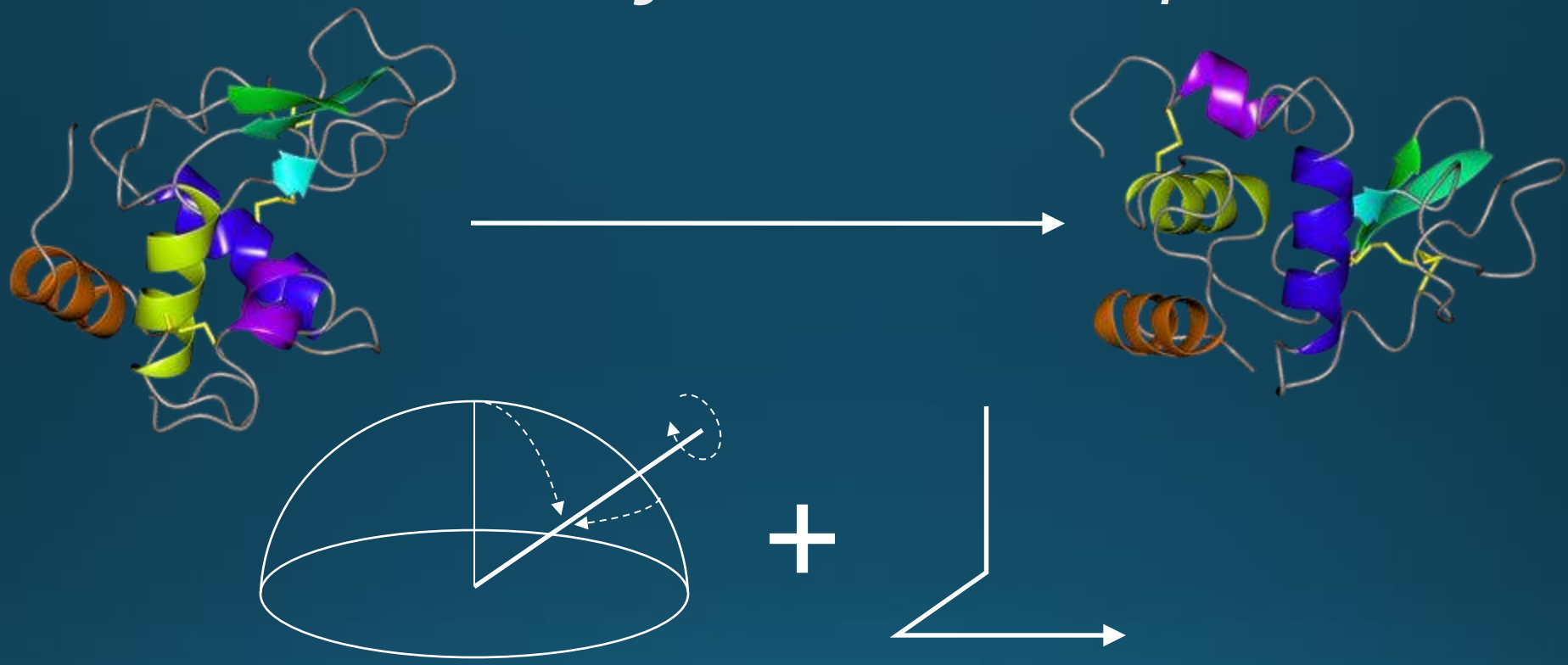
Introduction: *Basics of Molecular Replacement*



(slide from Gabor Bunkoczi)

CCP4

Introduction: *Basics of Molecular Replacement*



(φ, ψ, κ)

Rotation Function

(x, y, z)

Translation Function

(slide from Gabor Bunkoczi)

CCP4

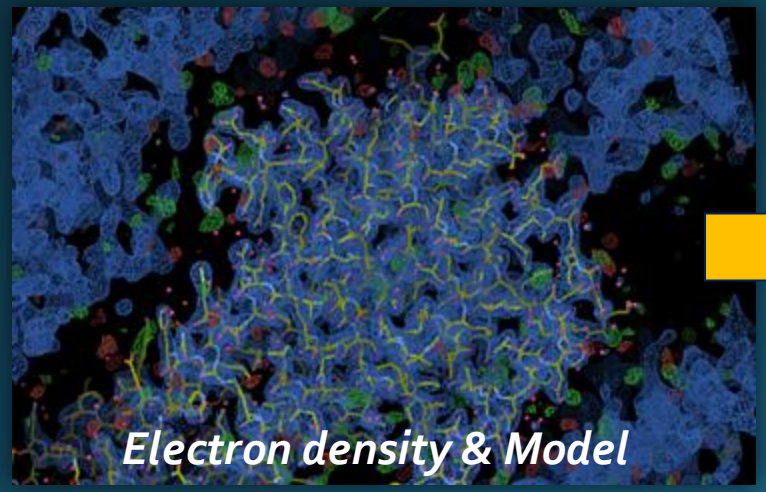
How do we find the correct positioning?

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

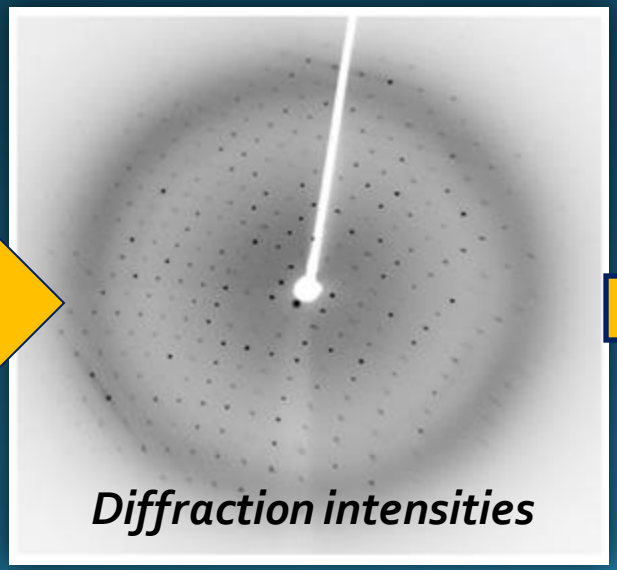
Electron density at (x,y,z)

Amplitudes

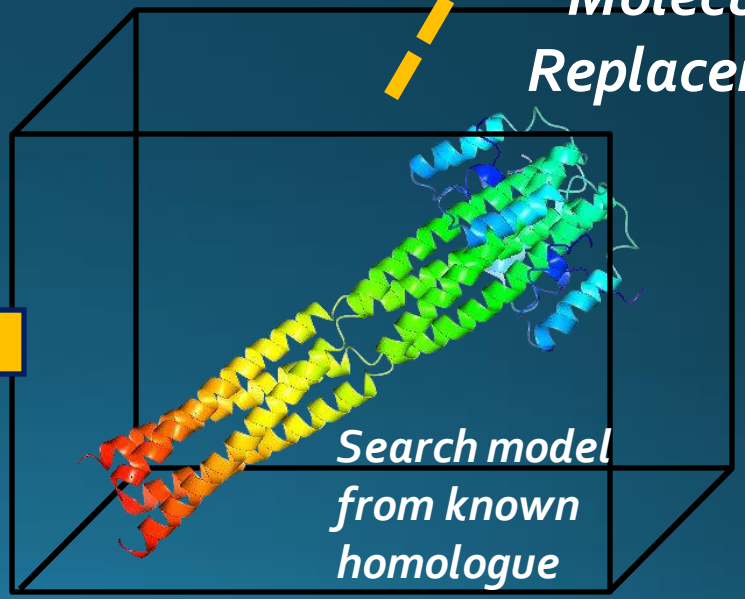
Molecular Replacement



Electron density & Model



Diffraction intensities



Search model from known homologue

CCP4

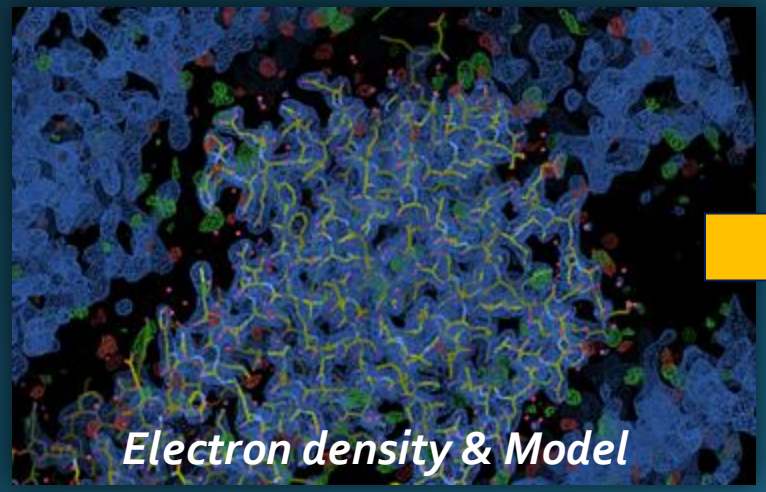
How do we find the correct positioning?

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

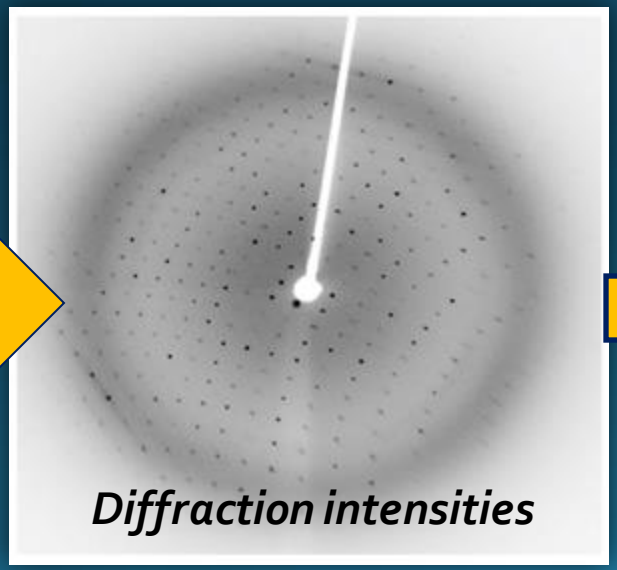
Electron density at (x,y,z)

Amplitudes

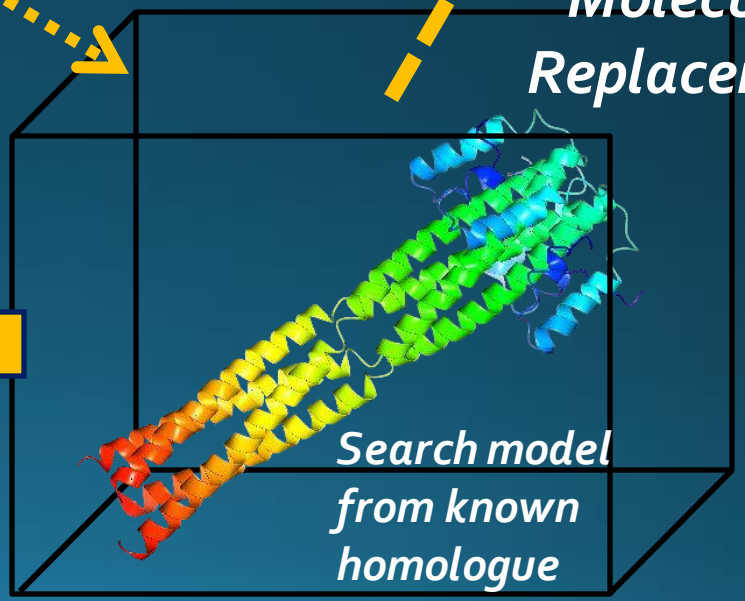
Molecular Replacement



Electron density & Model



Diffraction intensities

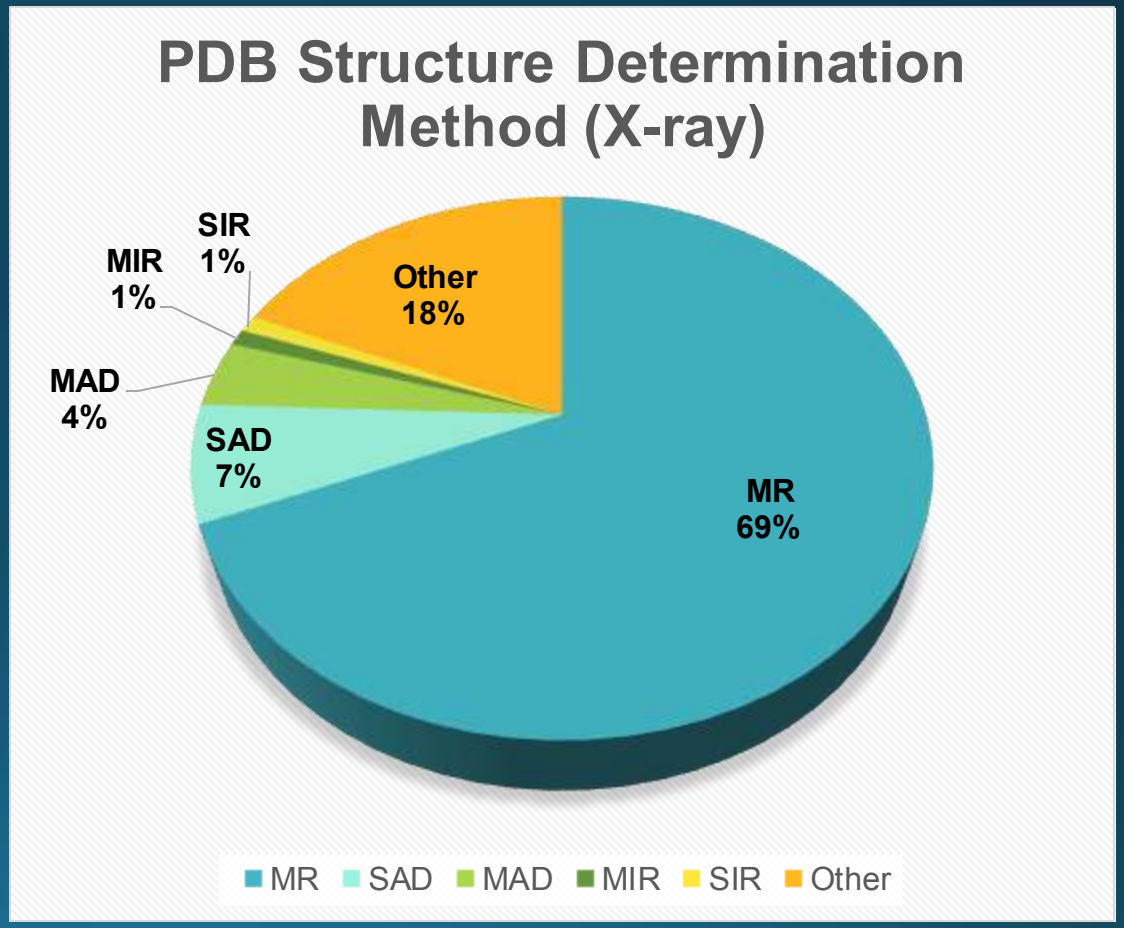


Search model from known homologue

CCP4

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

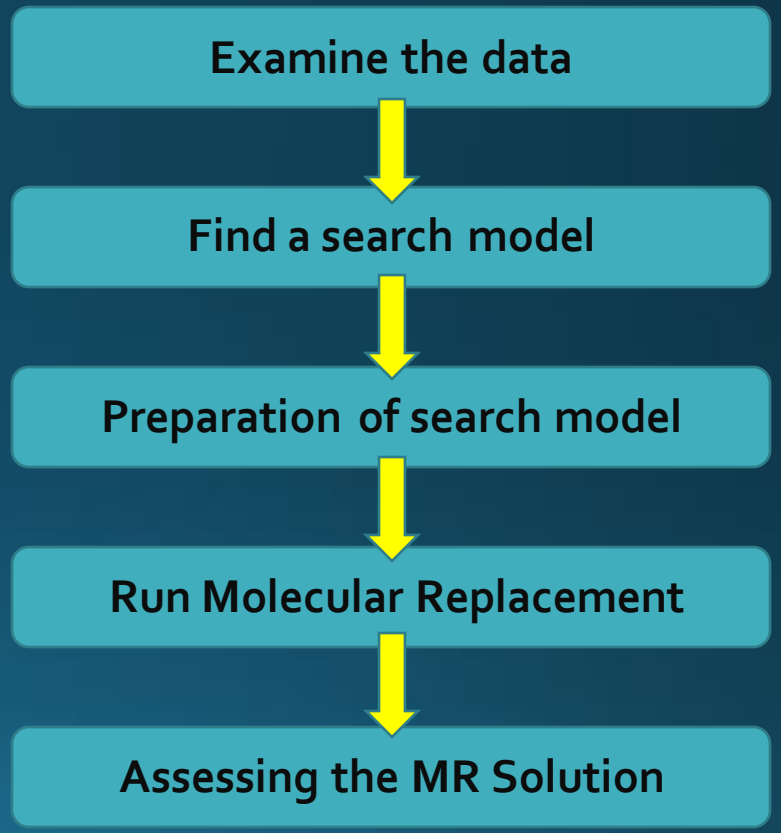


CCP4

MrBUMP

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



CCP4

Understanding the limitations of your data

Understanding the limitations of your data

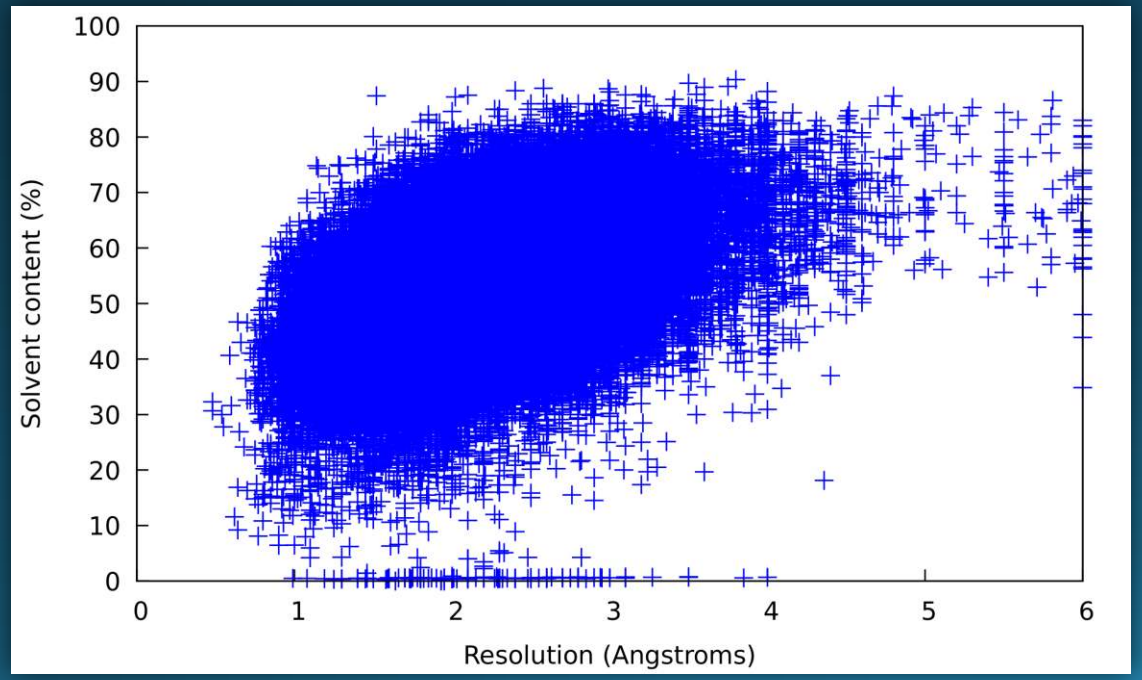
- Data issues can have an impact on how well MR will work

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



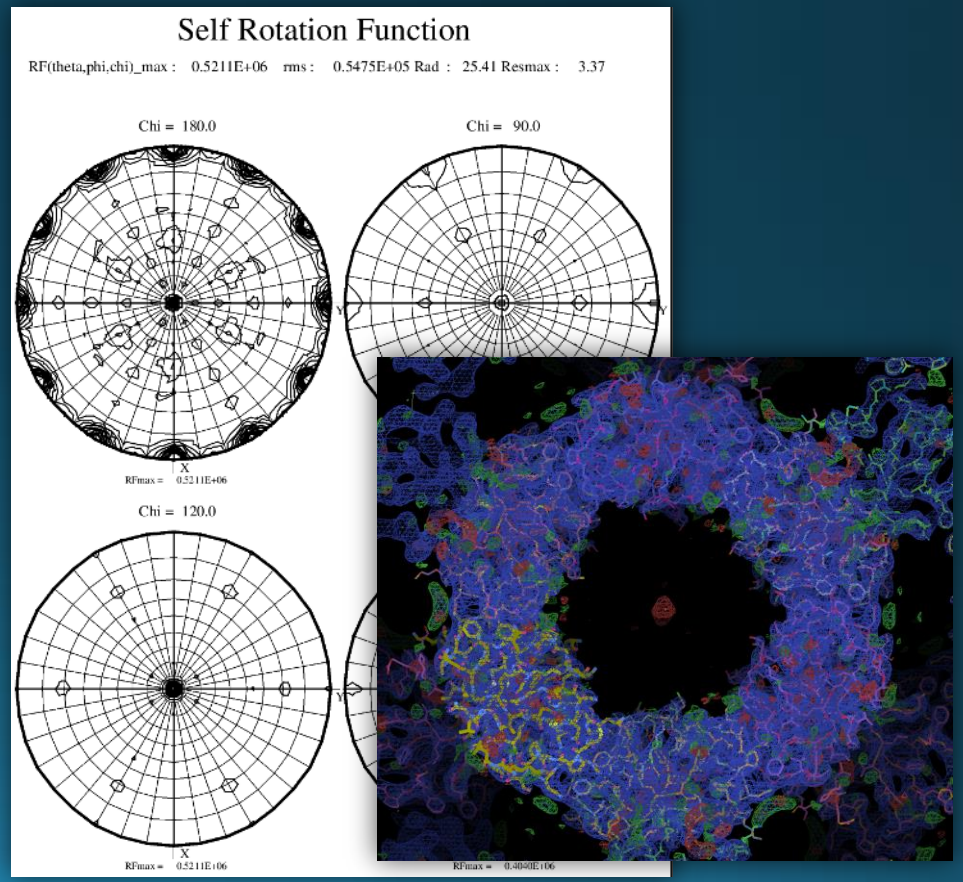
Distribution of solvent content across 156,000 X-ray datasets in the PDB

CCP4

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

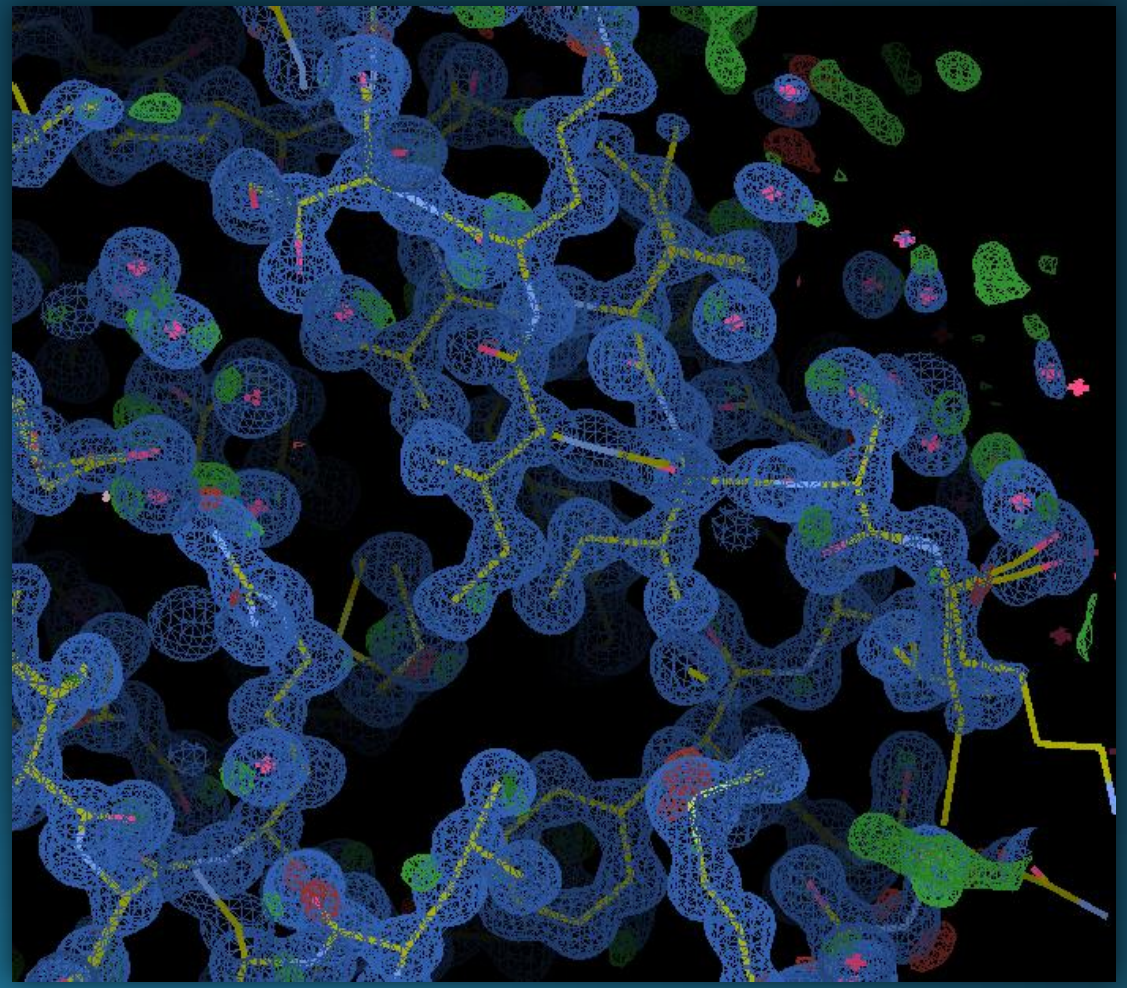
Understanding the limitations of your data

- 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

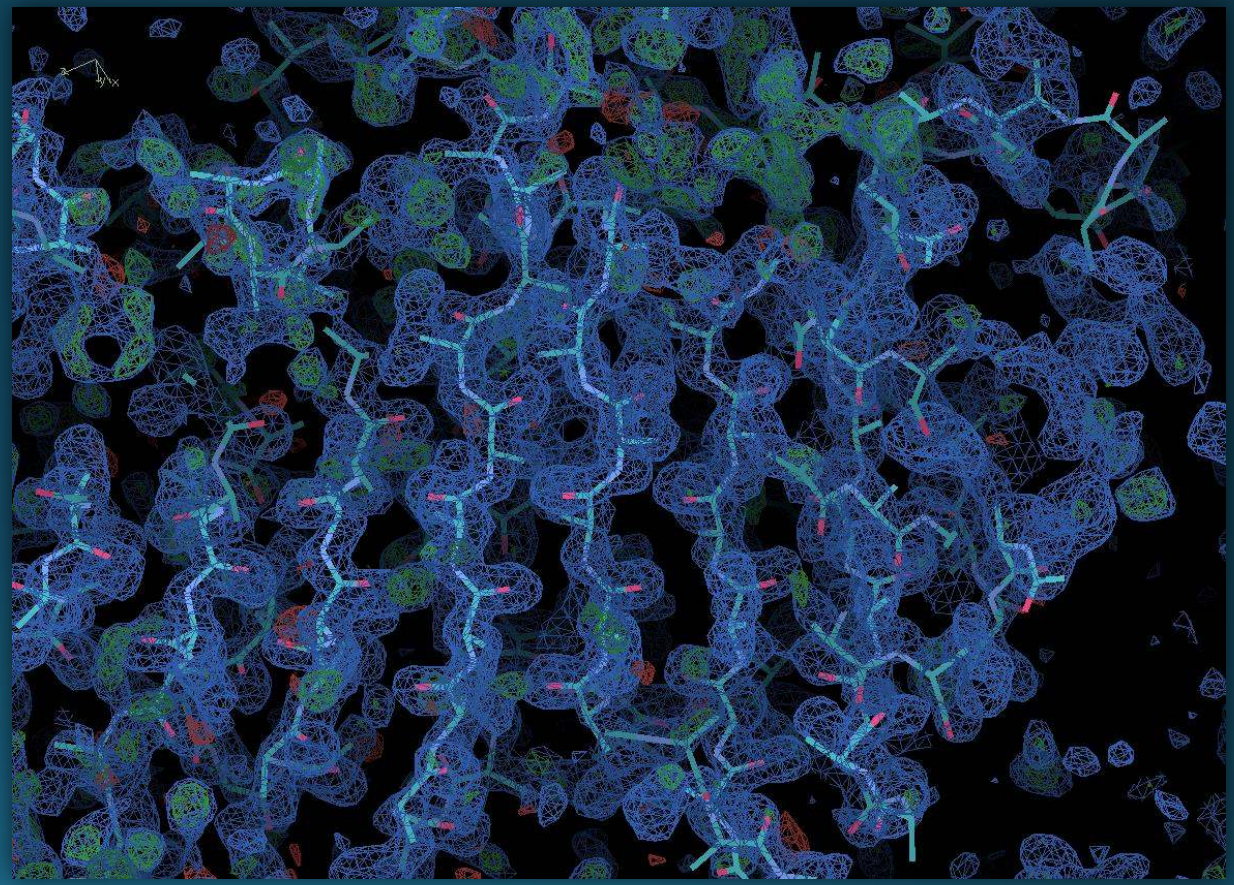
Resolution and Molecular Replacement

- 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

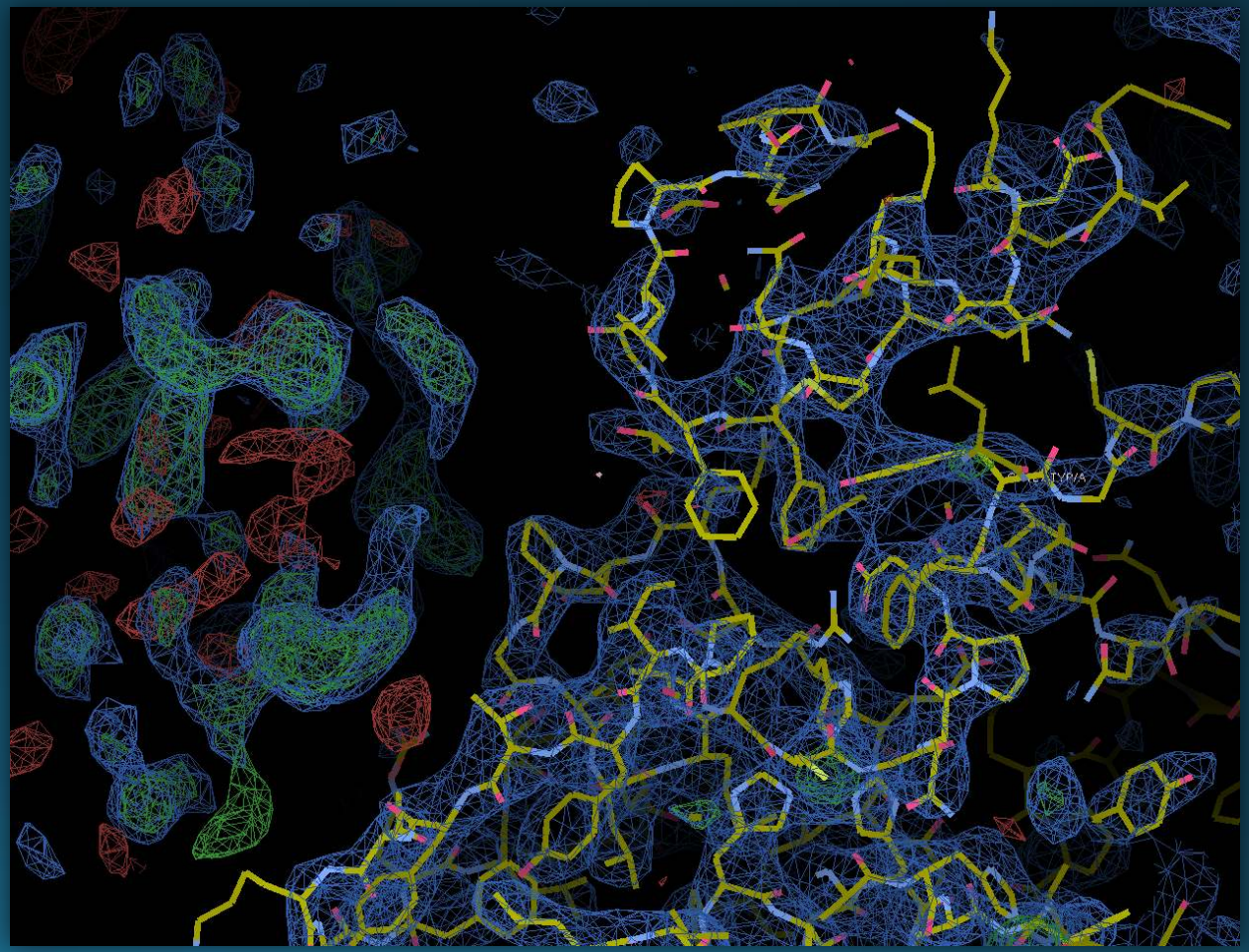
- At resolutions above ~ 2.5 Å
  - *Phaser* can place small fragments of total scattering
  - Applications like *SHELXE* and *Acorn* can improve through density modification (DM) and model building to a correct solution



CCP4

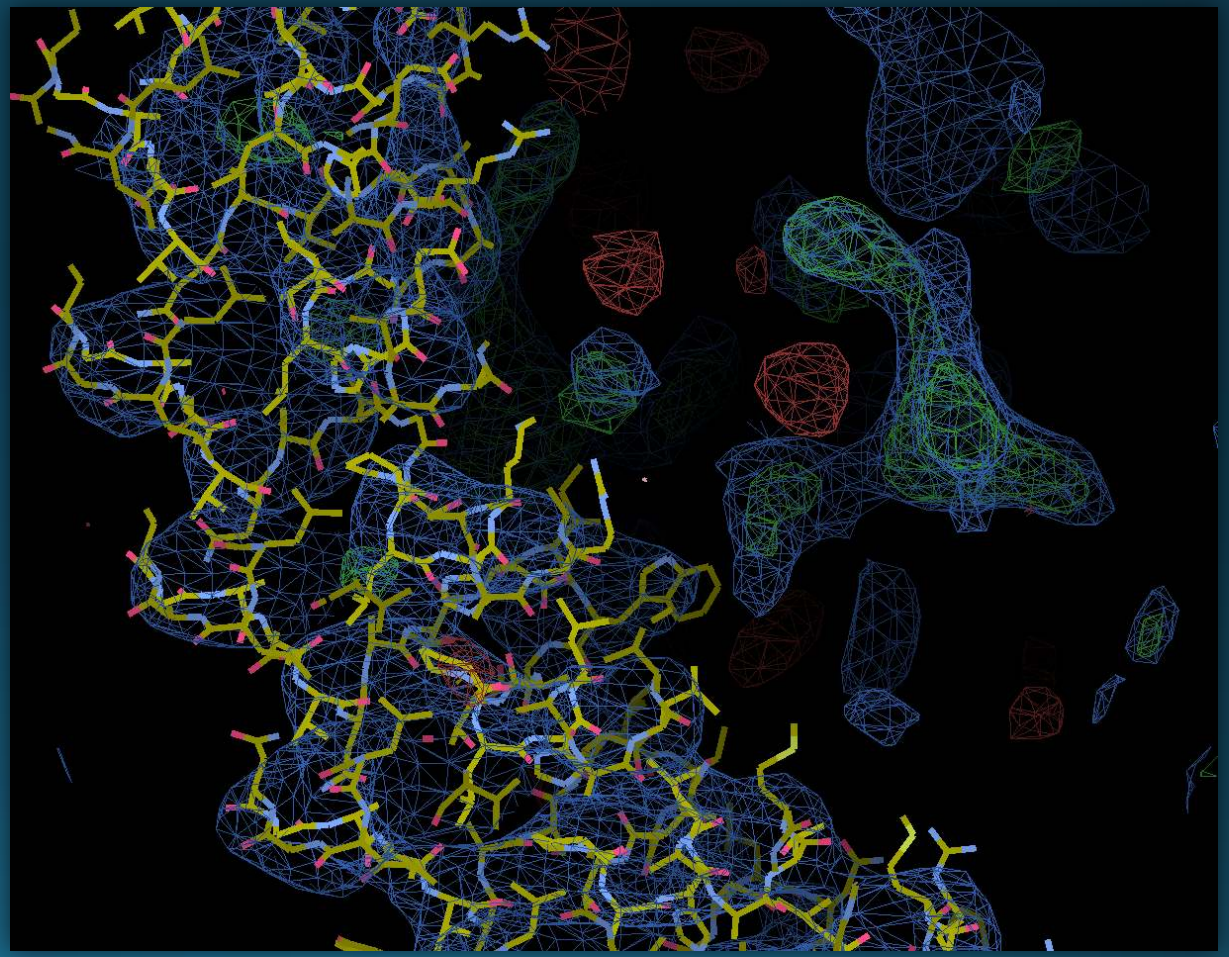
Resolution and Molecular Replacement

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



Resolution and Molecular Replacement

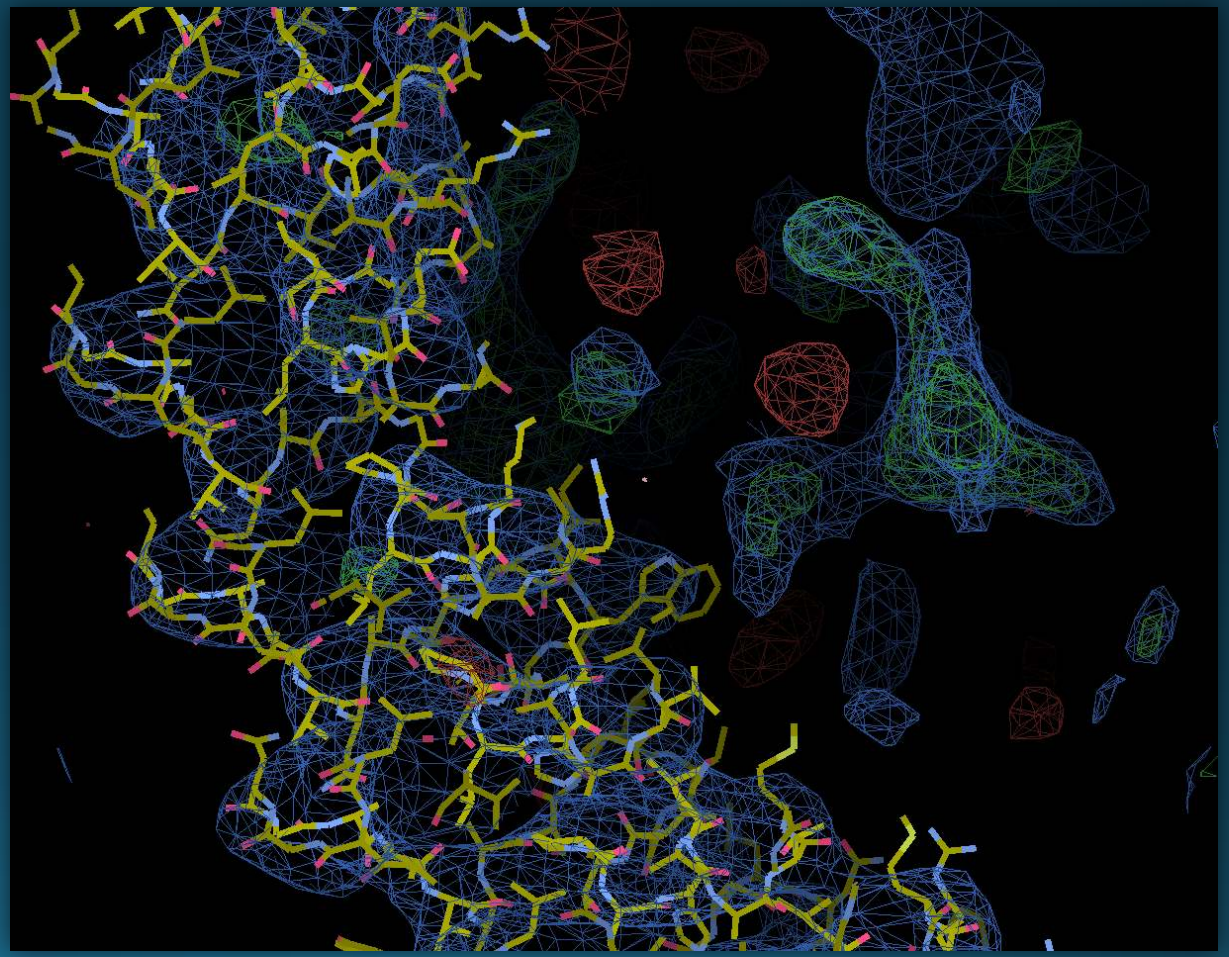
- Below 4Å automated model building becomes difficult



CCP4

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



CCP4

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

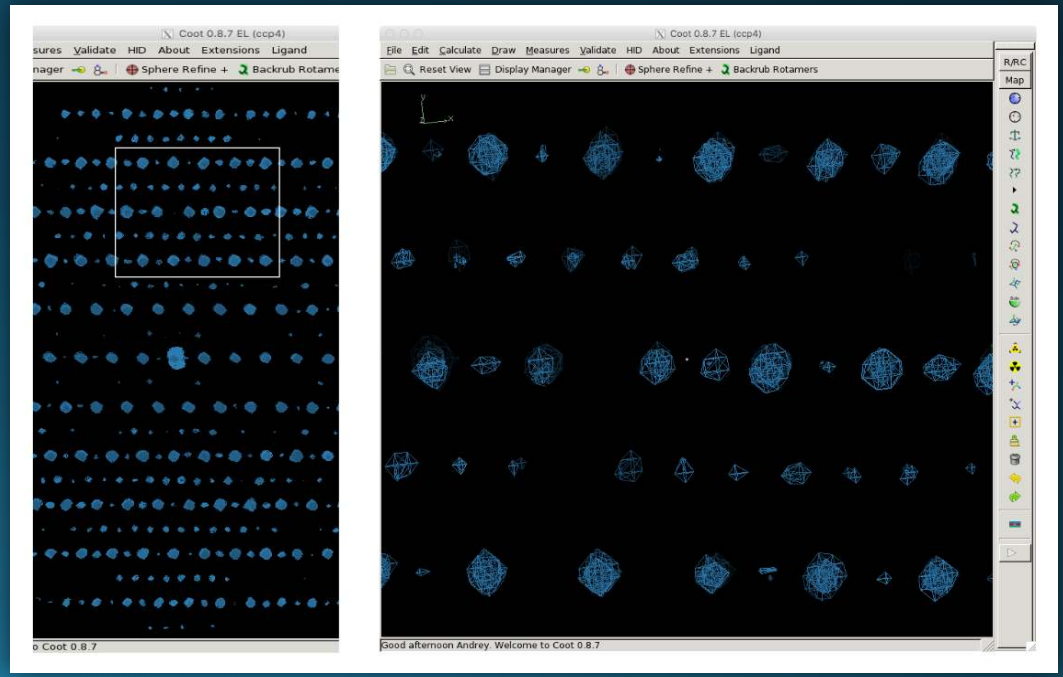
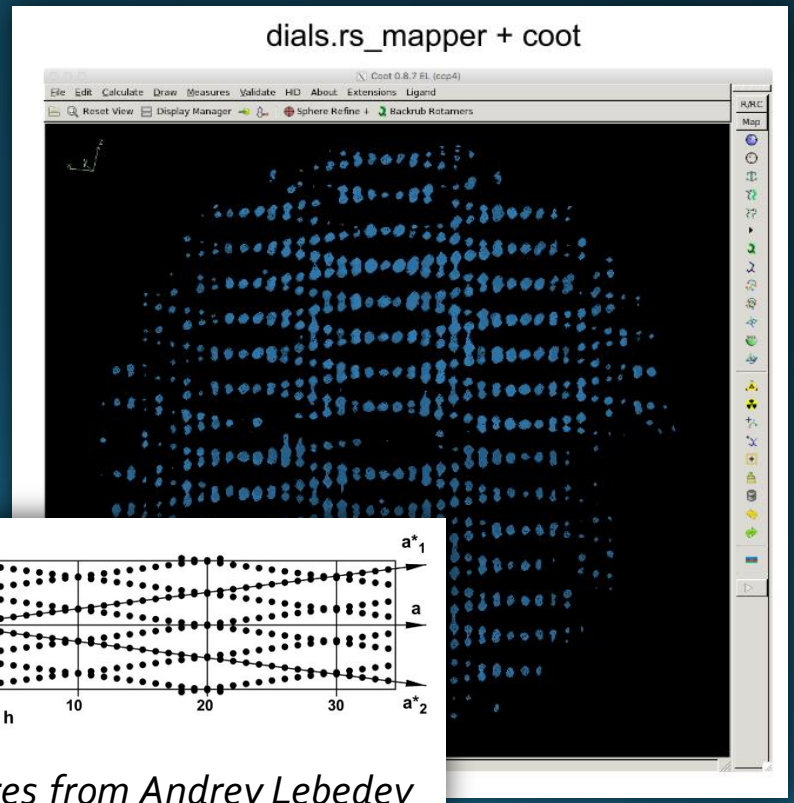


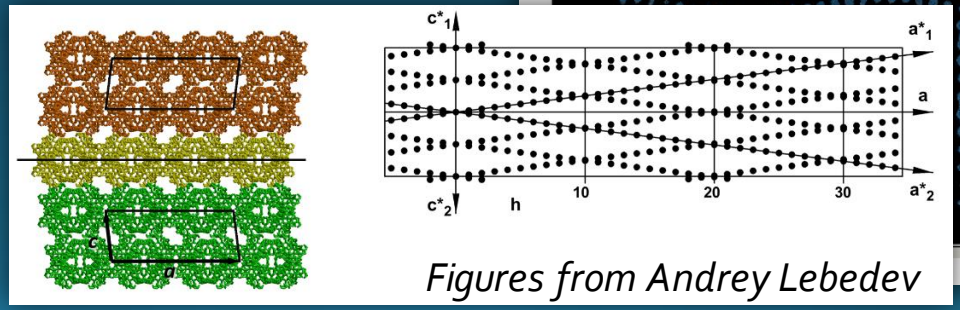
Figure from Andrey Lebedev

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





Figures 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?
  - Space group assignment correct?

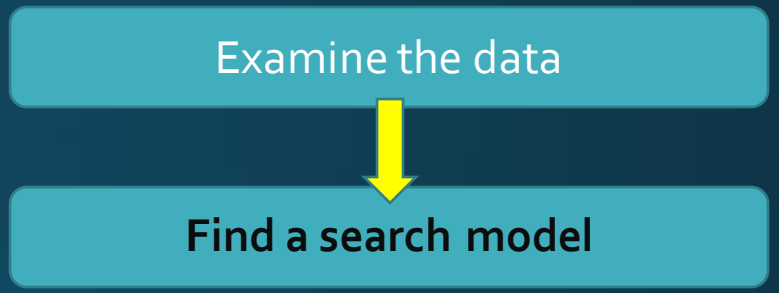
Examine the data

CCP4

Finding suitable Search Models

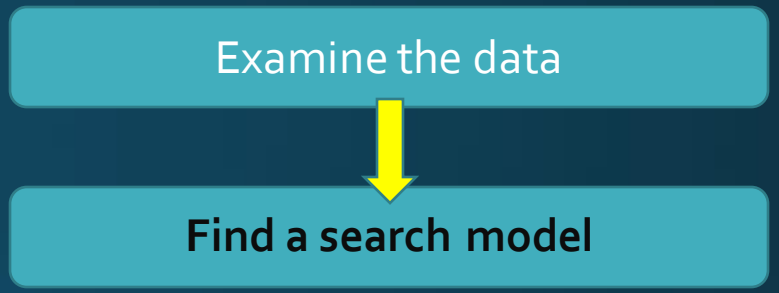
Finding suitable search models

- Considerations



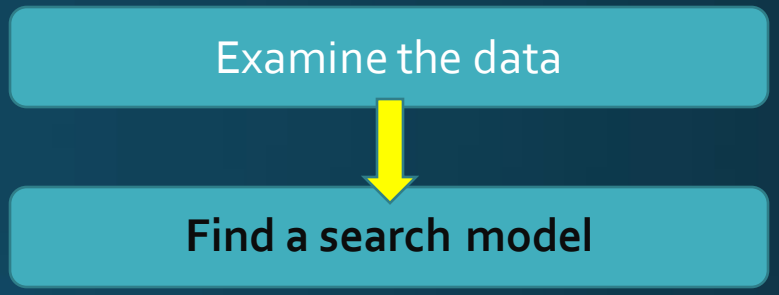
Finding suitable search models

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



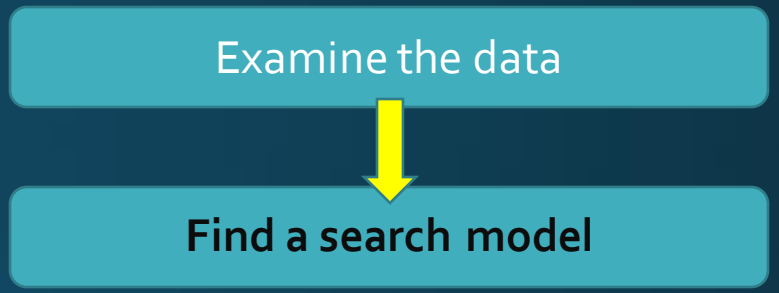
Finding suitable search models

- 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



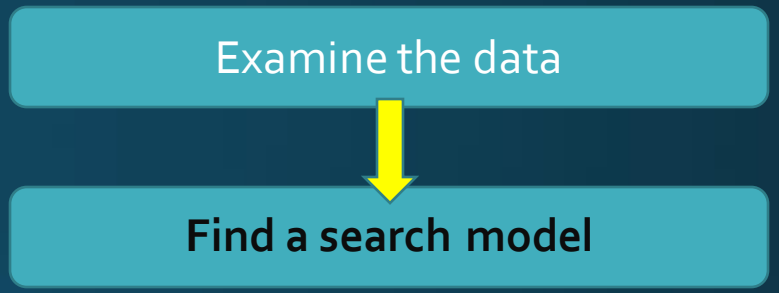
Finding suitable search models

- 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



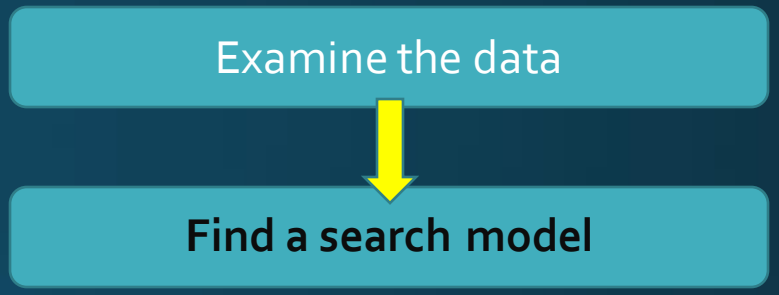
Finding suitable search models

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



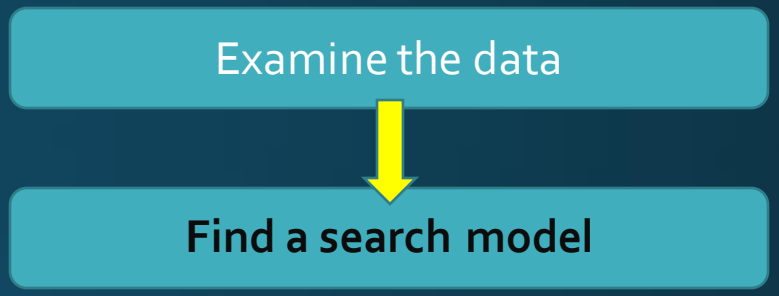
Finding suitable search models

- 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

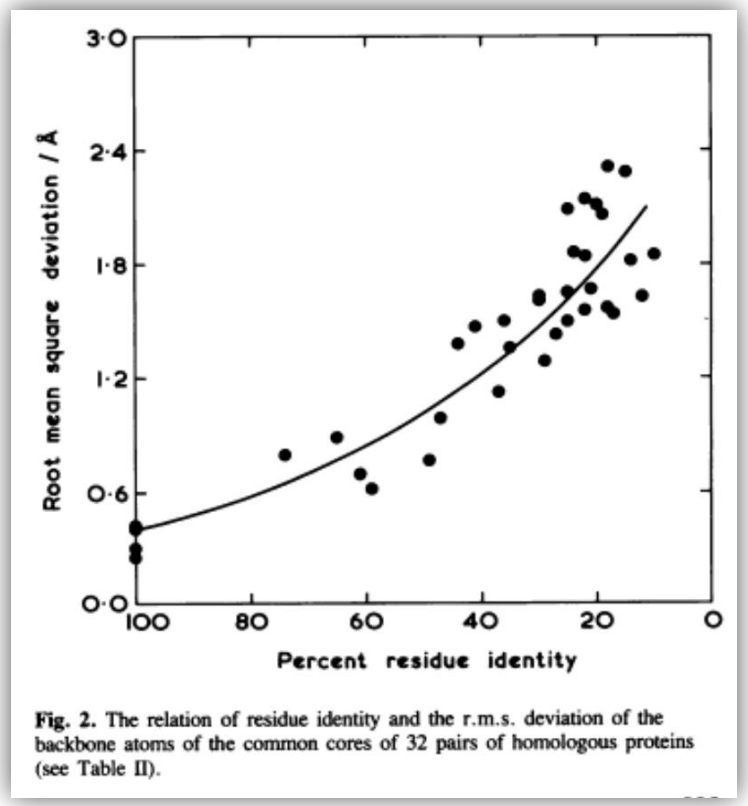


Fig. 2. The relation of residue identity and the r.m.s. deviation of the backbone atoms of the common cores of 32 pairs of homologous proteins (see Table II).

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

Examine the data

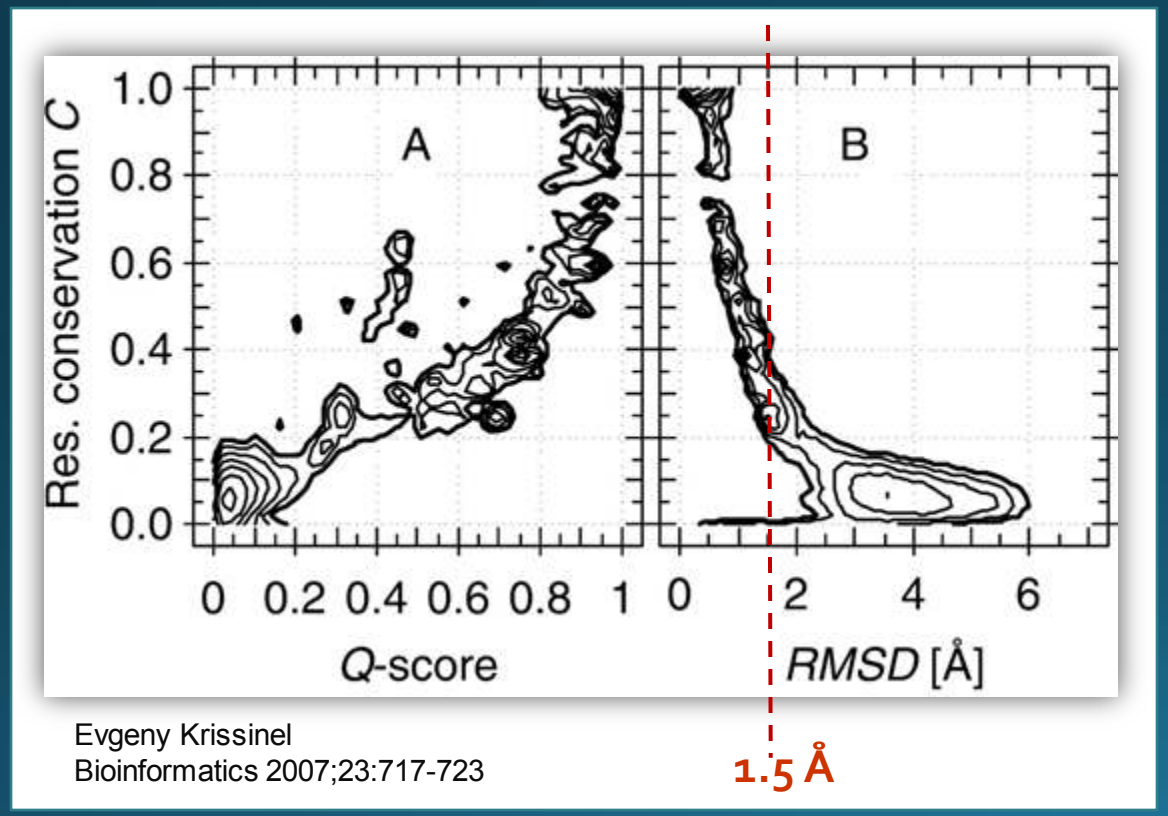
↓

Find a search model

CCP4

Finding suitable search models

- Amino acid sequence similarity often correlates well with structural similarity



Evgeny Krissinel
Bioinformatics 2007;23:717-723

Examine the data

Find a search model

CCP4

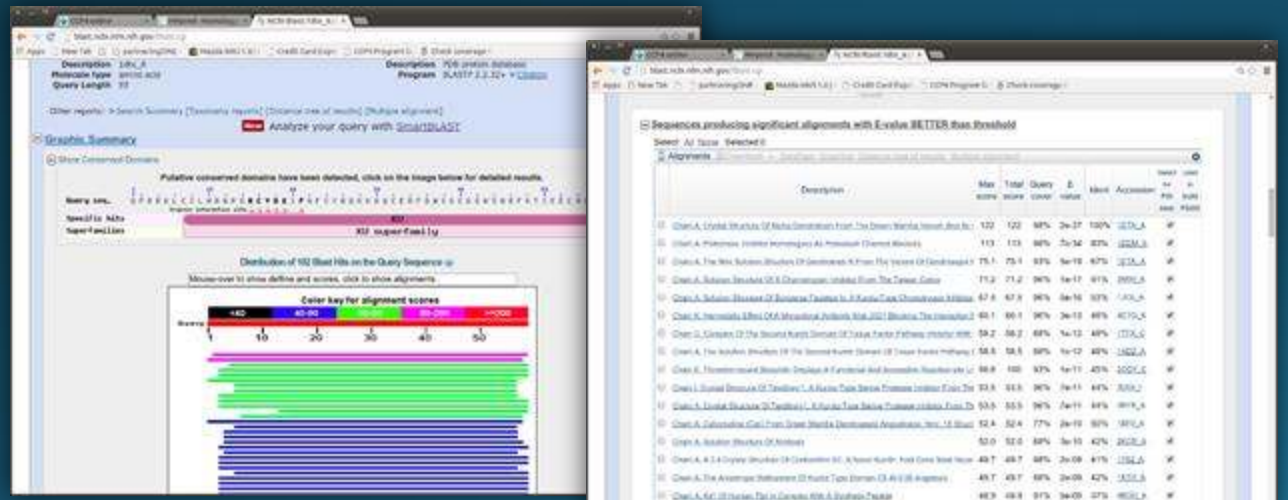
Finding suitable search models

- 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



CCP4

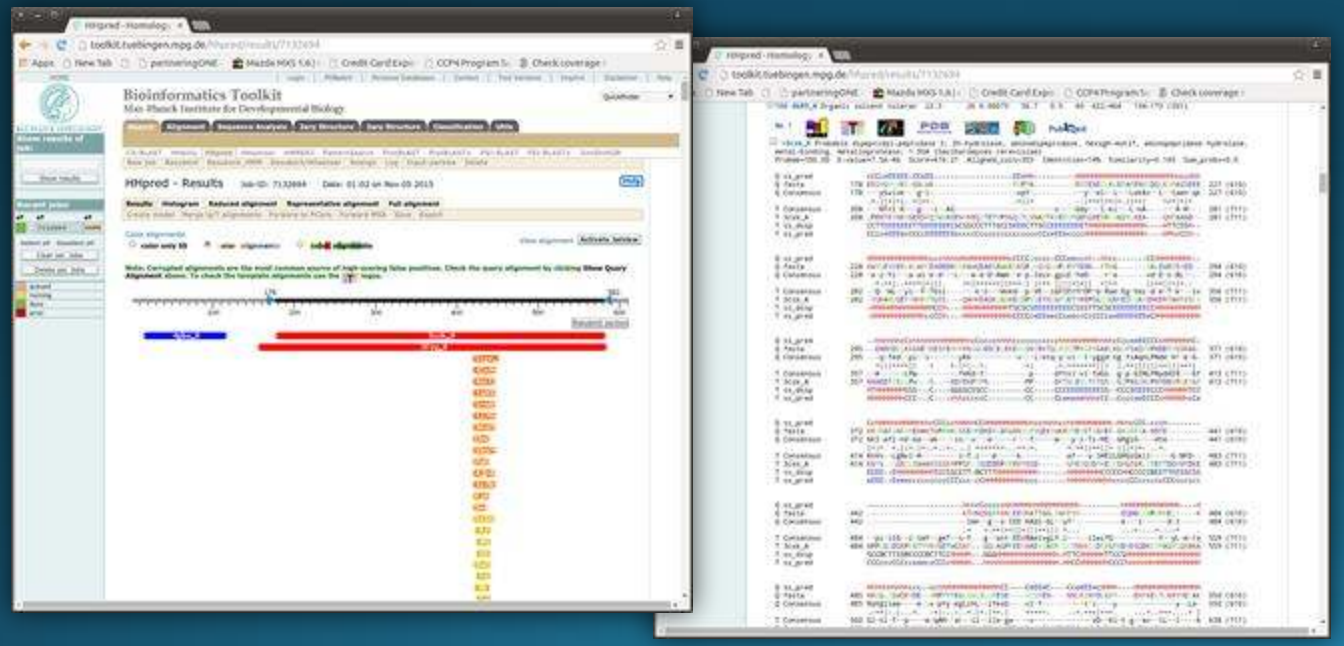
Finding suitable search models

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

Examine the data

↓

Find a search model

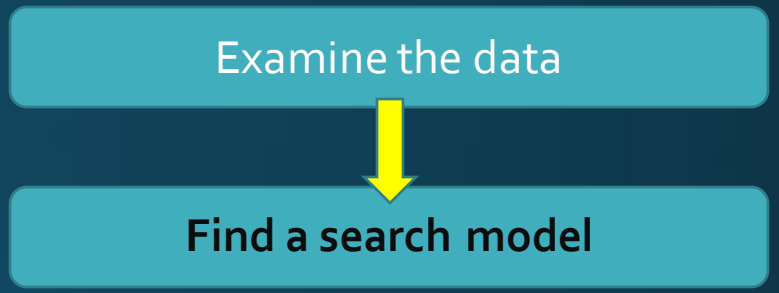


CCP4

The modern method..

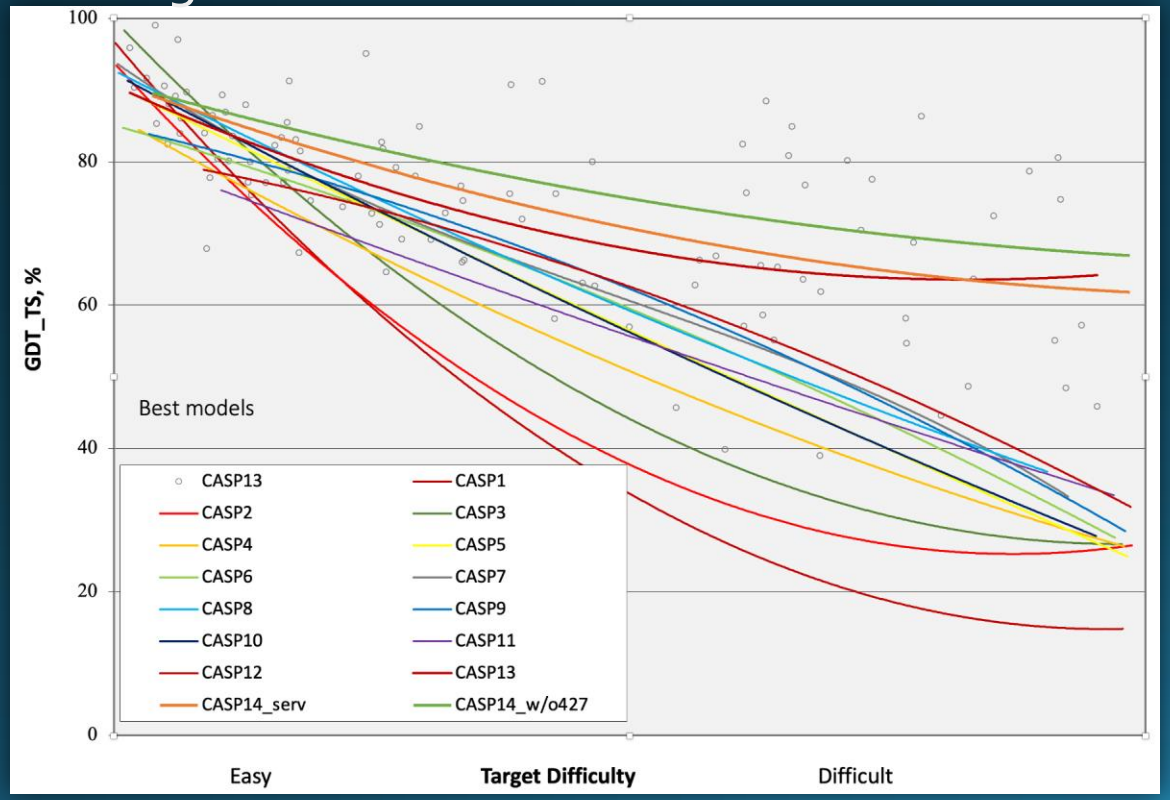
Finding suitable search models

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



Finding suitable search models

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



Examine the data

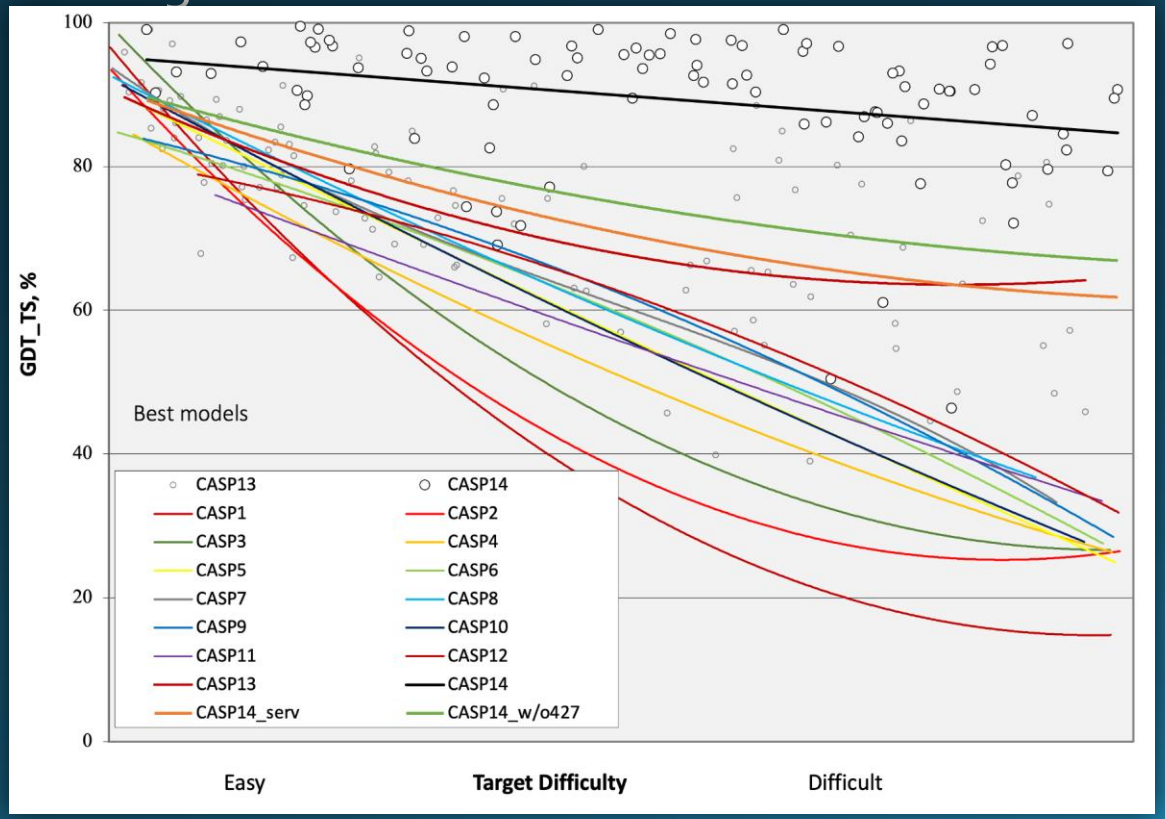
↓

Find a search model

Figure from John Moult (CASP)

Finding suitable search models

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

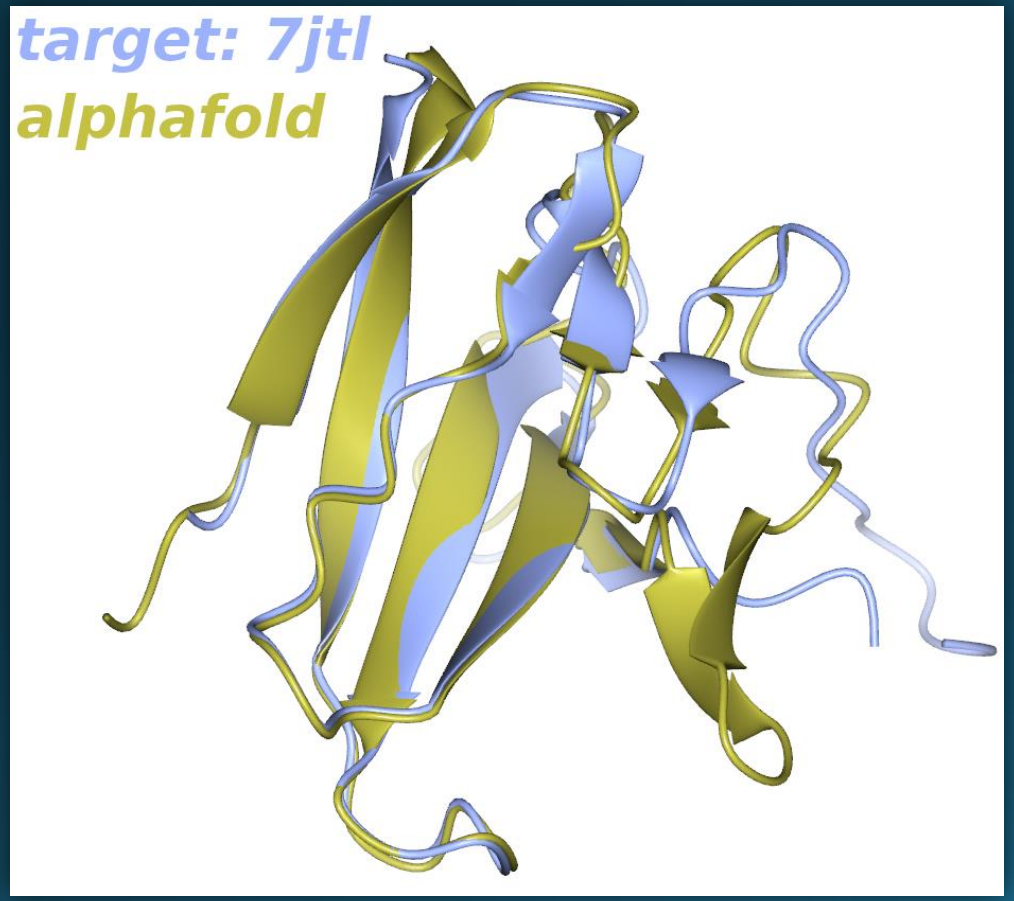


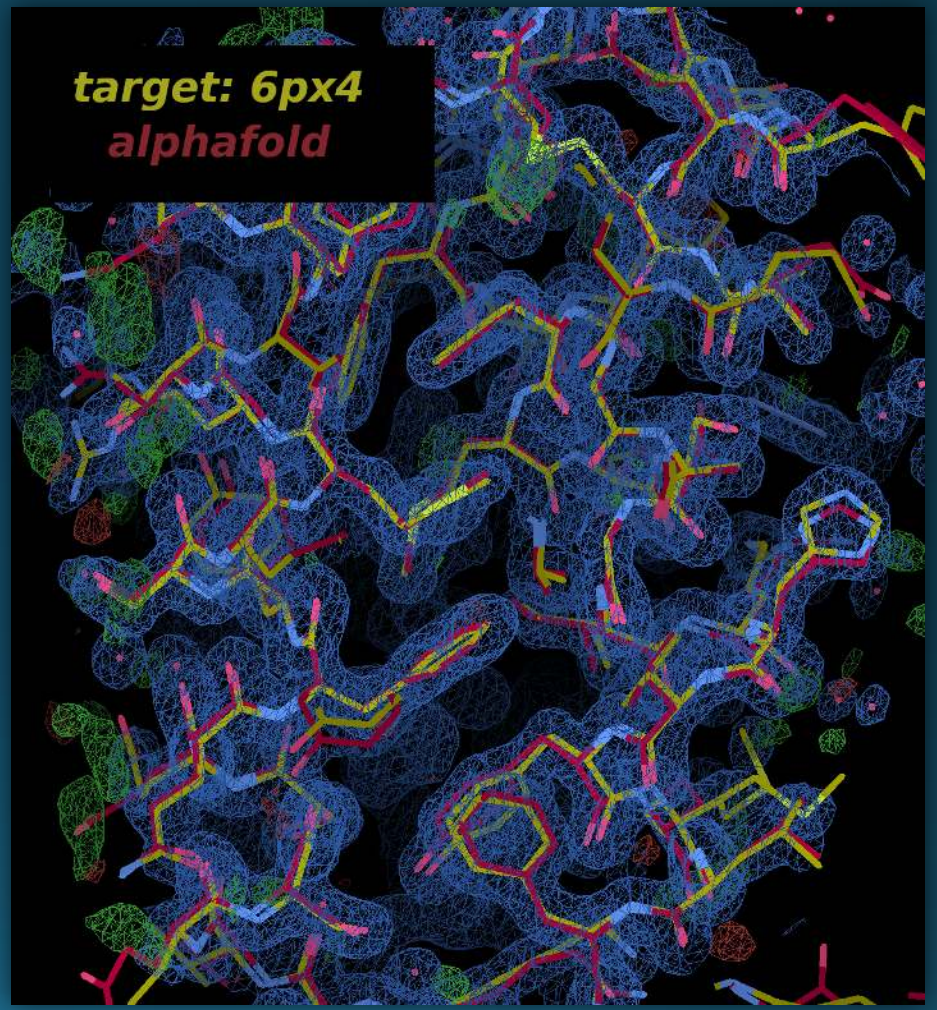
Examine the data

Find a search model

Figure from John Moult (CASP)

- *Alphafold CASP14 examples*

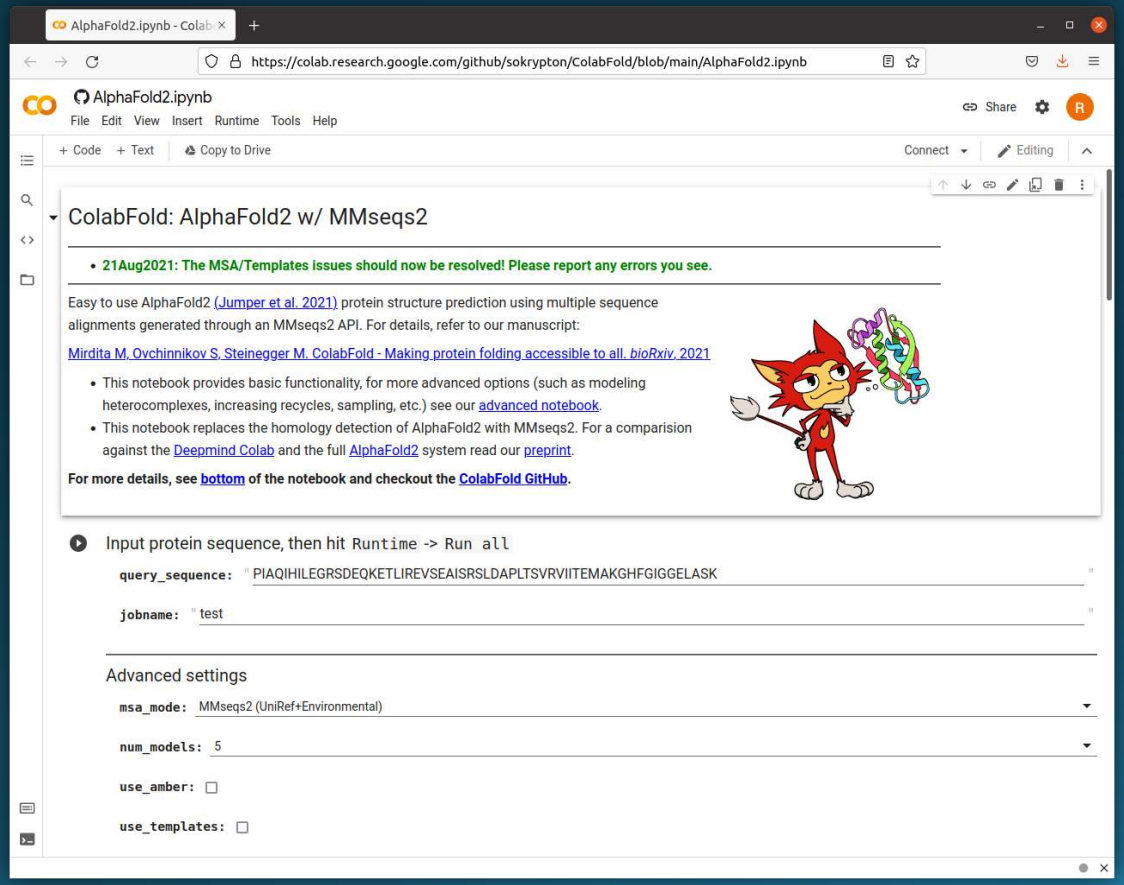




CCP4

Finding suitable search models

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



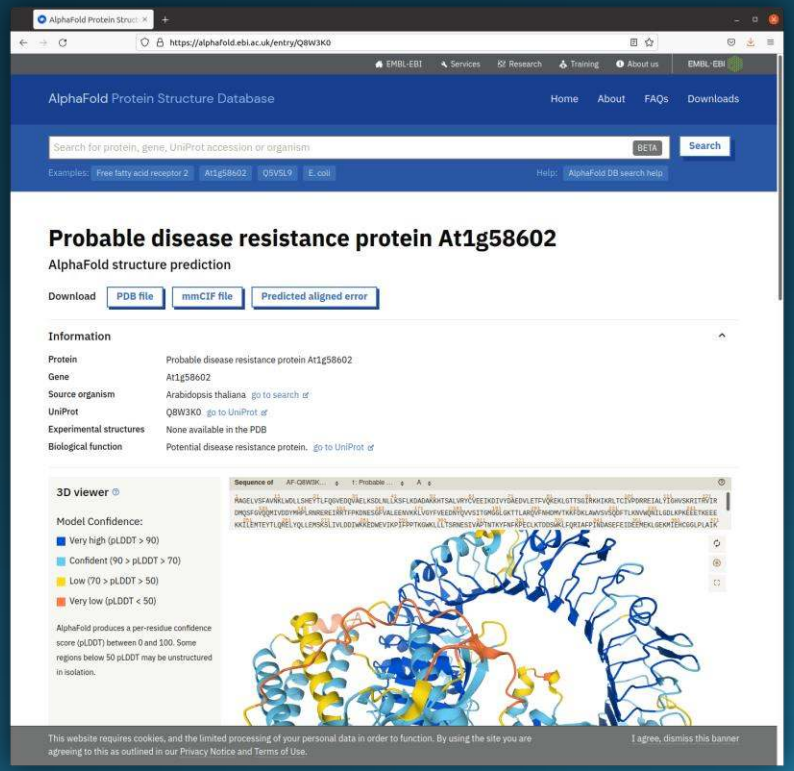
Examine the data

↓

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

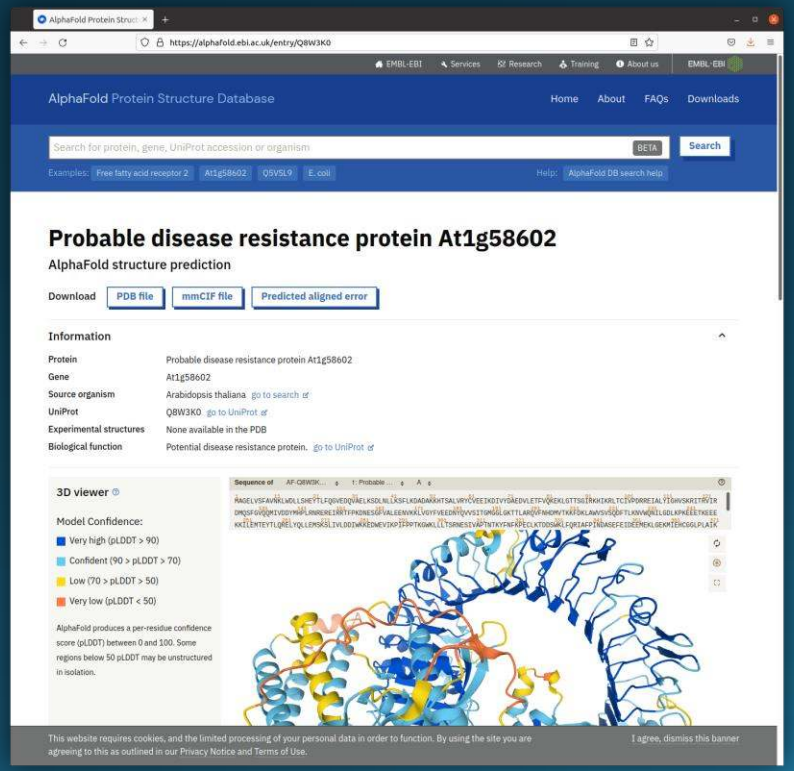
Finding suitable search models

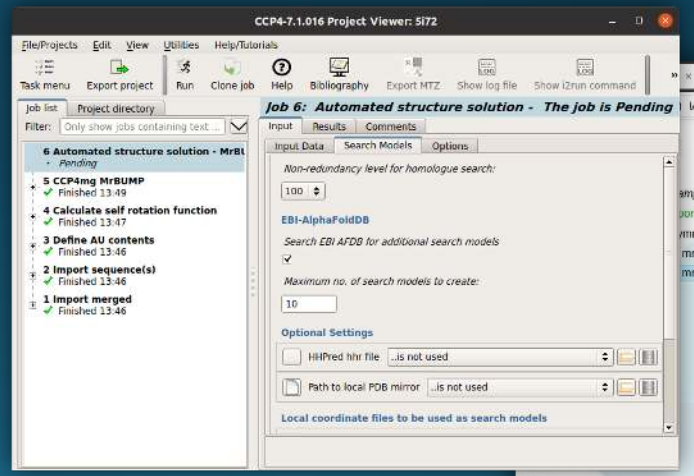
- Accessing Alphafold
  1. Google Colab Servers (see Claudia's talk next)
  2. EBI-Alphafold database
     1. 350000 predictions currently available
     2. 130 million in the near future (early 2022)

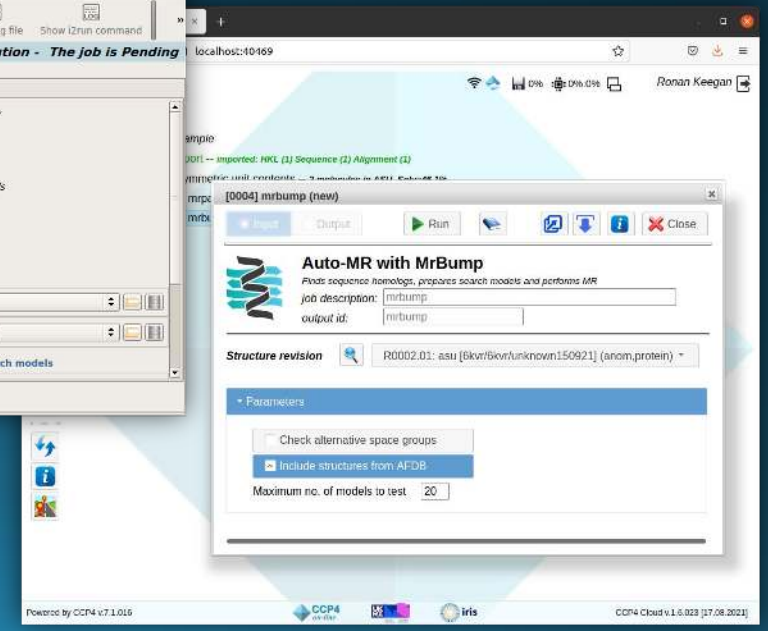
Examine the data

↓

Find a search model

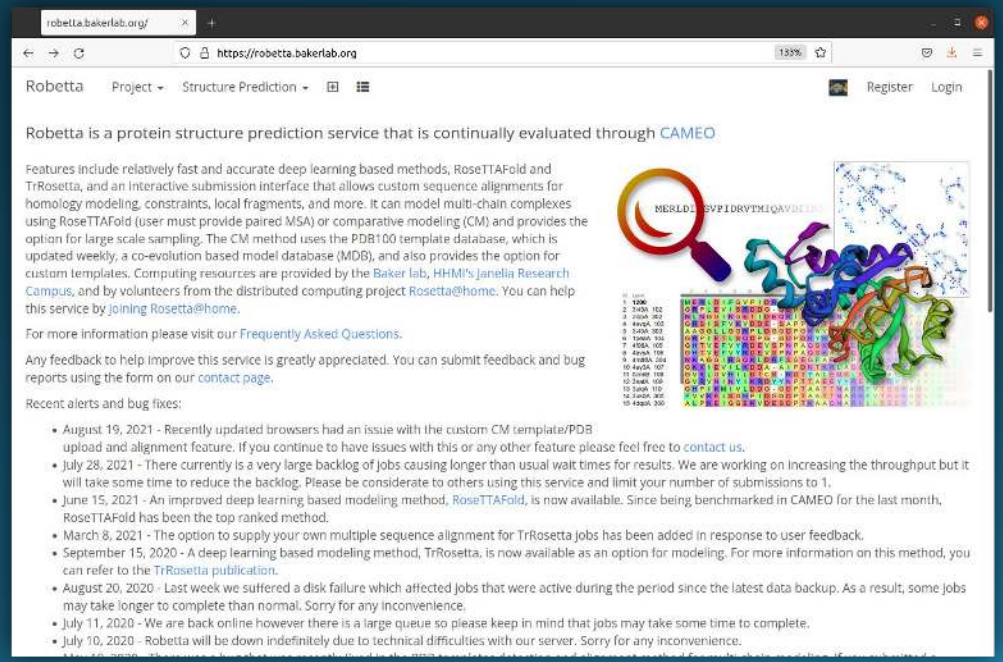






Finding suitable search models

- 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



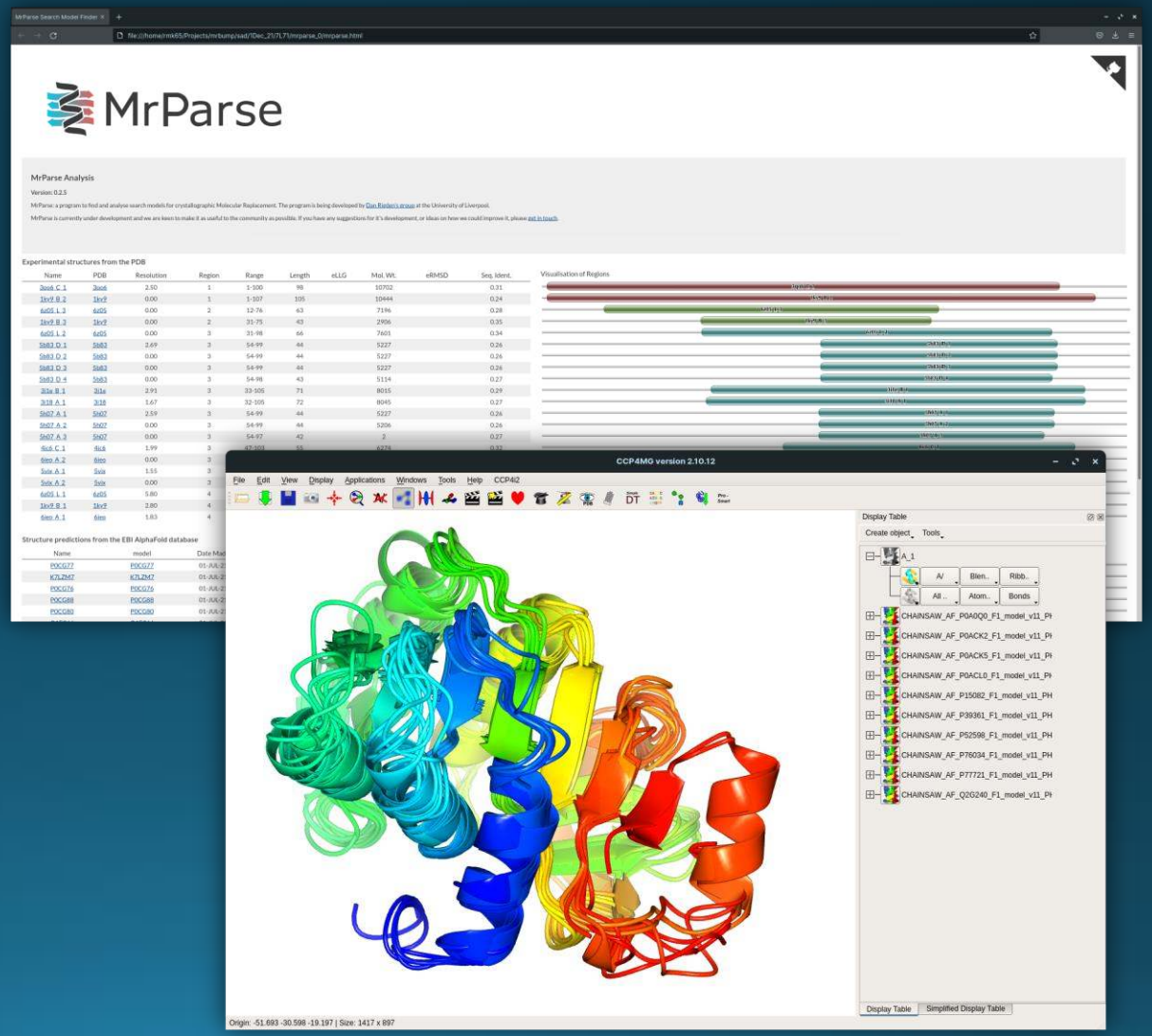
Examine the data

↓

Find a search model

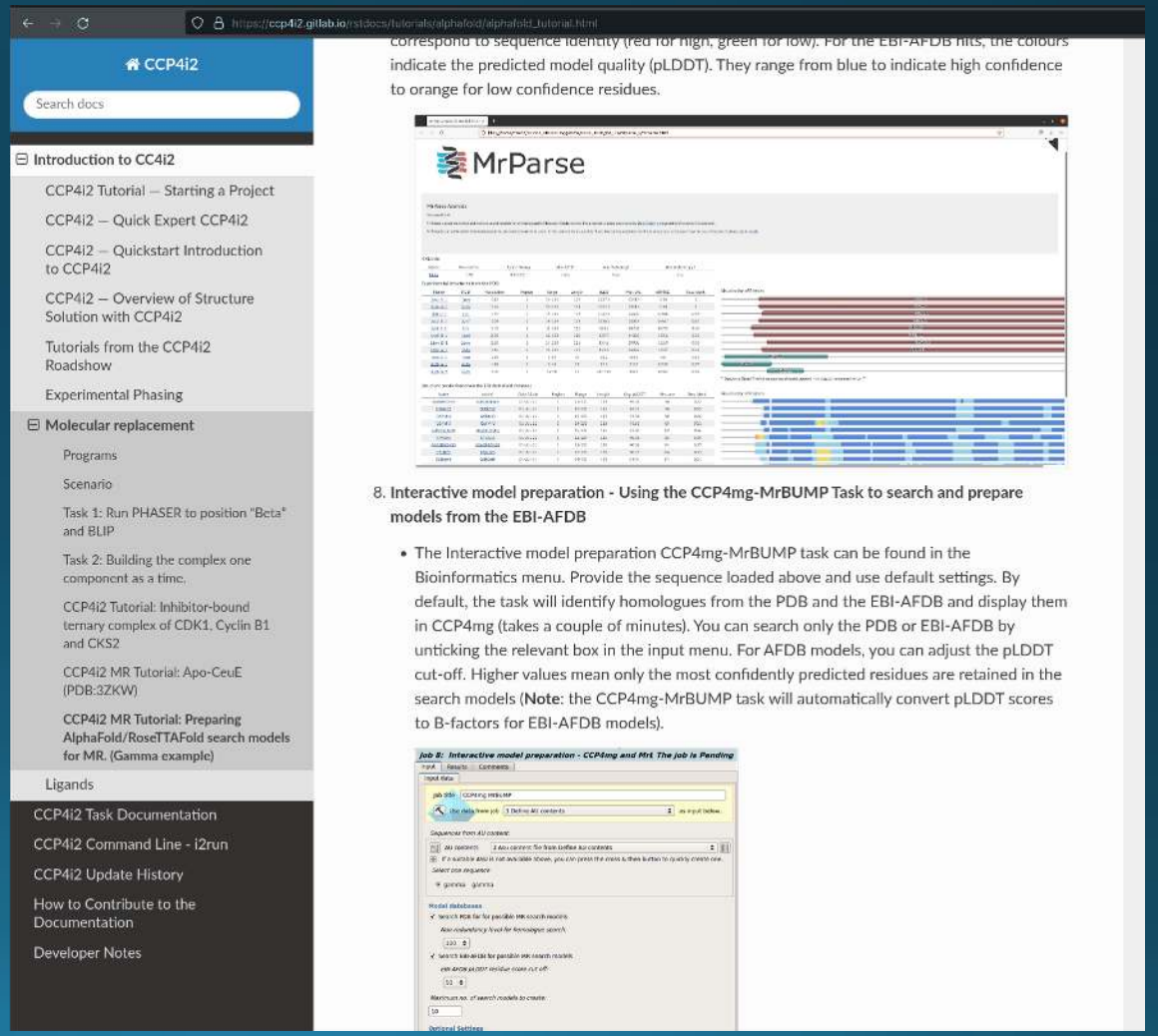
Using predicted models in CCP4

- Finding search models in the EBI-AFDB:
  - CCP4mg-MrBUMP
  - MrParse
- Importing from Colab servers
  - Correcting B-factors
- More from Claudia



Using predicted models in CCP4

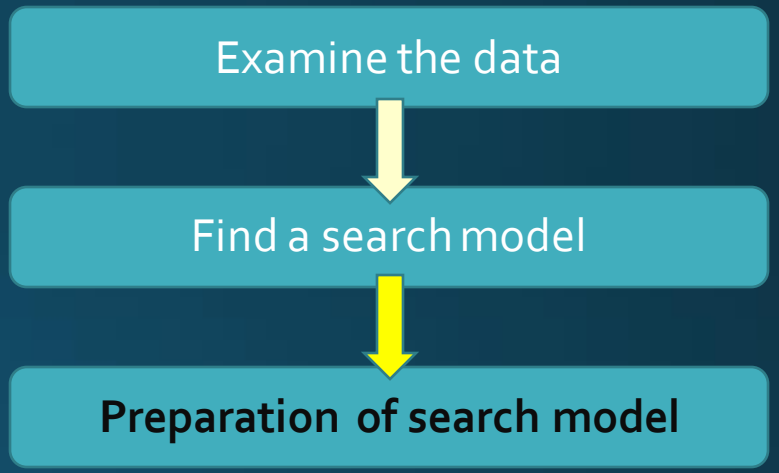
- Tutorials on how to use predicted models like those from Alphafold in CCP4
- https://ccp4i2.gitlab.io/rstdocs/tutorials/alphafold/alphafold_tutorial.html



Preparing Search Models for Molecular Replacement

Preparating models for Molecular Replacement

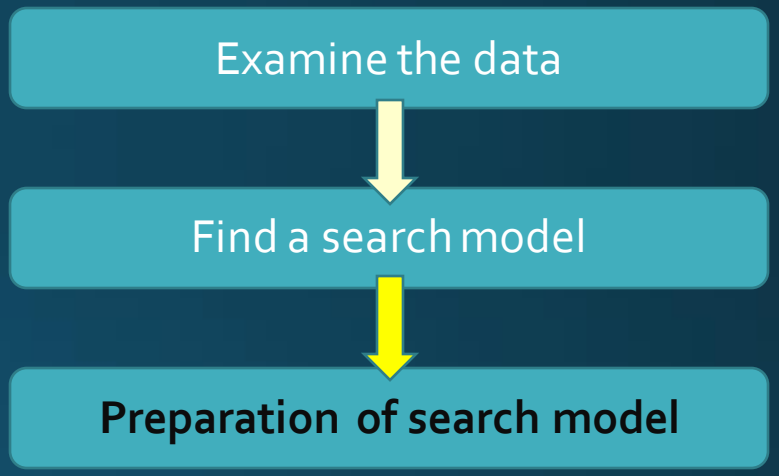
- Once a suitable homologue or set of homologues have been found they need to be prepared for use as search models



Preparating 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

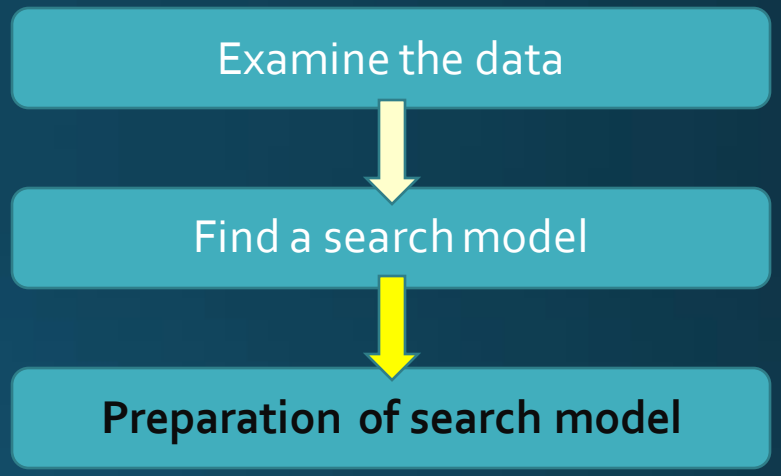
- The closer we can make the homologue to the target in terms of structural similarity the more likely it is to be successfully positioned in Molecular Replacement



CCP4

Preparating 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



CCP4

Preparating models for Molecular Replacement

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

Examine the data

↓

Find a search model

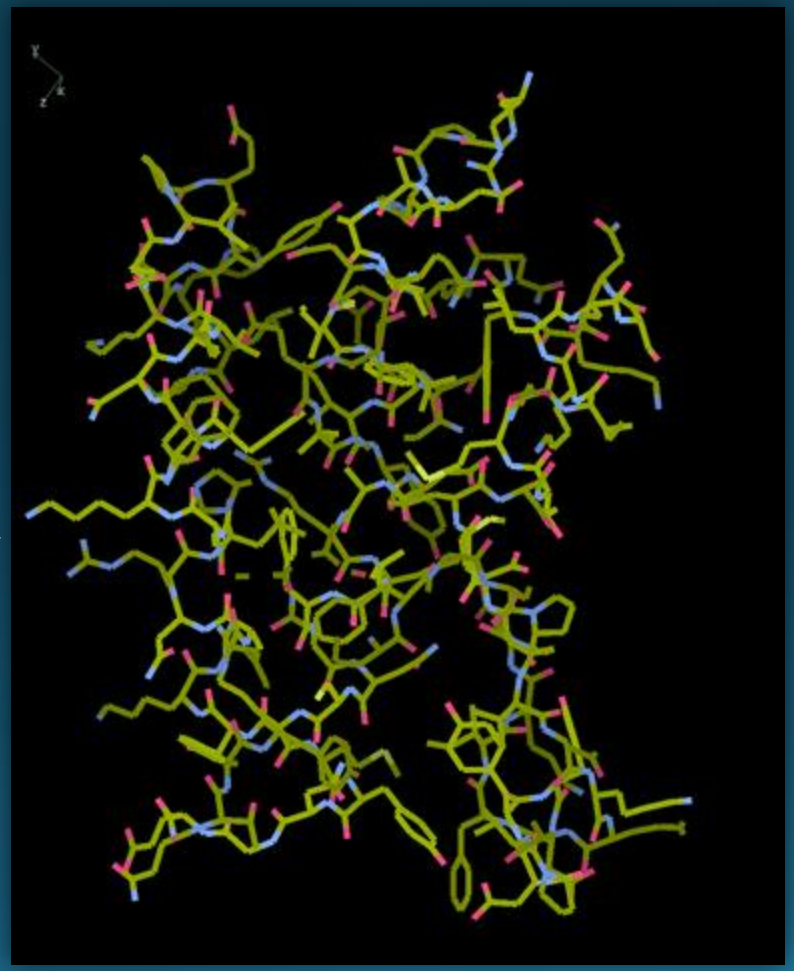
↓

Preparation of search model

CCP4

Preparating 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

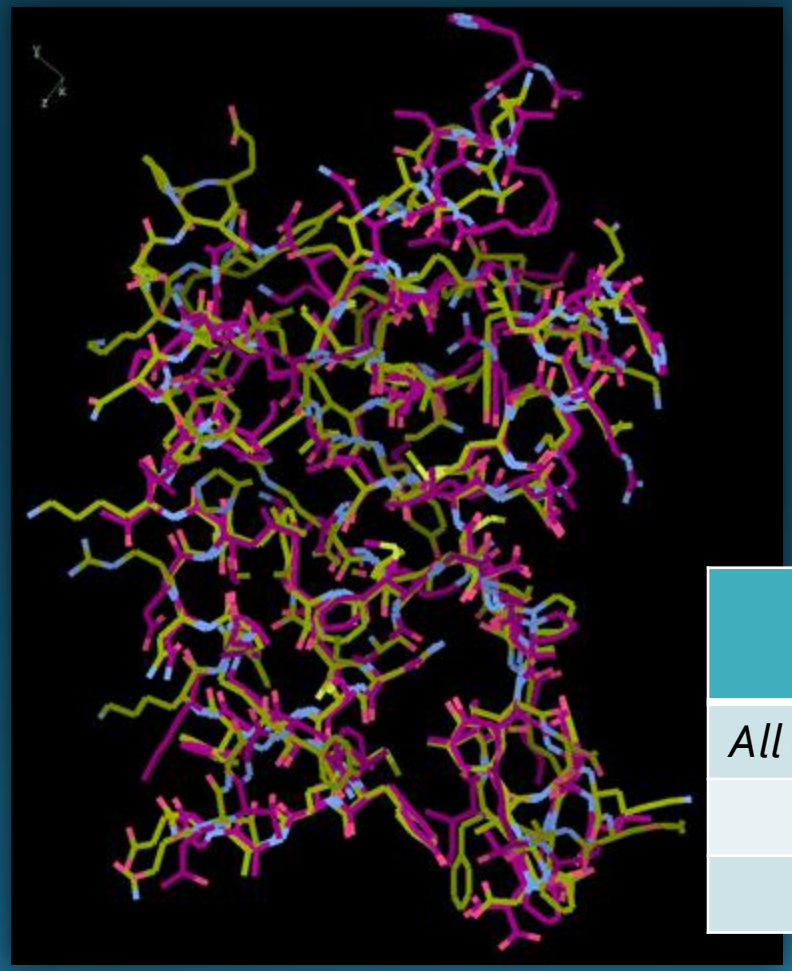
↓

Preparation of search model

CCP4

Preparating models for Molecular Replacement

- Search Model preparation and MR:
  - All atoms?



Examine the data

↓

Find a search model

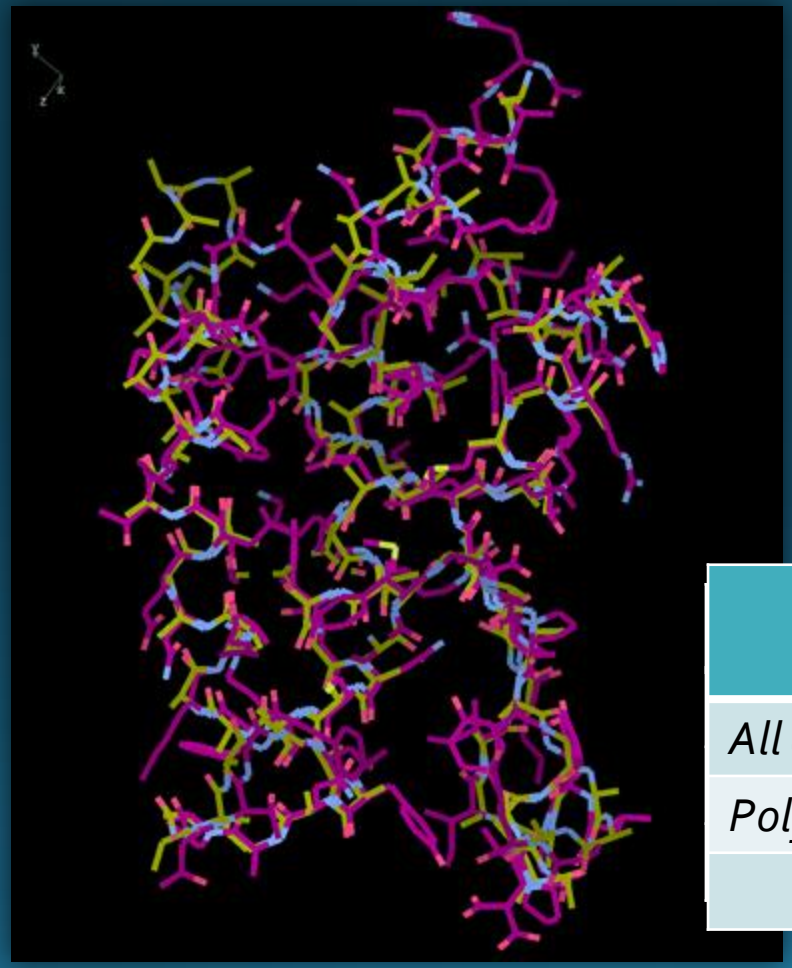
↓

Preparation of search model

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

Preparating models for Molecular Replacement

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



Examine the data

↓

Find a search model

↓

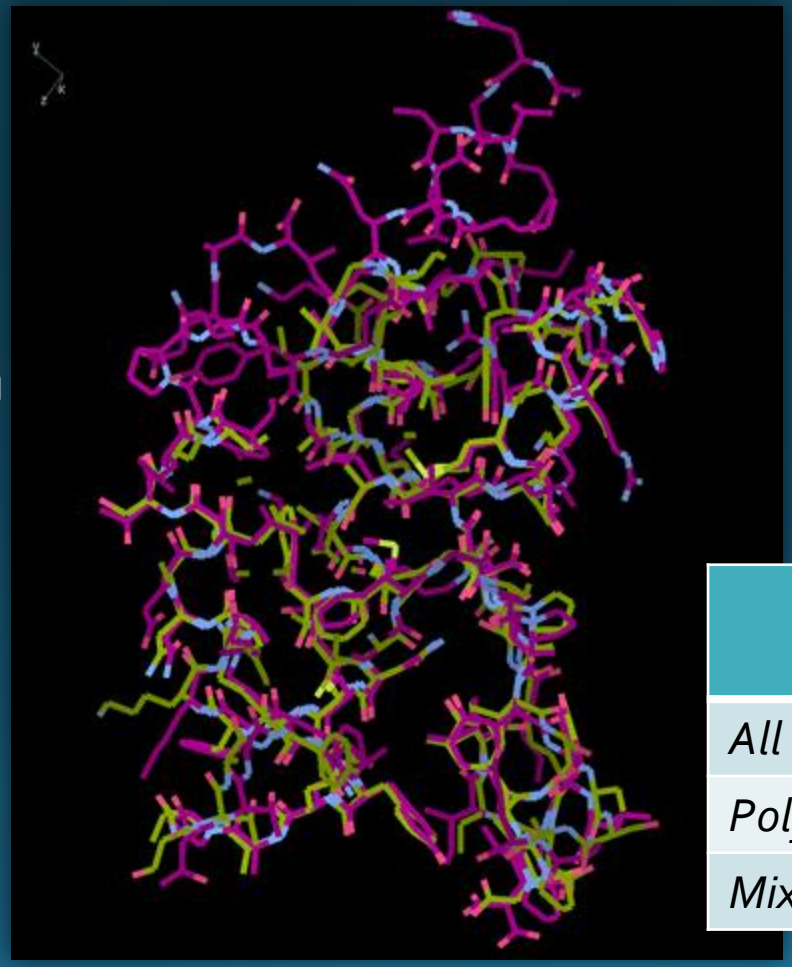
Preparation of search model

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

CCP4

Preparating 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



Examine the data

↓

Find a search model

↓

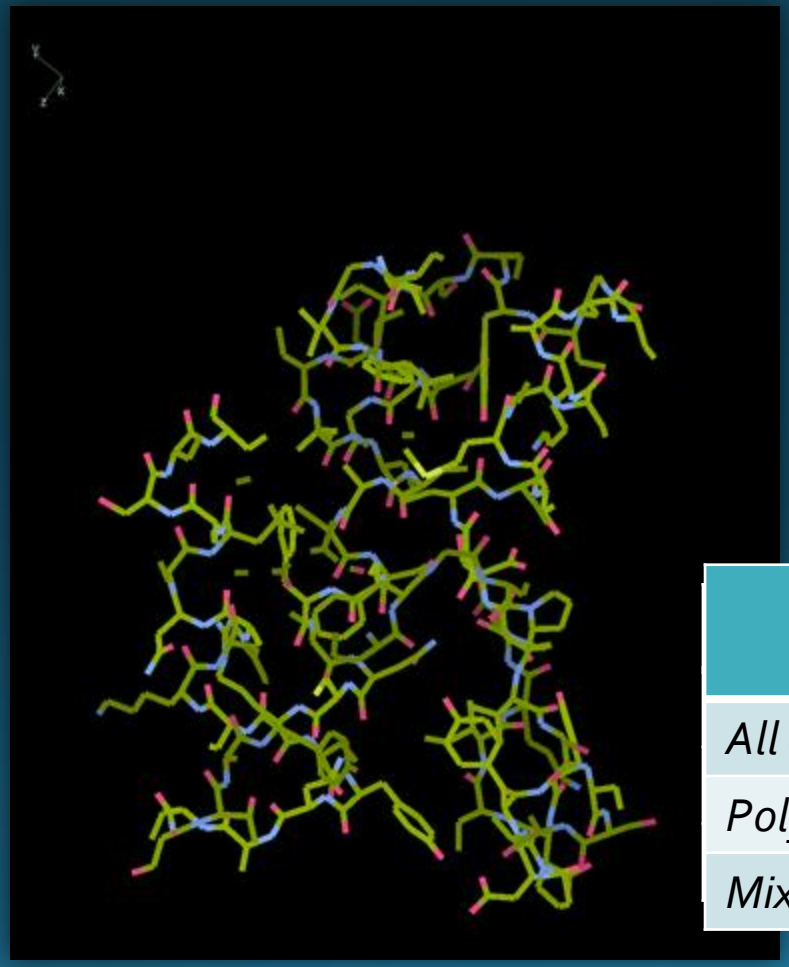
Preparation of search model

	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

Preparating 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



Examine the data

↓

Find a search model

↓

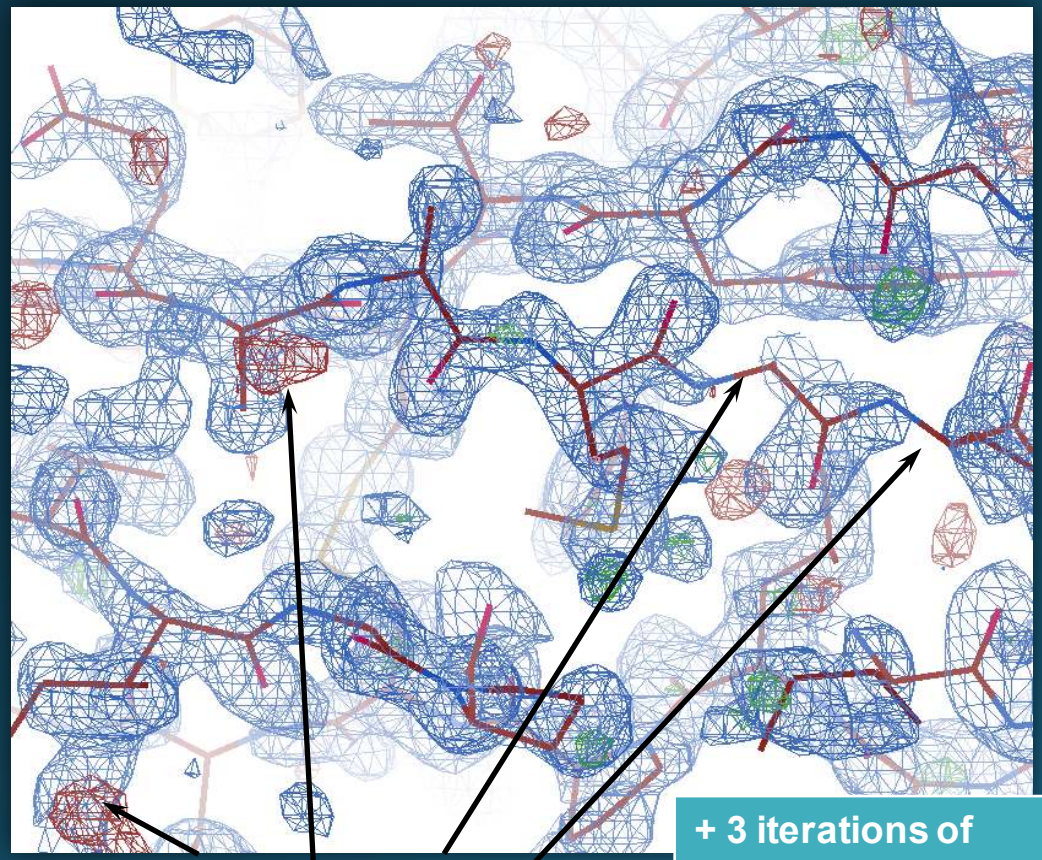
Preparation of search model

	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

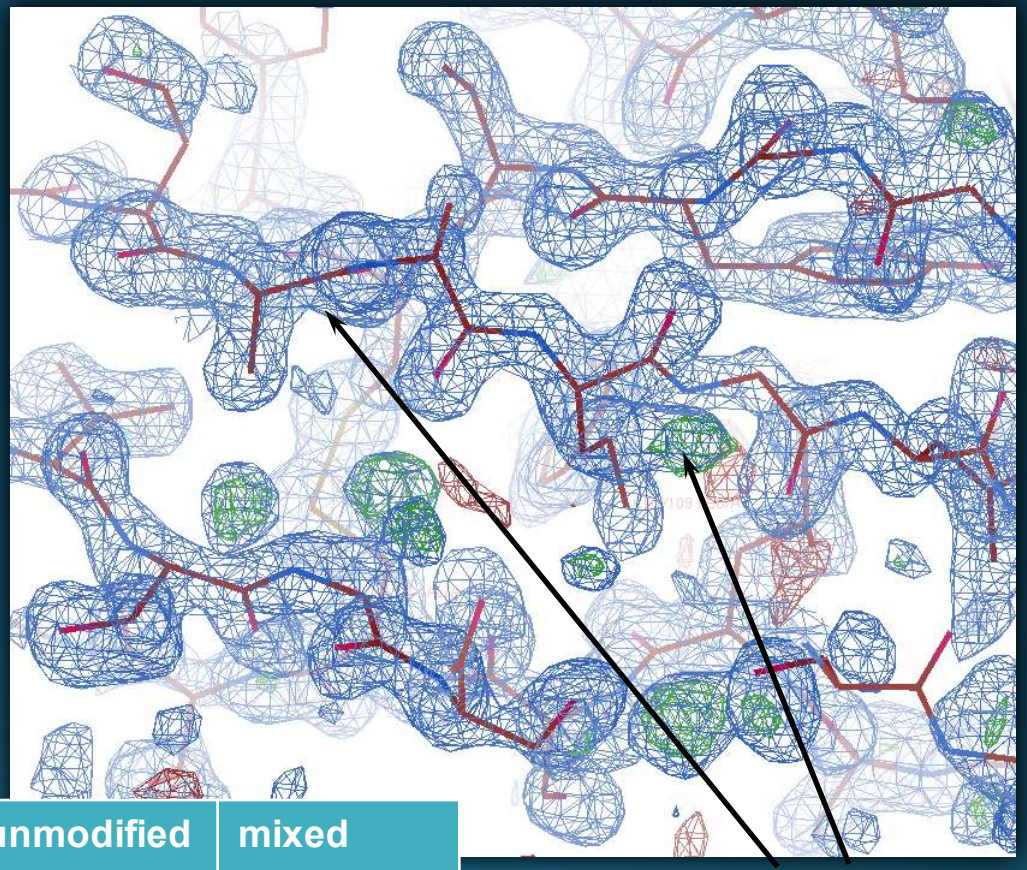
Preparating models for Molecular Replacement

Unmodified model:



bad density features

Mixed model:



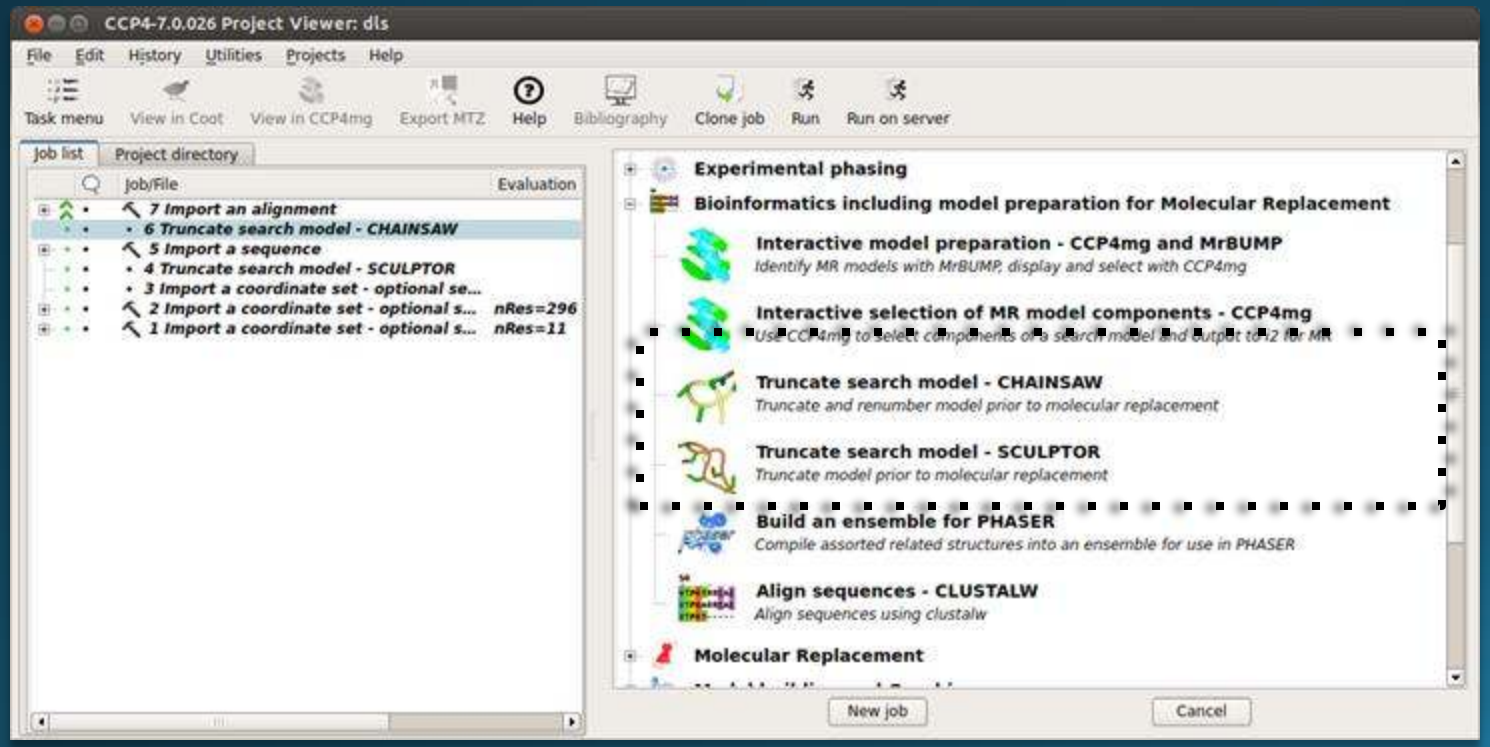
good density features

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

CCP4

Preparating models for Molecular Replacement

- Mixed-model generation in CCP4
  - Chainsaw & Sculptor



Examine the data

↓

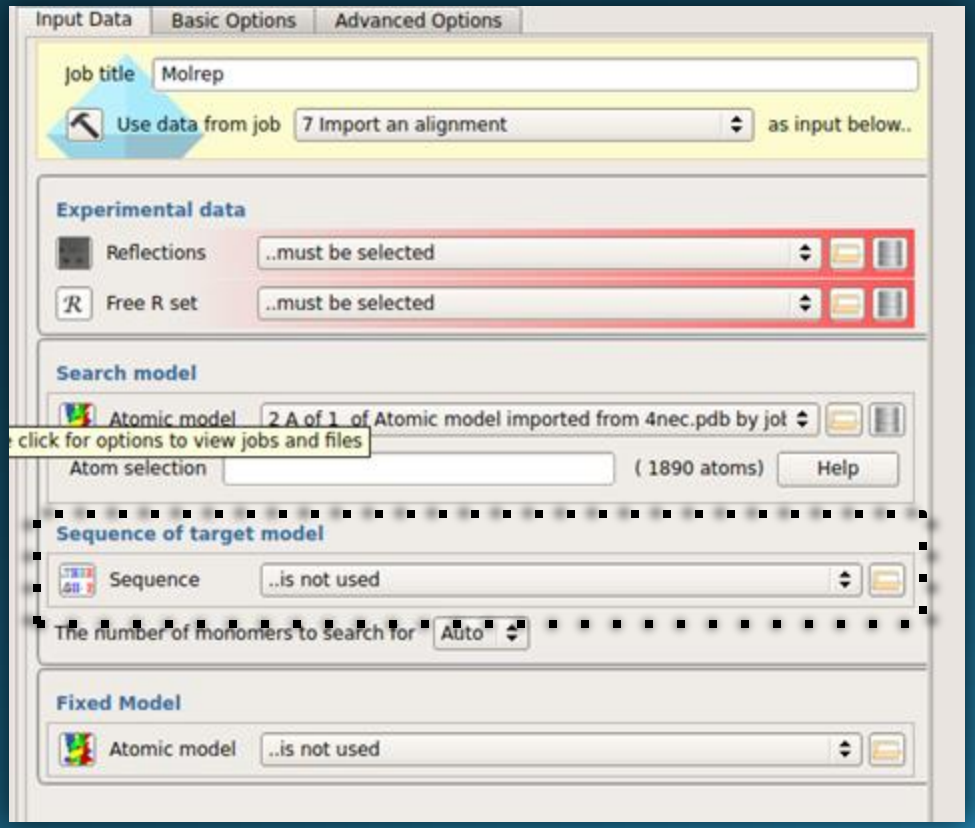
Find a search model

↓

Preparation of search model

Preparating models for Molecular Replacement

- Mixed-model generation in CCP4
  - Molrep



Examine the data

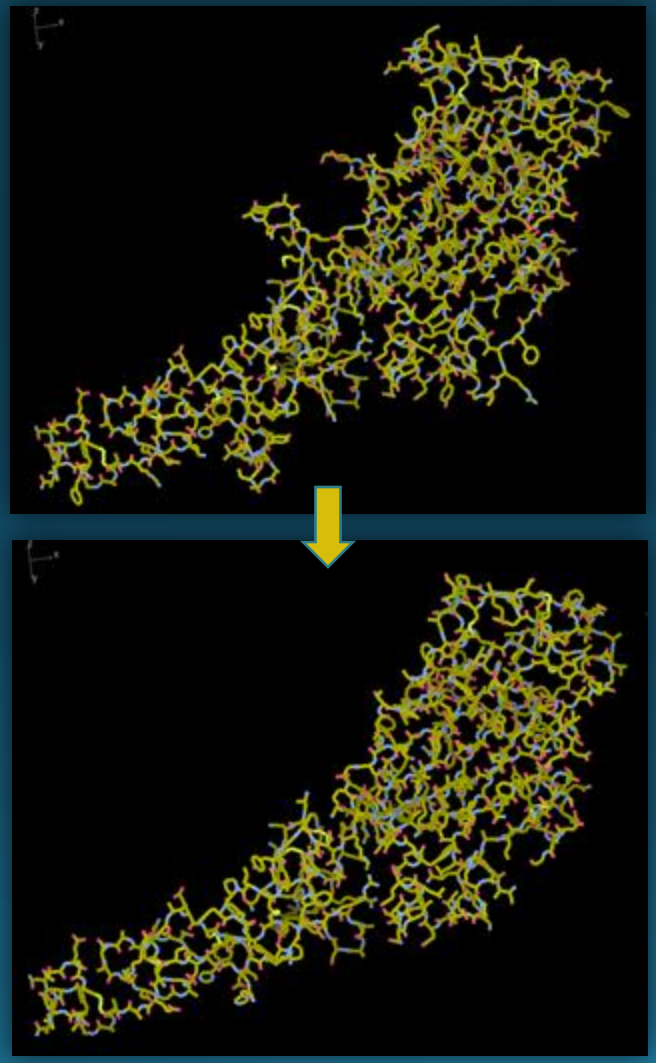
Find a search model

Preparation of search model

CCP4

Preparating 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

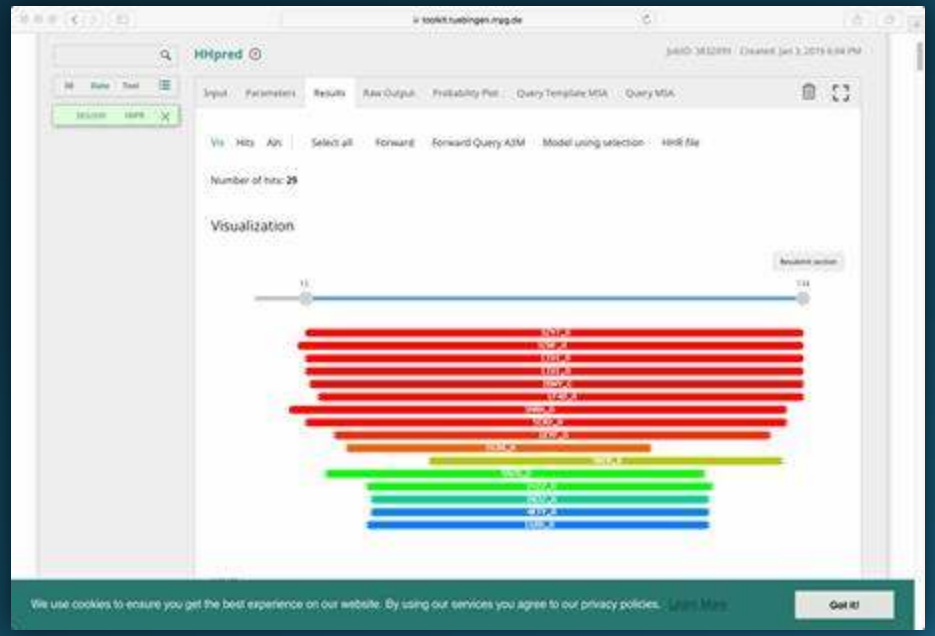
↓

Preparation of search model

CCP4

Preparating models for Molecular Replacement

- Ensembles: alignment of search models : several homologues available



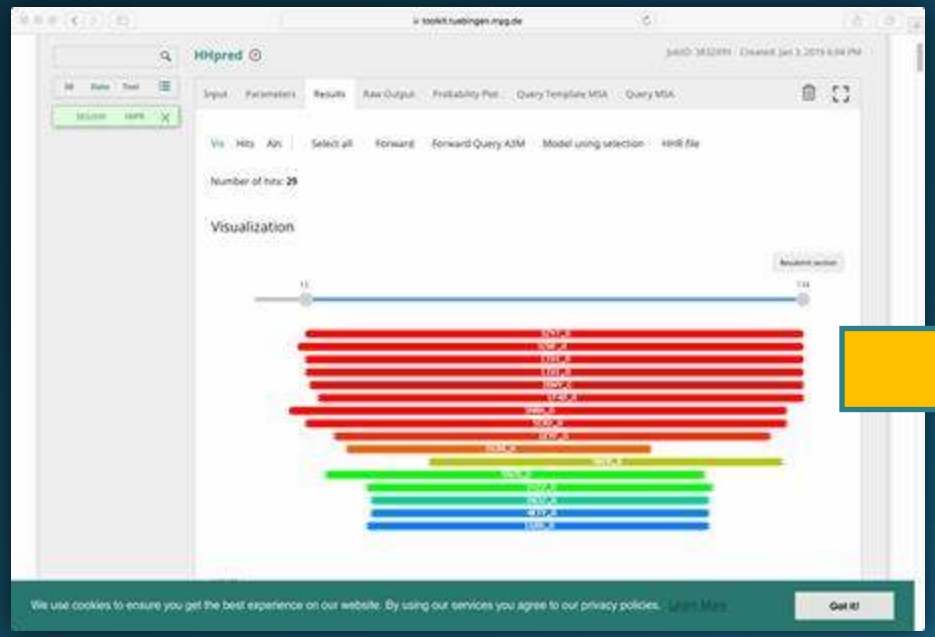
Examine the data

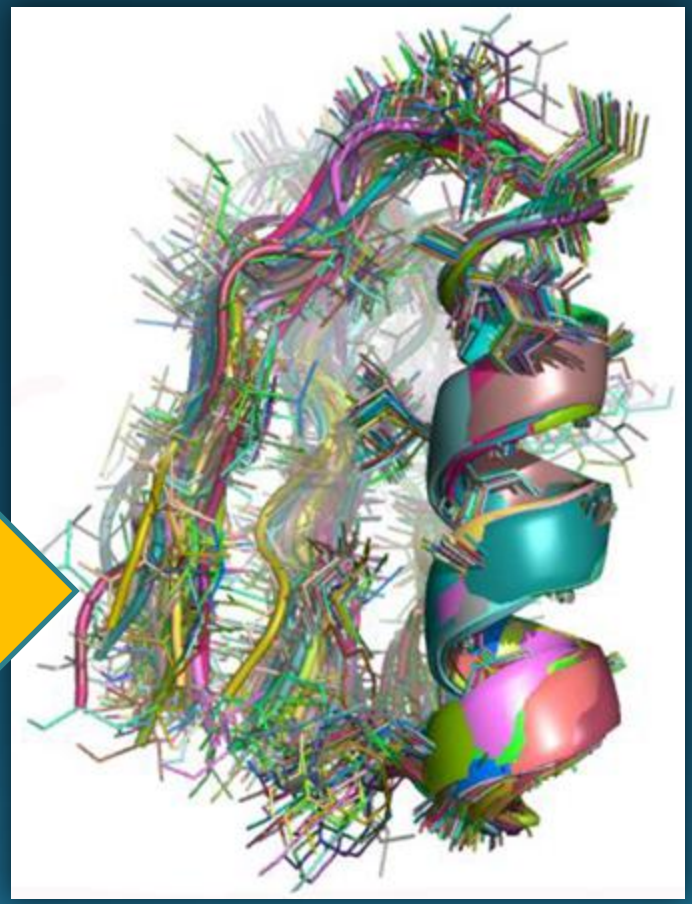
Find a search model

Preparation of search model

Preparating models for Molecular Replacement

- Ensembles: alignment of search models





Examine the data

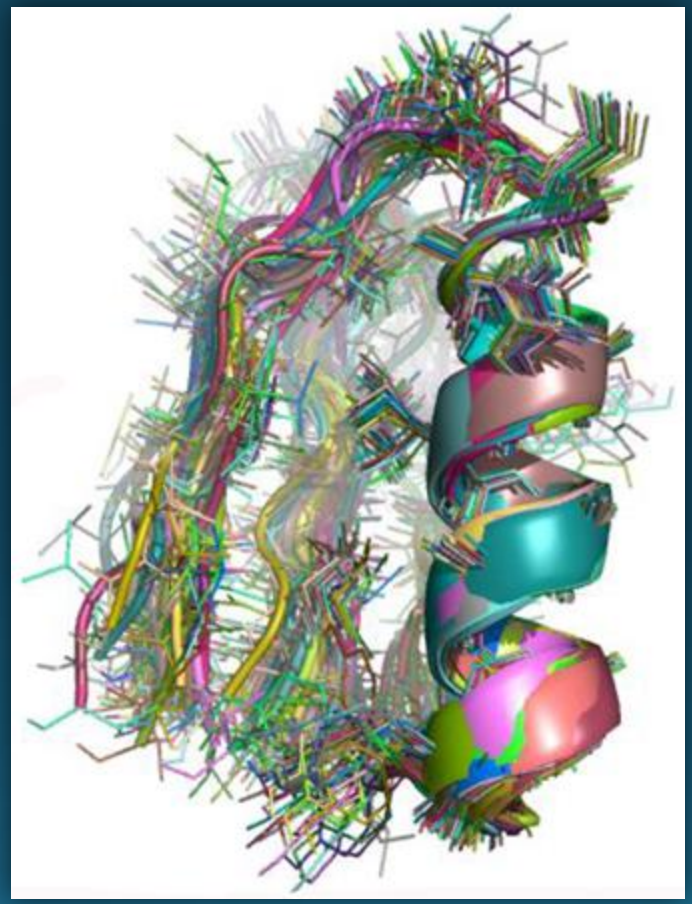
Find a search model

Preparation of search model

CCP4

Preparating 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

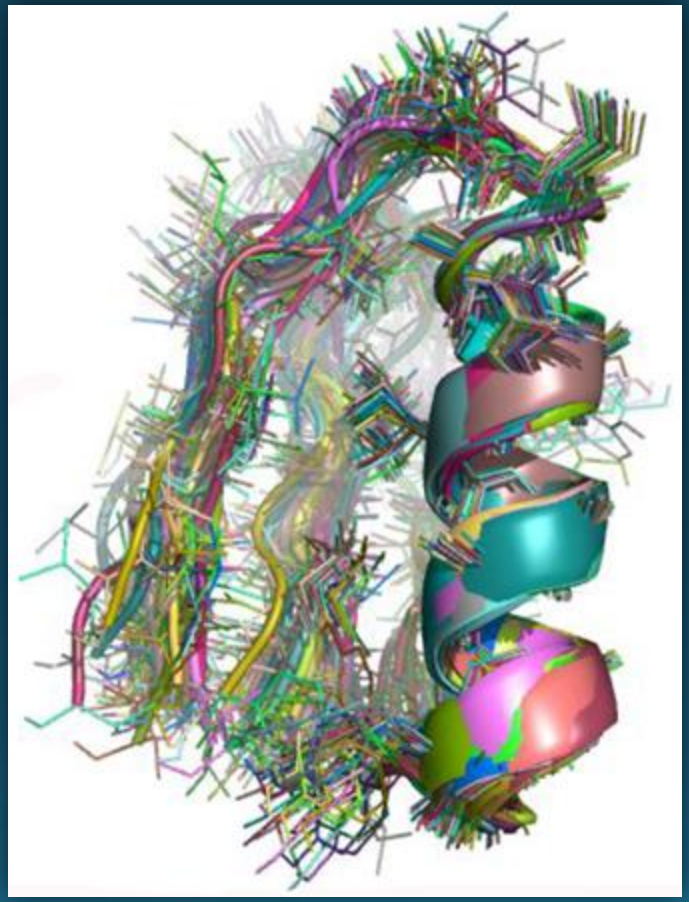
↓

Preparation of search model

CCP4

Preparating 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

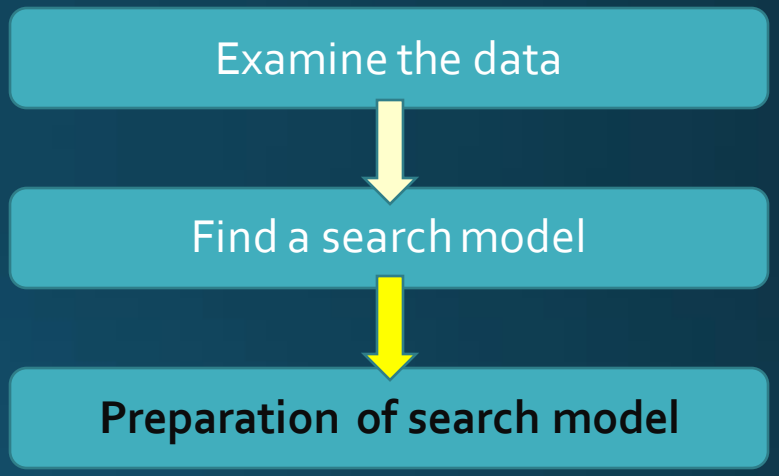
↓

Preparation of search model

CCP4

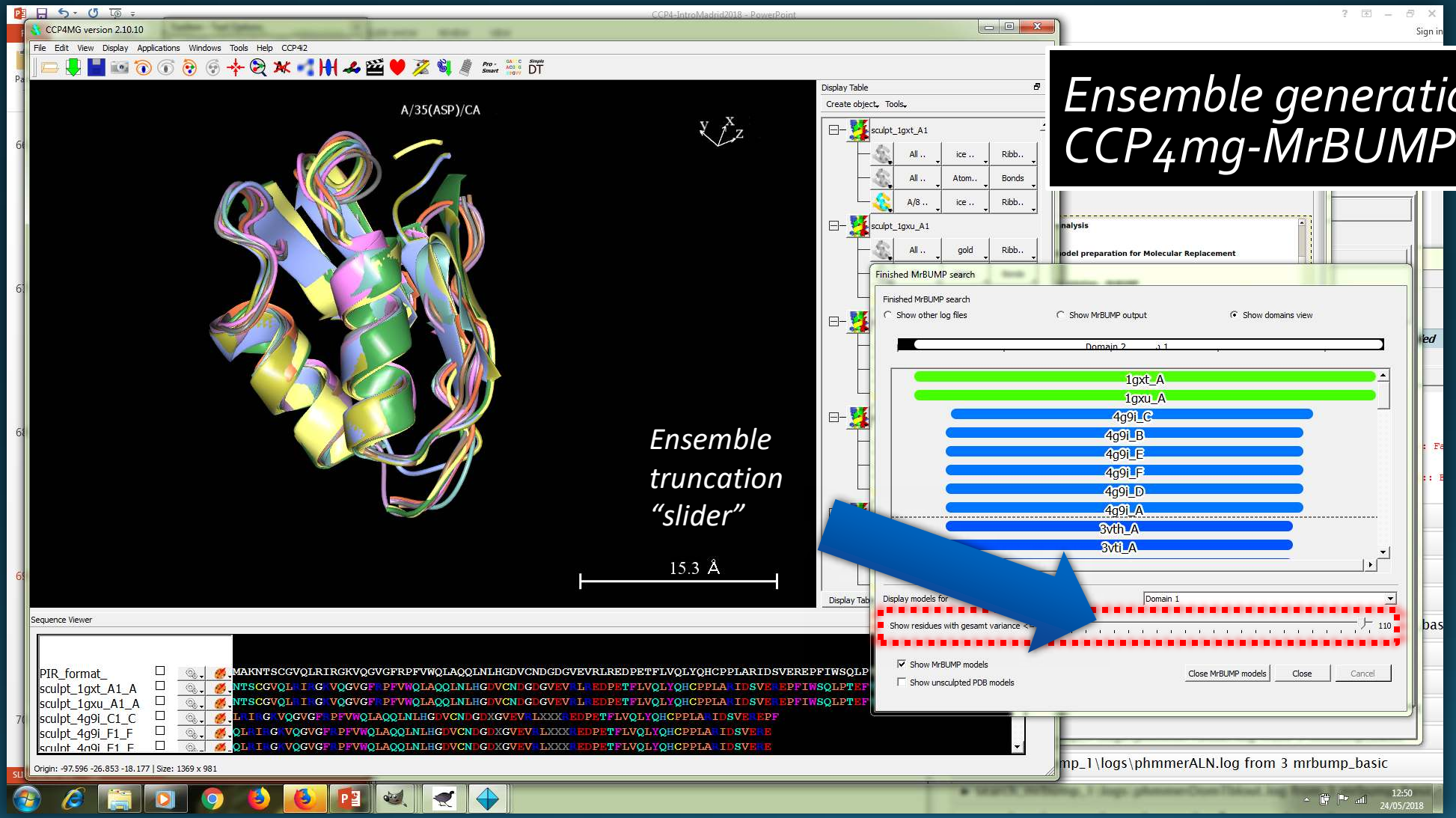
Preparating models for Molecular Replacement

- 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



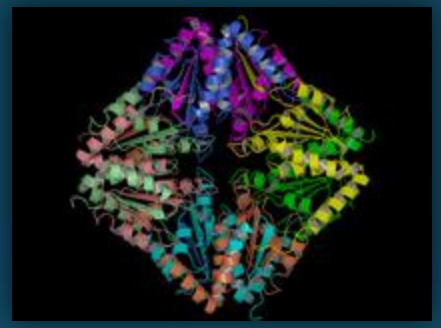
CCP4

Preparating models for Molecular Replacement



Preparaing 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

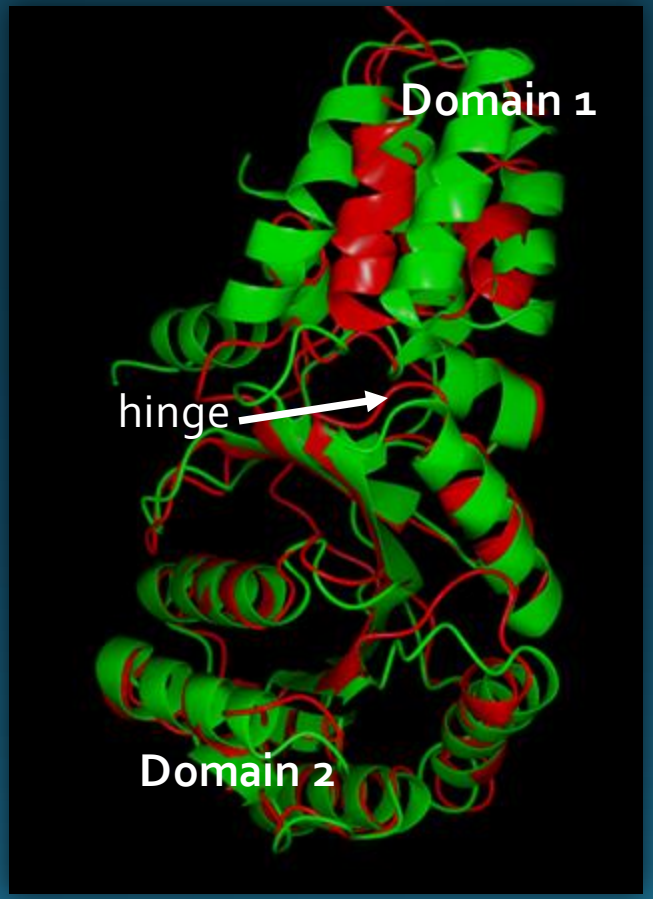
↓

Preparation of search model

CCP4

Preparating 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



Examine the data

Find a search model

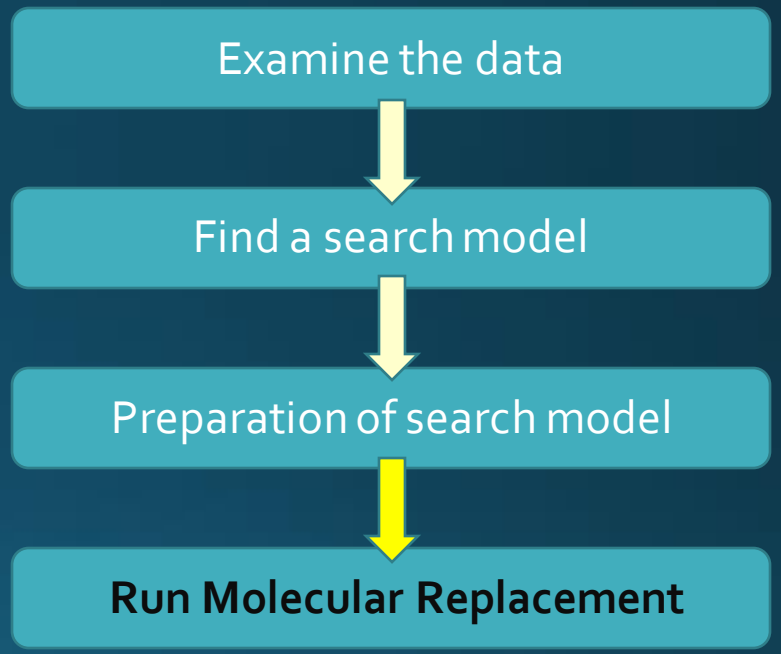
Preparation of search model

CCP4

Running Molecular Replacement

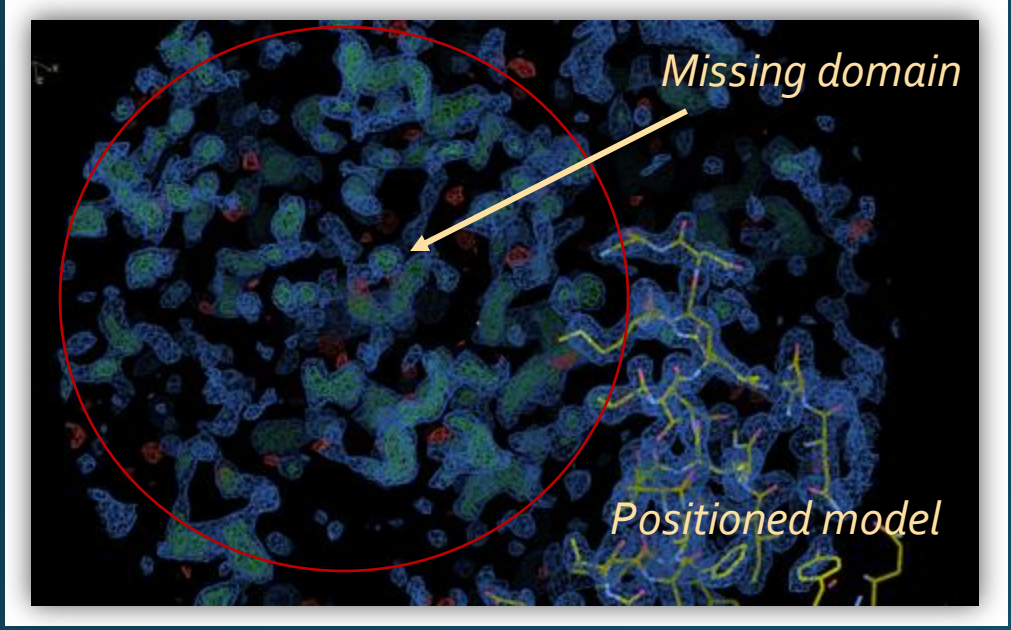
Molecular Replacement

- 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



CCP4

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

Examine the data

Find a search model

Preparation of search model

Run Molecular Replacement

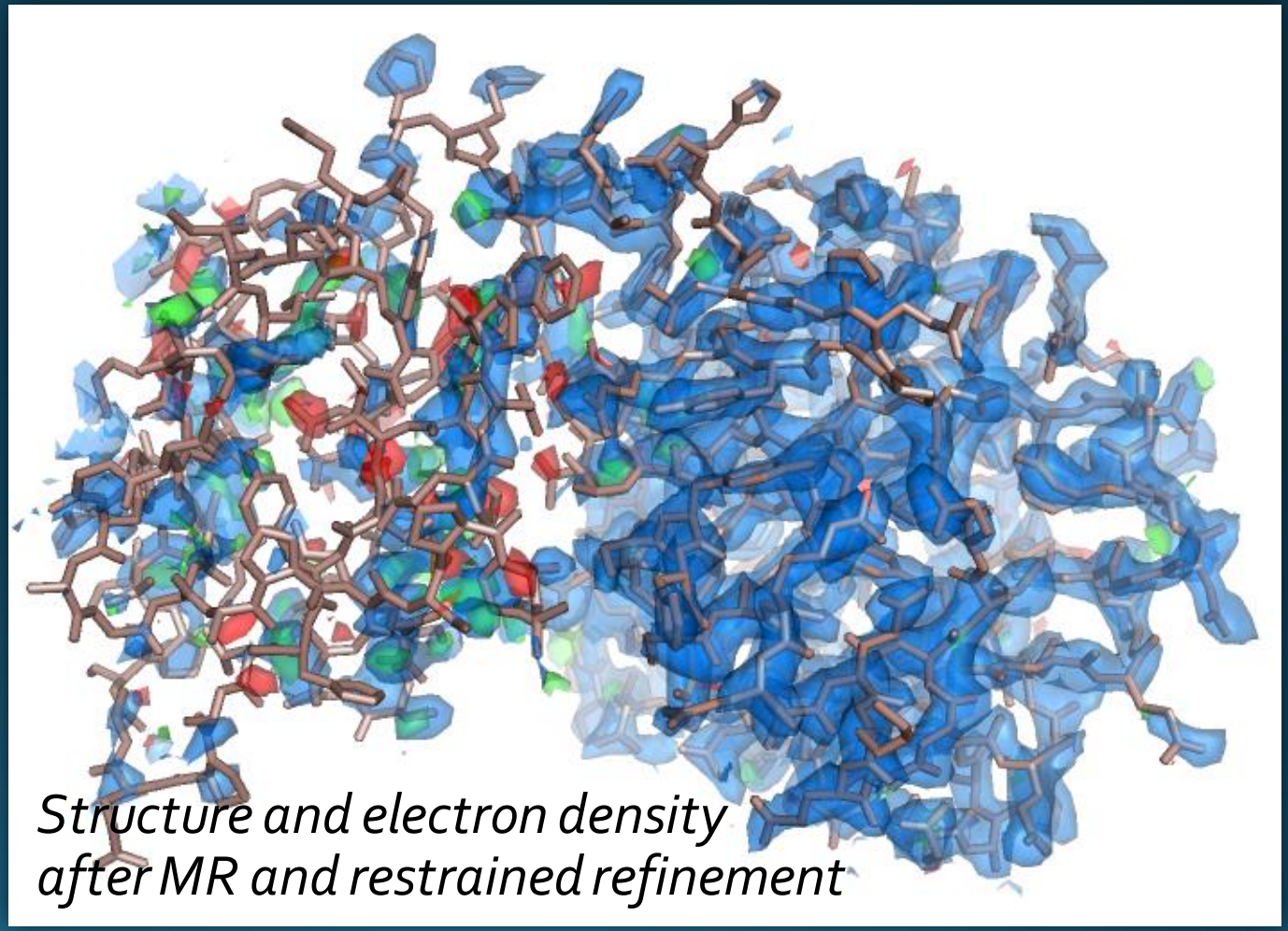
CCP4

Molecular Replacement

Phased translation search:

Example: ***1tj3***

Search model: 1s20, chain A



Structure and electron density after MR and restrained refinement

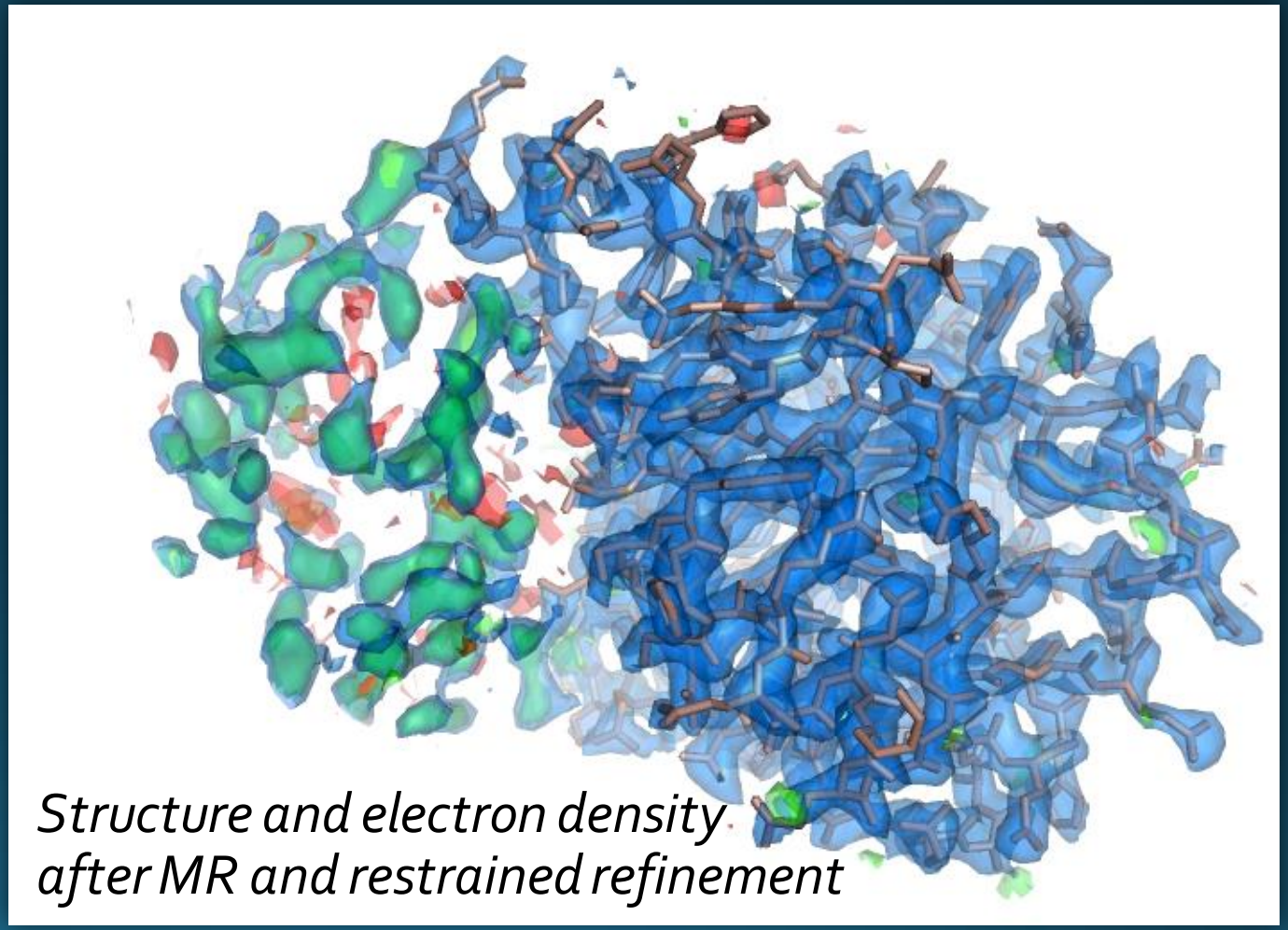
Molecular Replacement

Phased translation search:

Example: ***1tj3***

Search model: 1s20, chain A

Large Domain only



Structure and electron density after MR and restrained refinement

CCP4

Molecular Replacement

Phased translation search:

Example: ***1tj3***

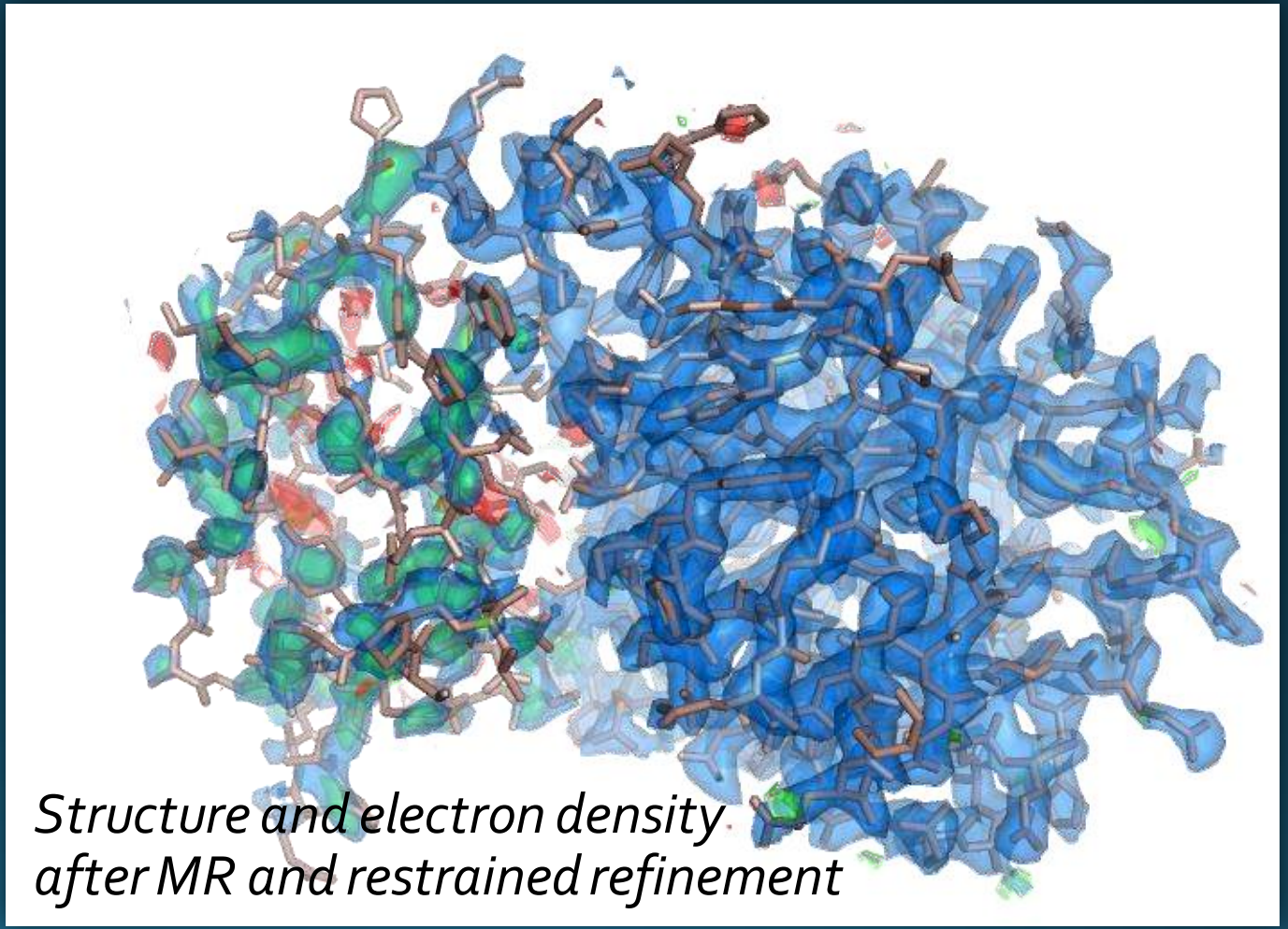
Search model: 1s20, chain A

Large Domain fixed

Small Domain search

Fit into density

(Molrep & Phaser)



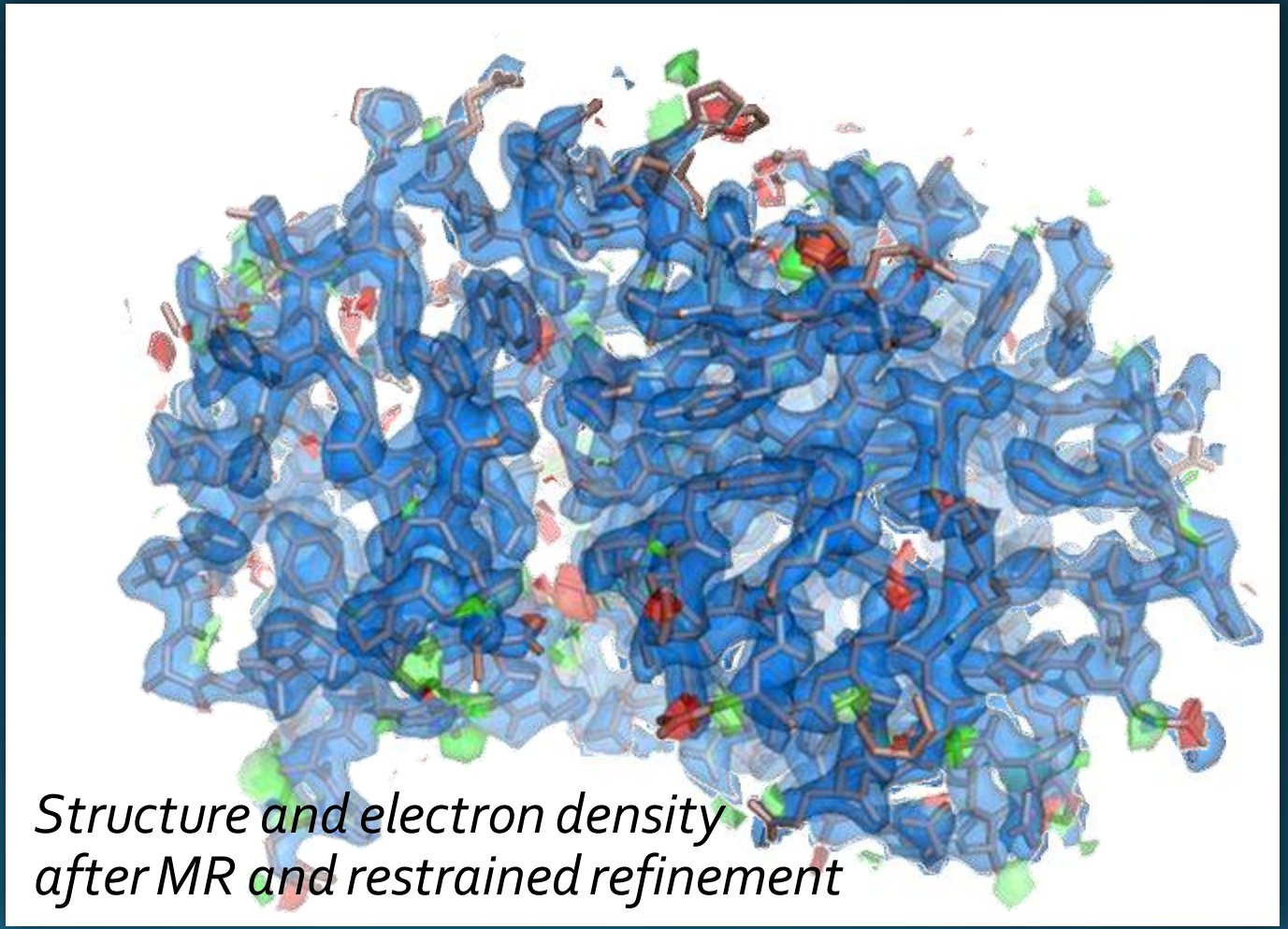
Structure and electron density after MR and restrained refinement

Molecular Replacement

Phased translation search:

Example: ***1tj3***

Complete structure and electron density after restrained refinement

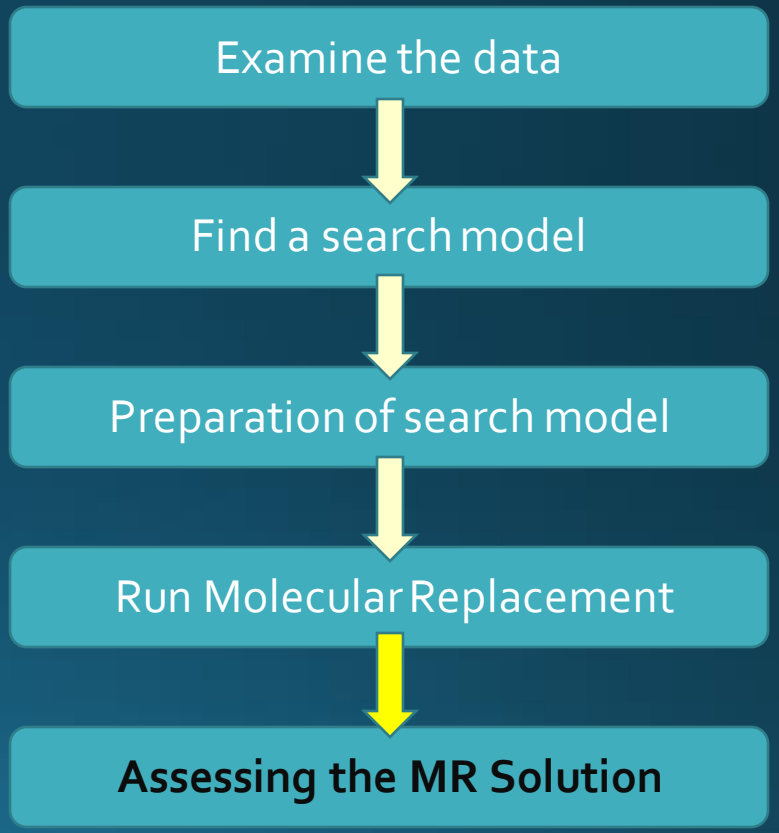


Structure and electron density after MR and restrained refinement

How do I know my MR solution is successful?

Assessing the solution

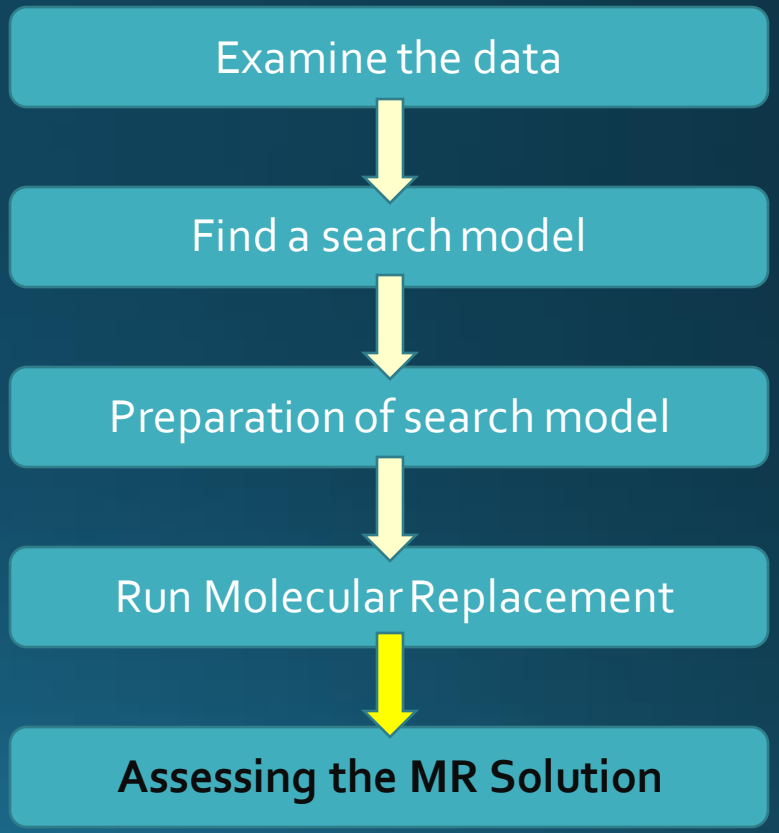
- Terminology
  - MR solution
  - Successful MR solution
- What is a successful MR solution?
  - A search model placed in the target unit cell such that it will provide us with sufficiently accurate phase estimates for the target
  - Enables us to complete the model through model building and refinement



CCP4

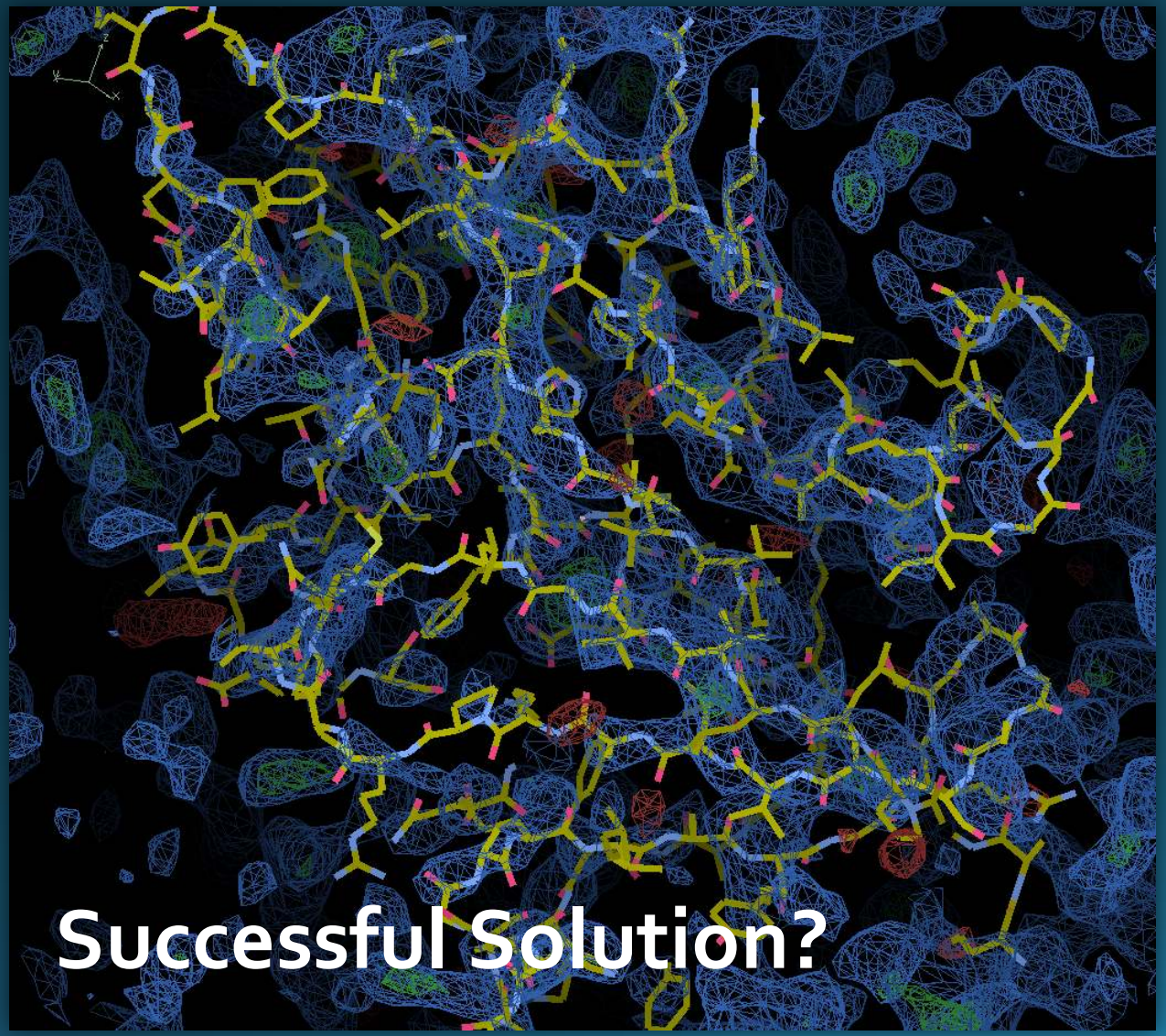
Assessing the solution

- 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



CCP4

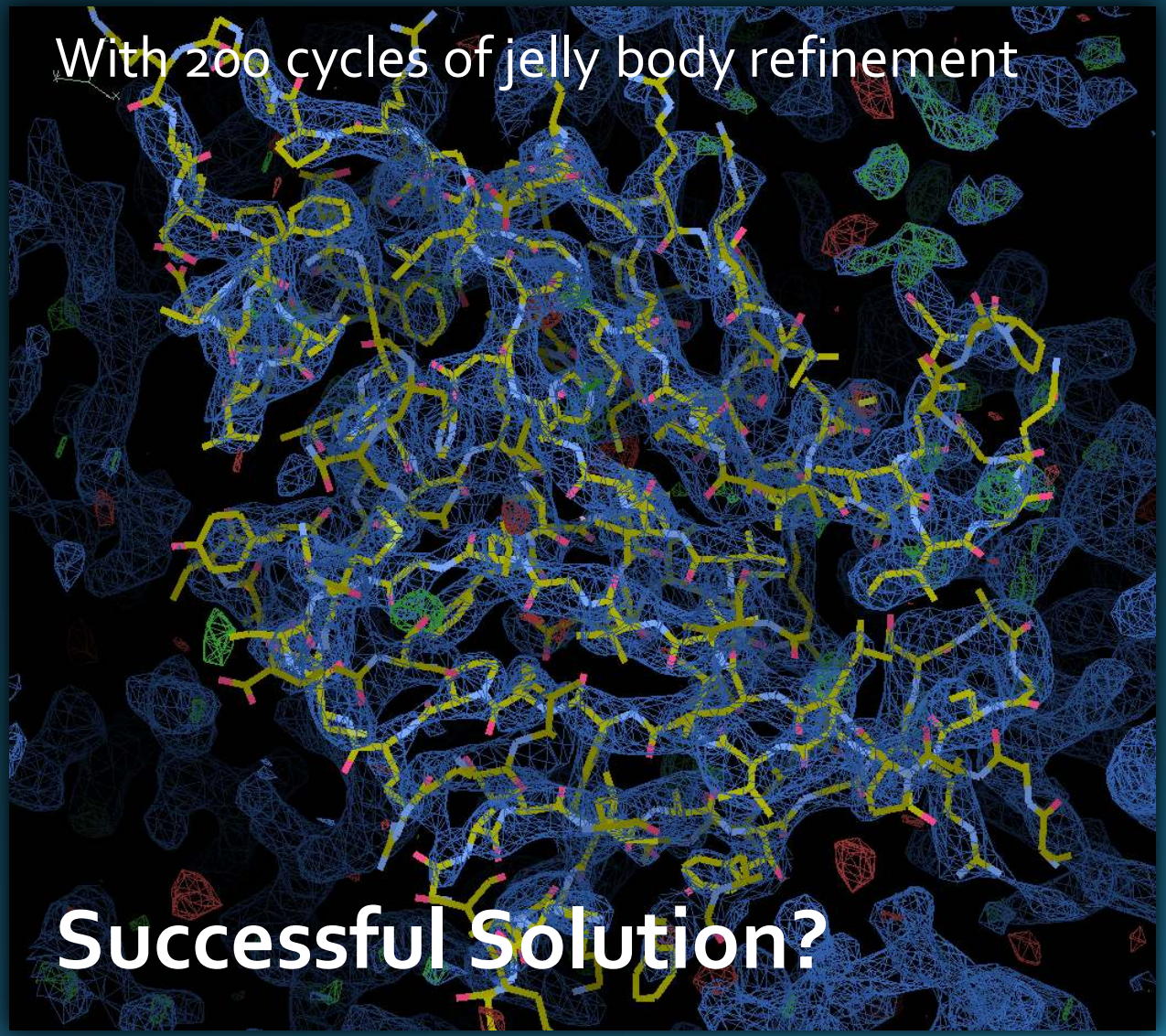
Assessing the solution



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

CCP4

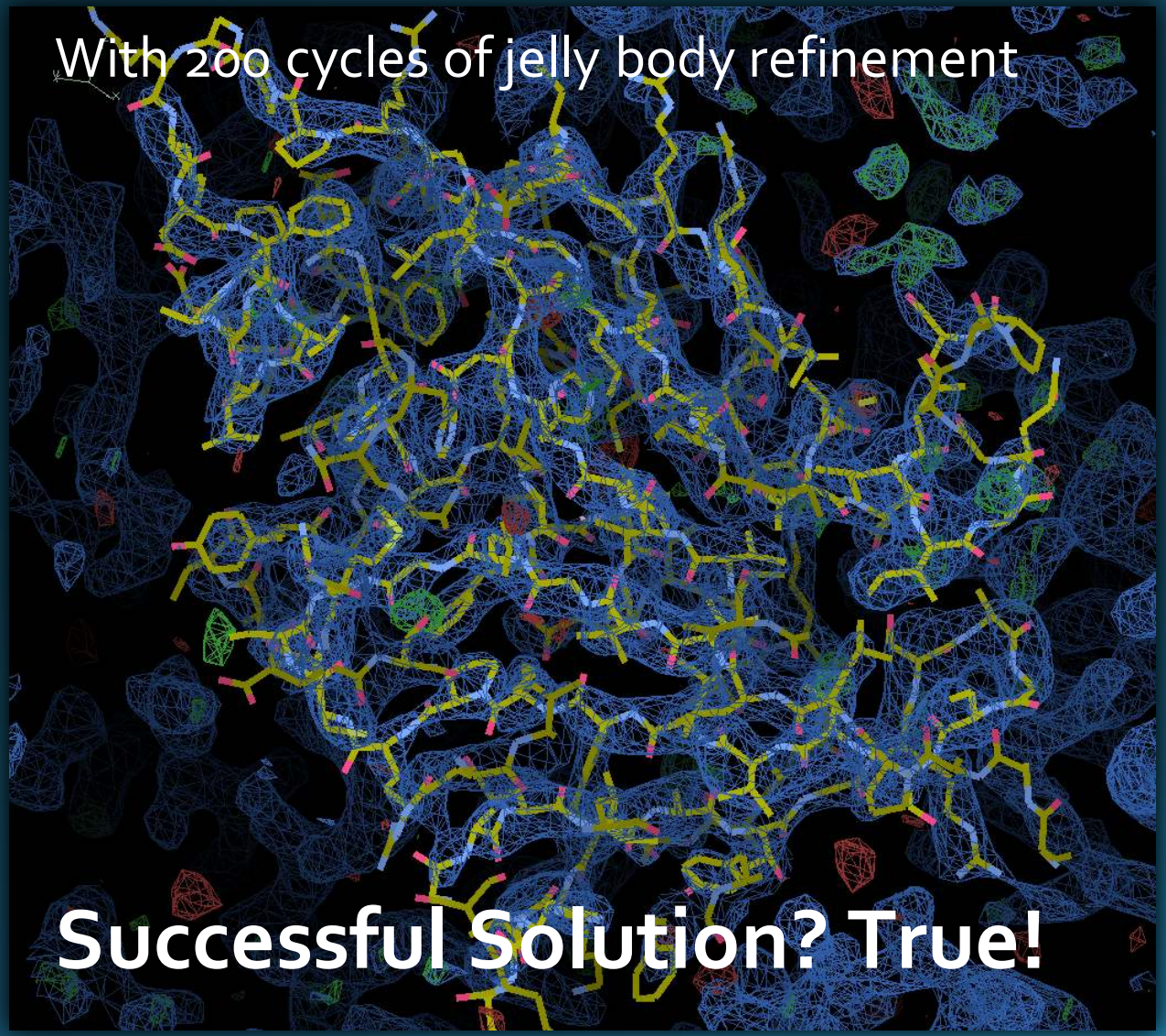
Assessing the solution



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

CCP4

Assessing the solution

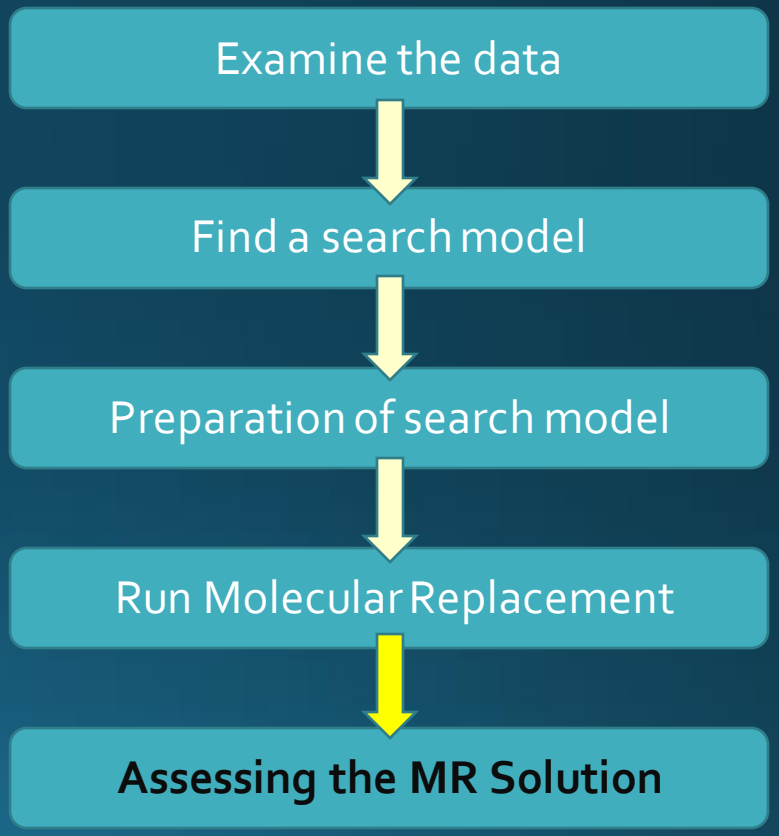


- Examine the data
- Find a search model
- Preparation of search model
- Run Molecular Replacement
- **Assessing the MR Solution**

CCP4

Assessing the solution

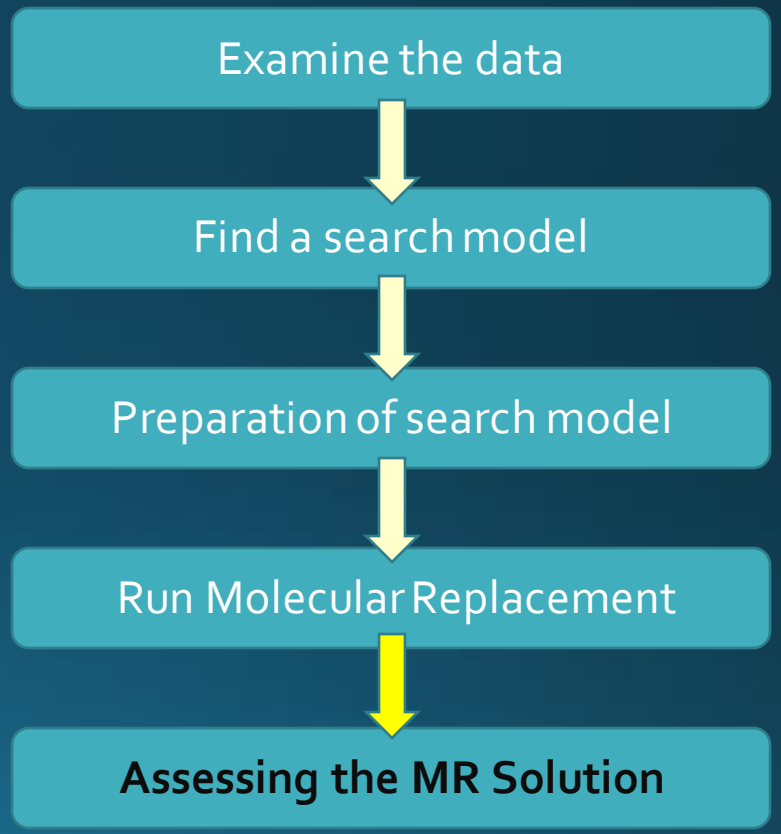
- Rough guide to MR program scoring



CCP4

Assessing the solution

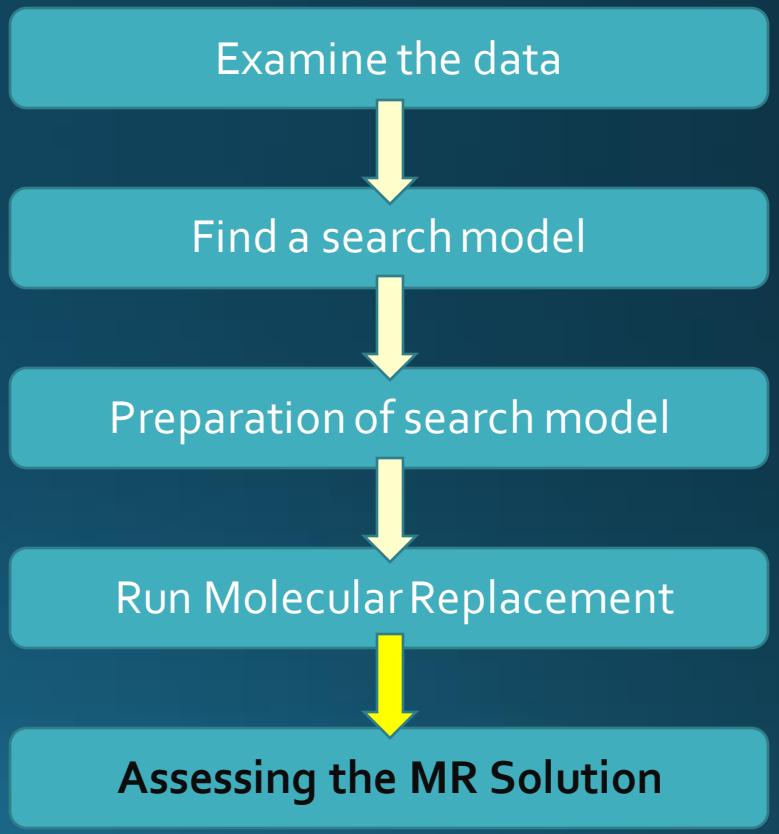
- Rough guide to MR program scoring
  - Phaser scores
    - LLG scores – has it increased by 6σ or more after the placement of a new molecule?
      (resolution and space group dependent)
    - TFZ – greater than 8?
    - Few or single solution almost always indicative of success



CCP4

Assessing the solution

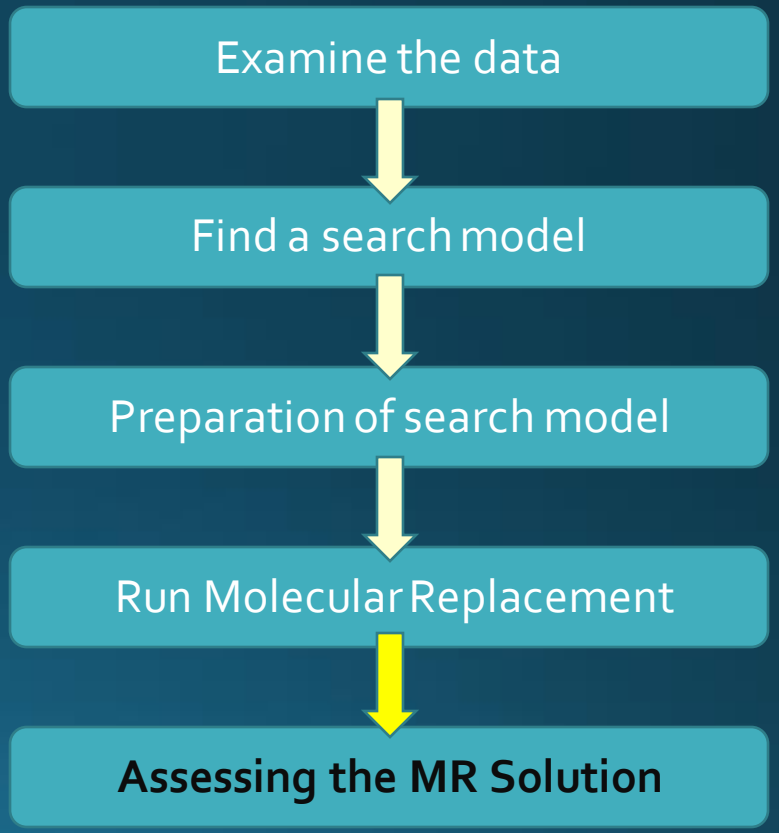
- Rough guide to MR program scoring
  - Phaser scores
    - LLG scores – has it increased by 60 or more after the placement of a new molecule?
      (resolution and space group dependent)
    - TFZ – greater than 8?
    - Few or single solution almost always indicative of success
  - Molrep scores
 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



CCP4

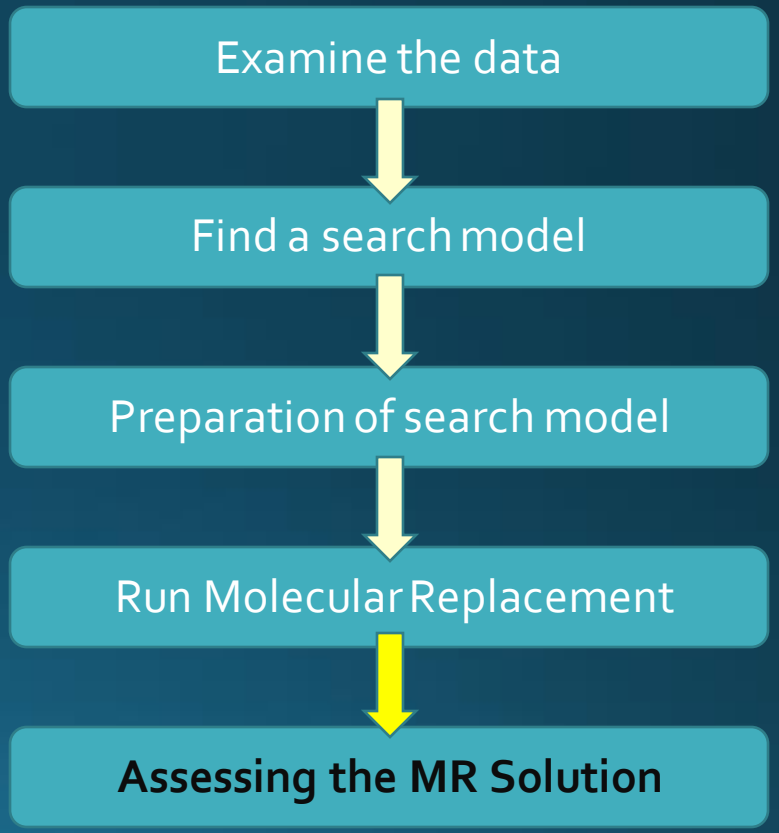
Assessing the solution

- Refinement



Assessing the solution

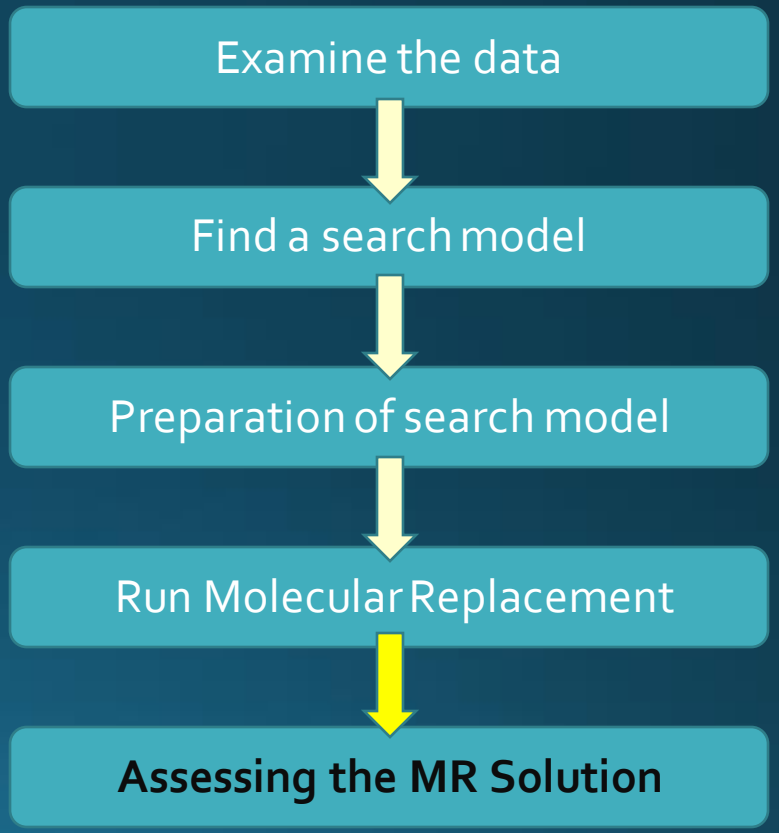
- Refinement
  - Look at Rfactor/Rfree
    - are they falling? Is Rfree below 0.5?



CCP4

Assessing the solution

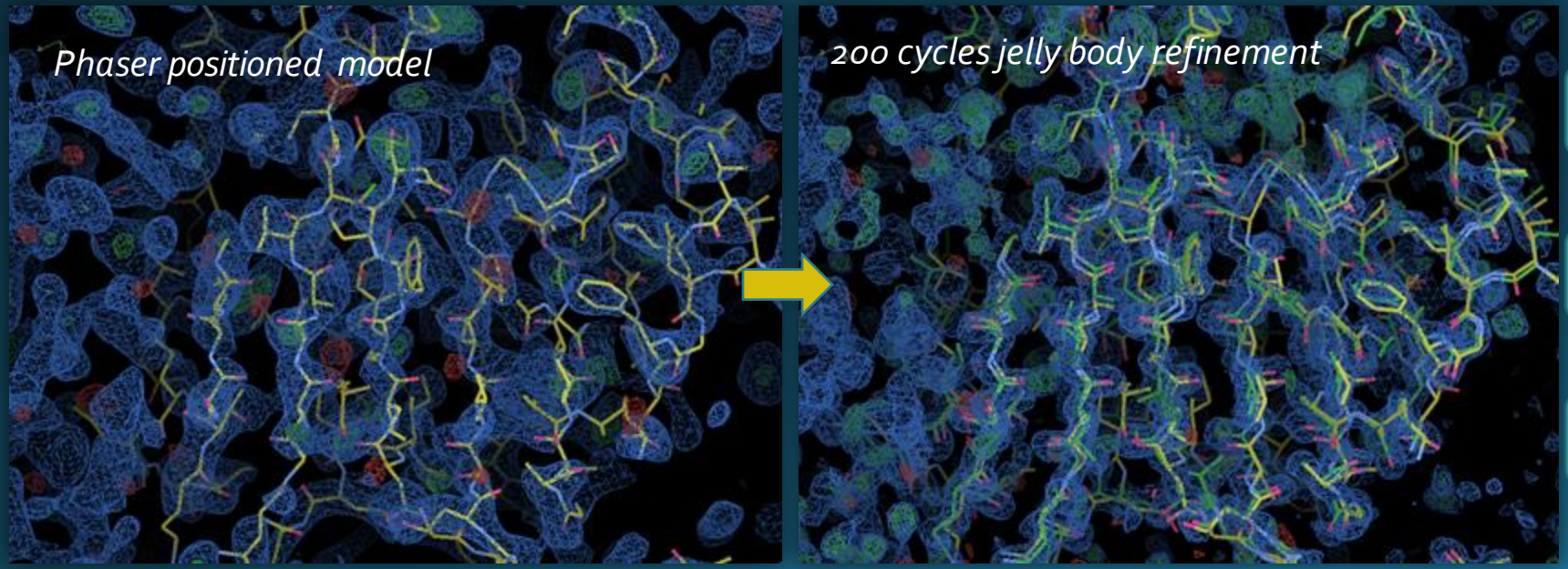
- Refinement
  - Look at Rfactor/Rfree
    - are they falling? Is Rfree below 0.5?
  - Use 200 cycles of jelly body refinement option in Refmac post MR



CCP4

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



Examine the data

Find a search model

Preparation of search model

Run Molecular Replacement

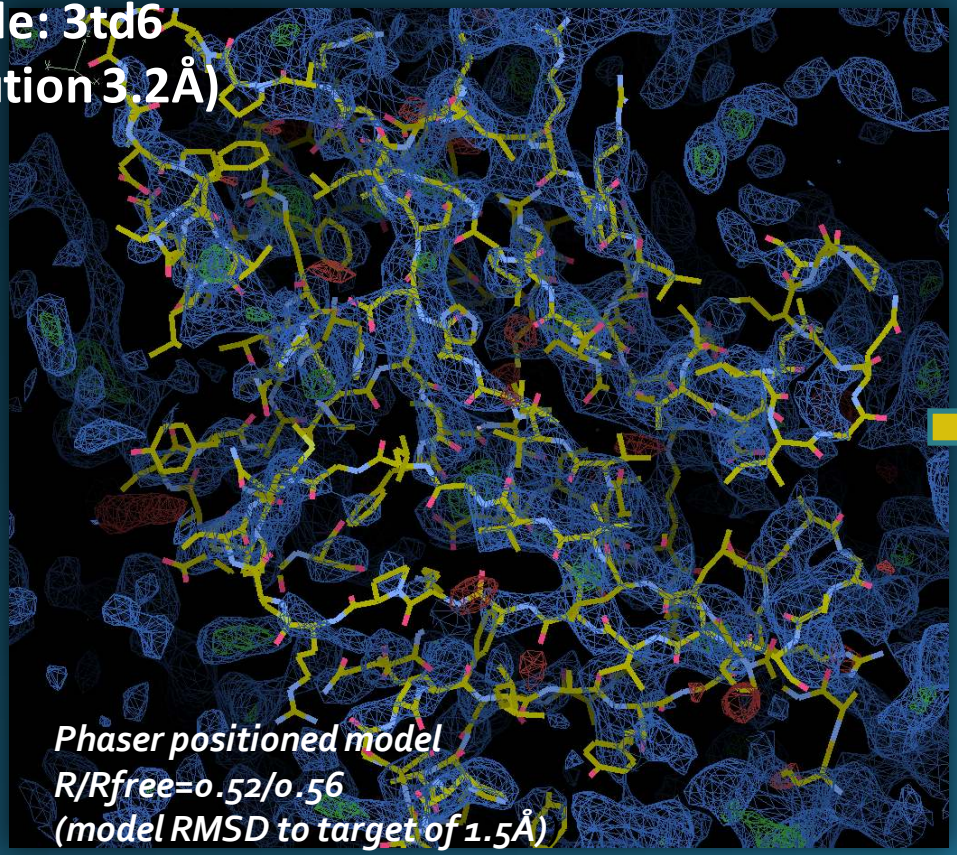
Assessing the MR Solution

CCP4

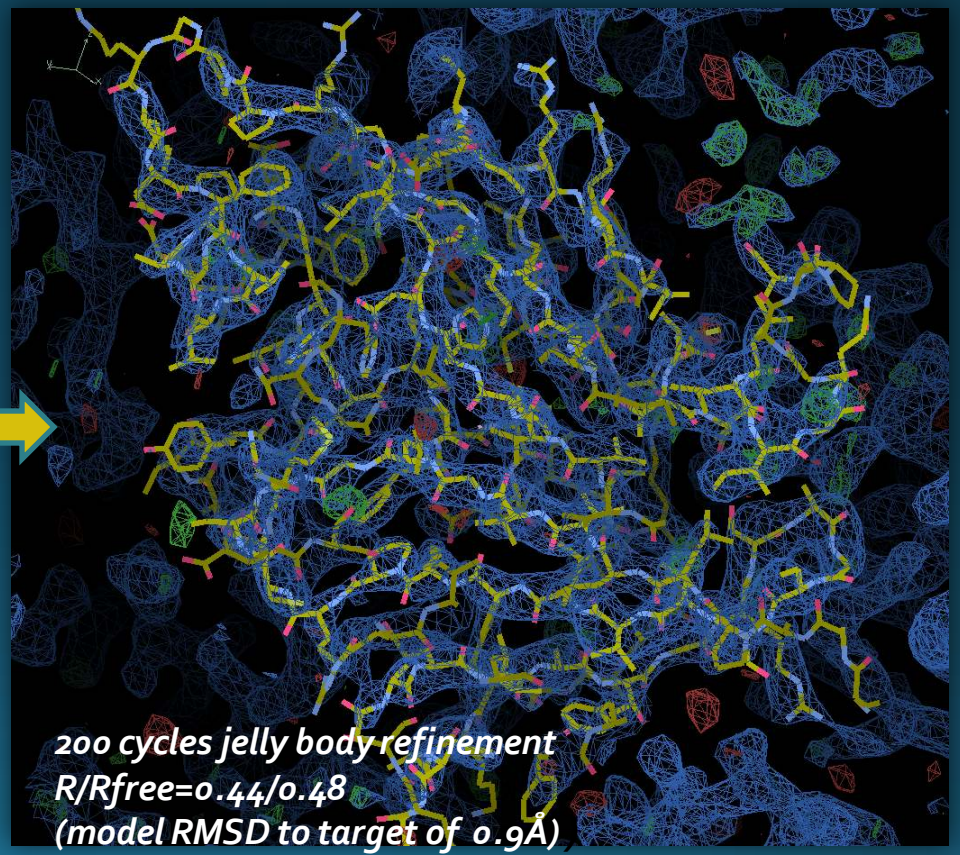
Assessing the solution

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

Example: 3td6 (resolution 3.2Å)



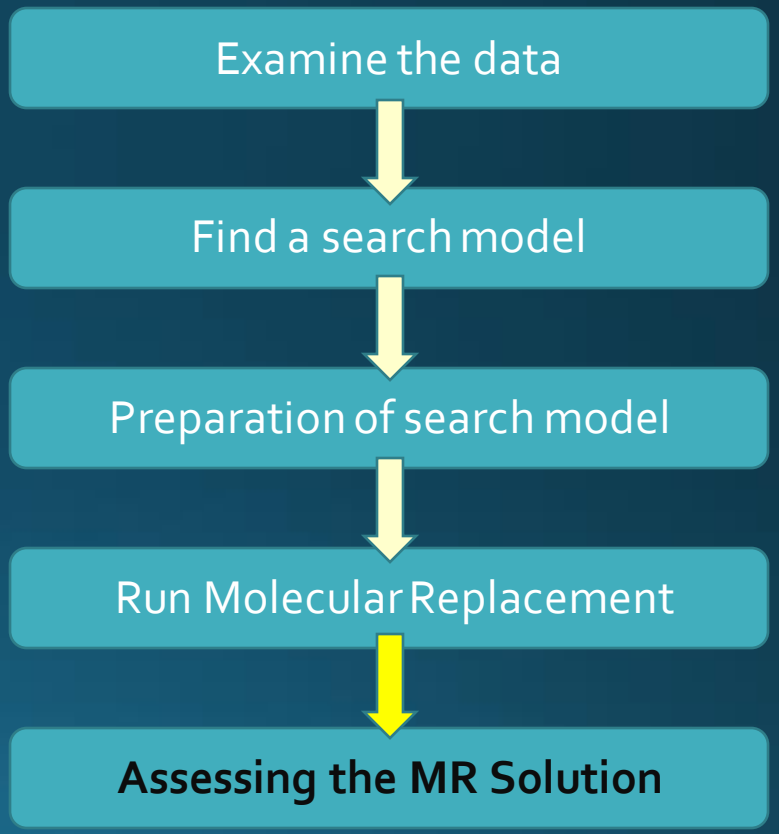
Phaser positioned model
R/Rfree=0.52/0.56
(model RMSD to target of 1.5Å)



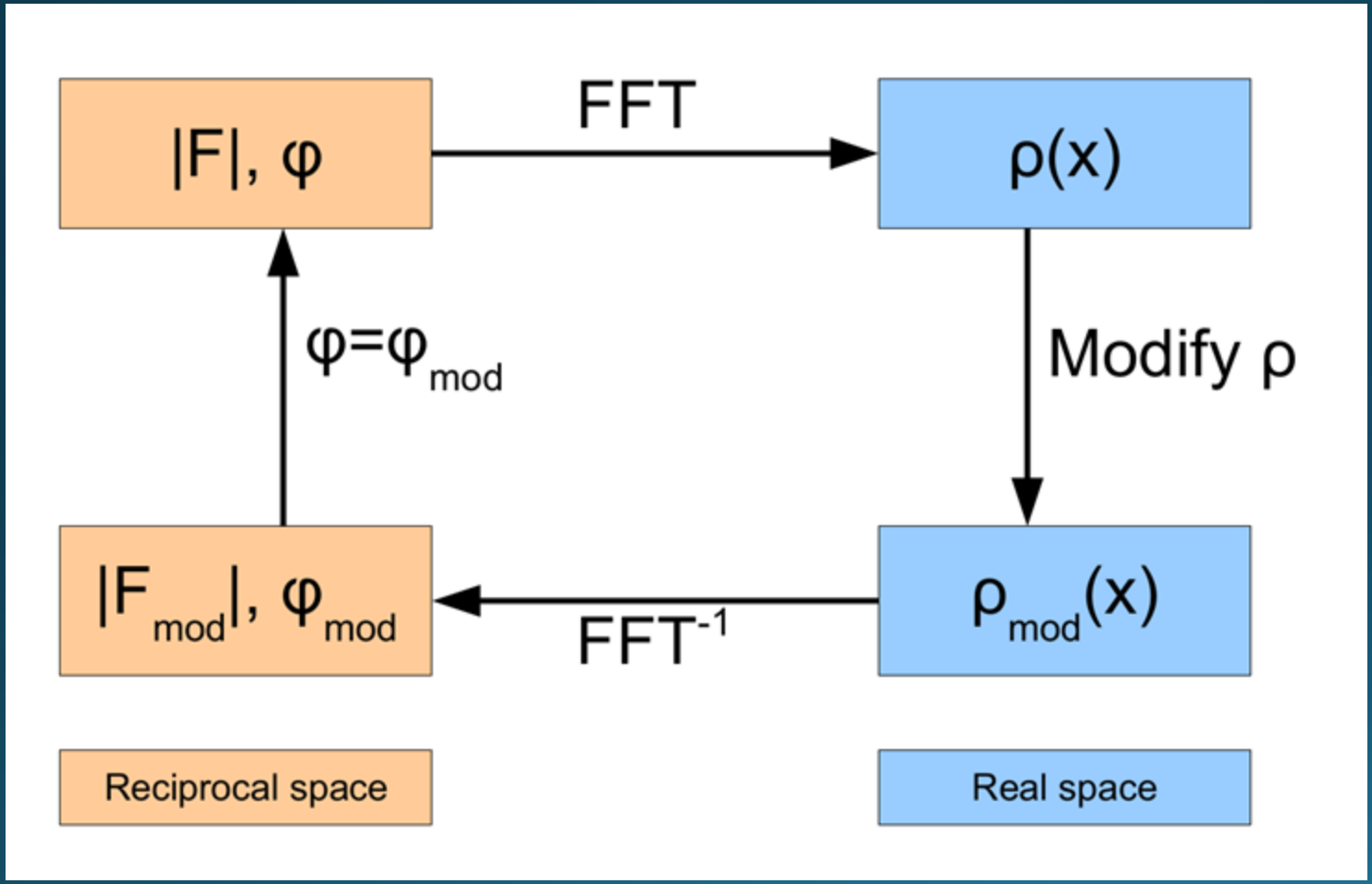
200 cycles jelly body refinement
R/Rfree=0.44/0.48
(model RMSD to target of 0.9Å)

Assessing the solution

- Density Modification & C-alpha tracing



Density Modification



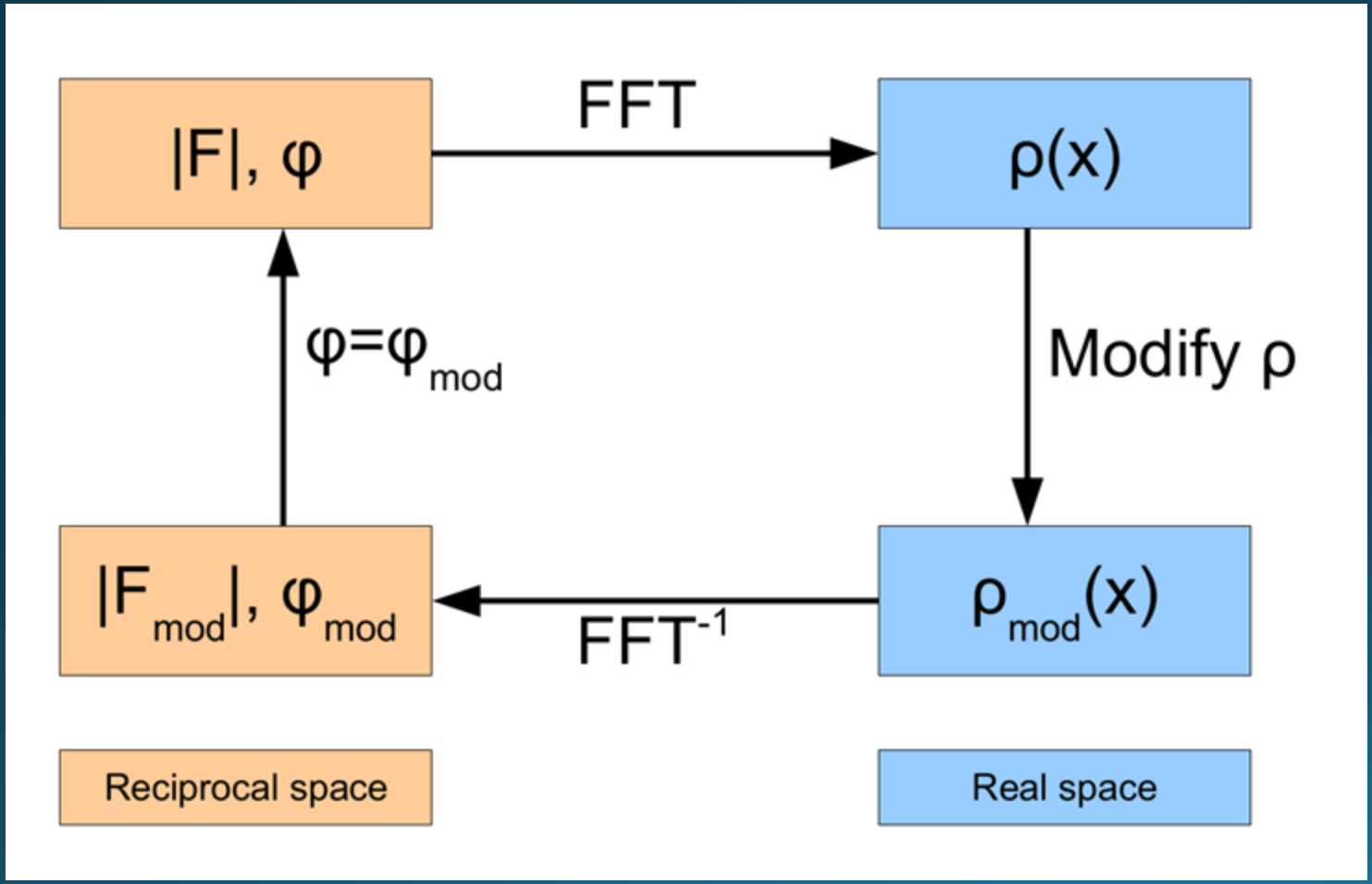
Main techniques:

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

(slide from Kevin Cowtan)

CCP4

Density Modification



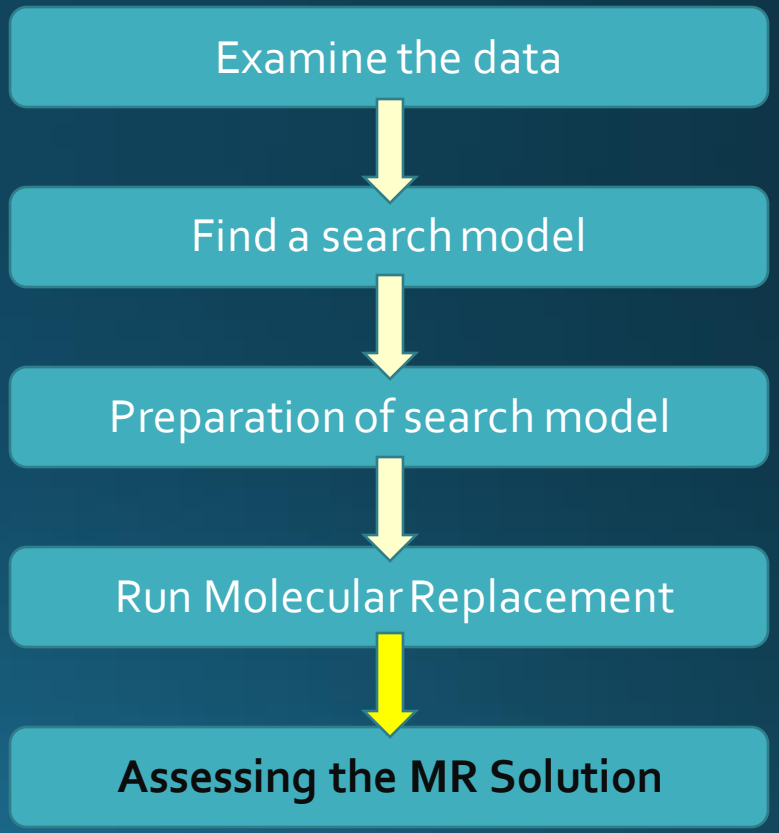
CCP4 Applications:

- 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



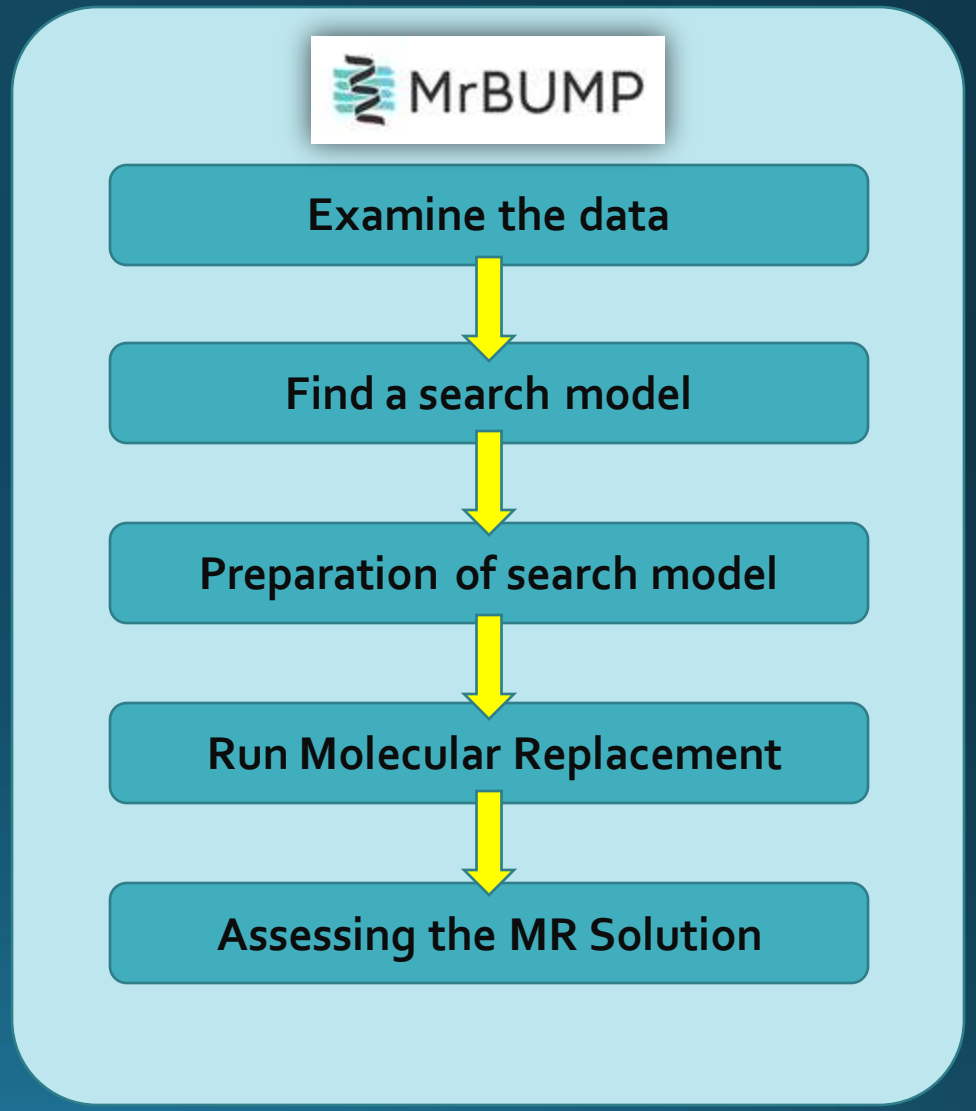
CCP4

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

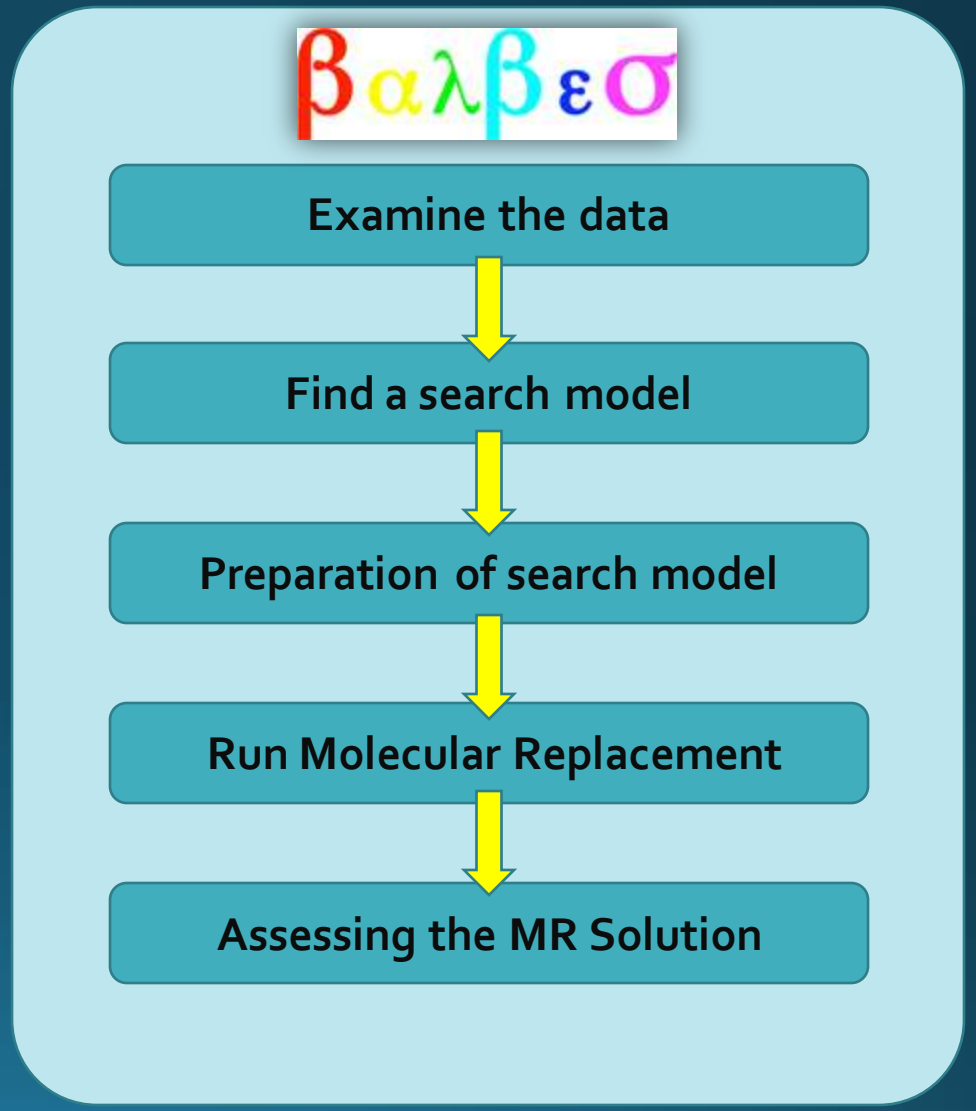
- Covers all stages from data examination through to initial model building
- Two modes:
  1. Model generation only
  2. Model generation and Molecular Replacement through to model building
- Can be run from:
  - CCP4i2, CCP4 Cloud & CCP4 Online
  - CCP4mg – model generation step with graphical view



CCP4

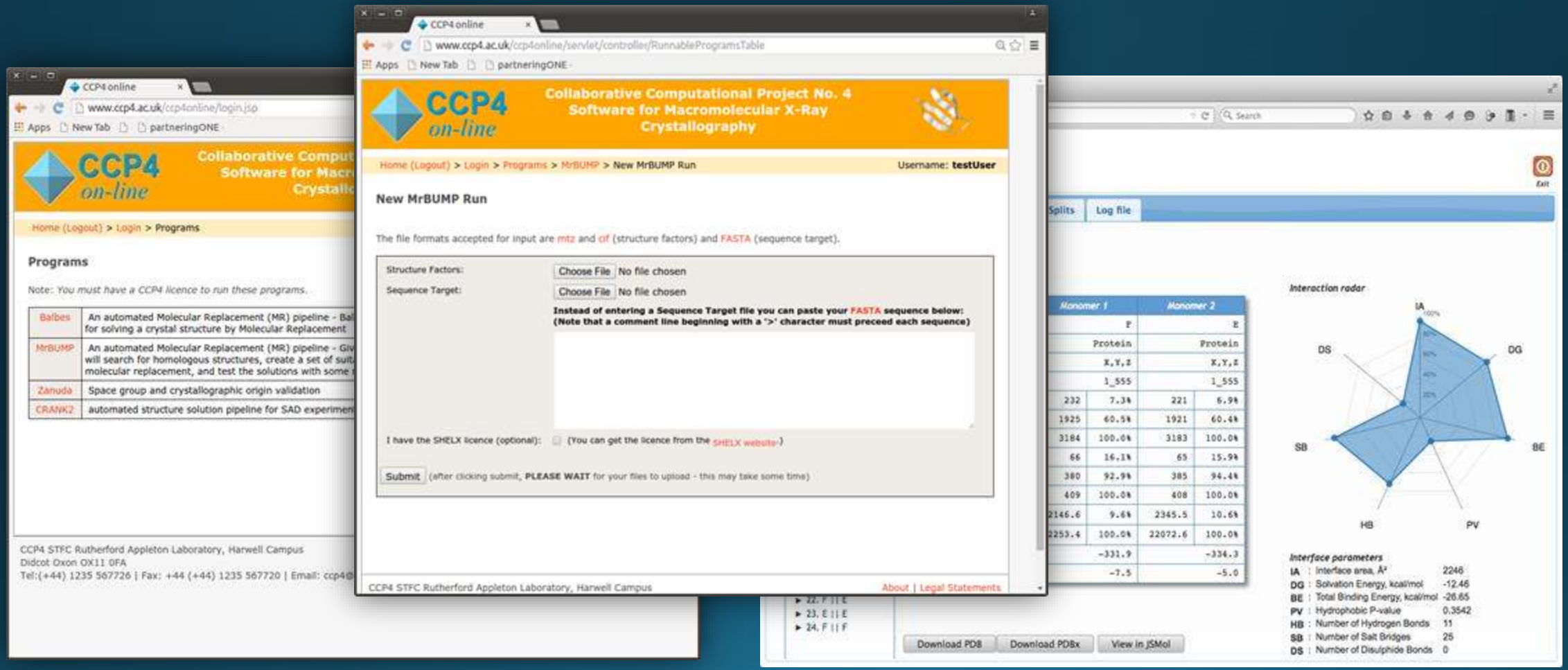
BALBES/MoRDa

- 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

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

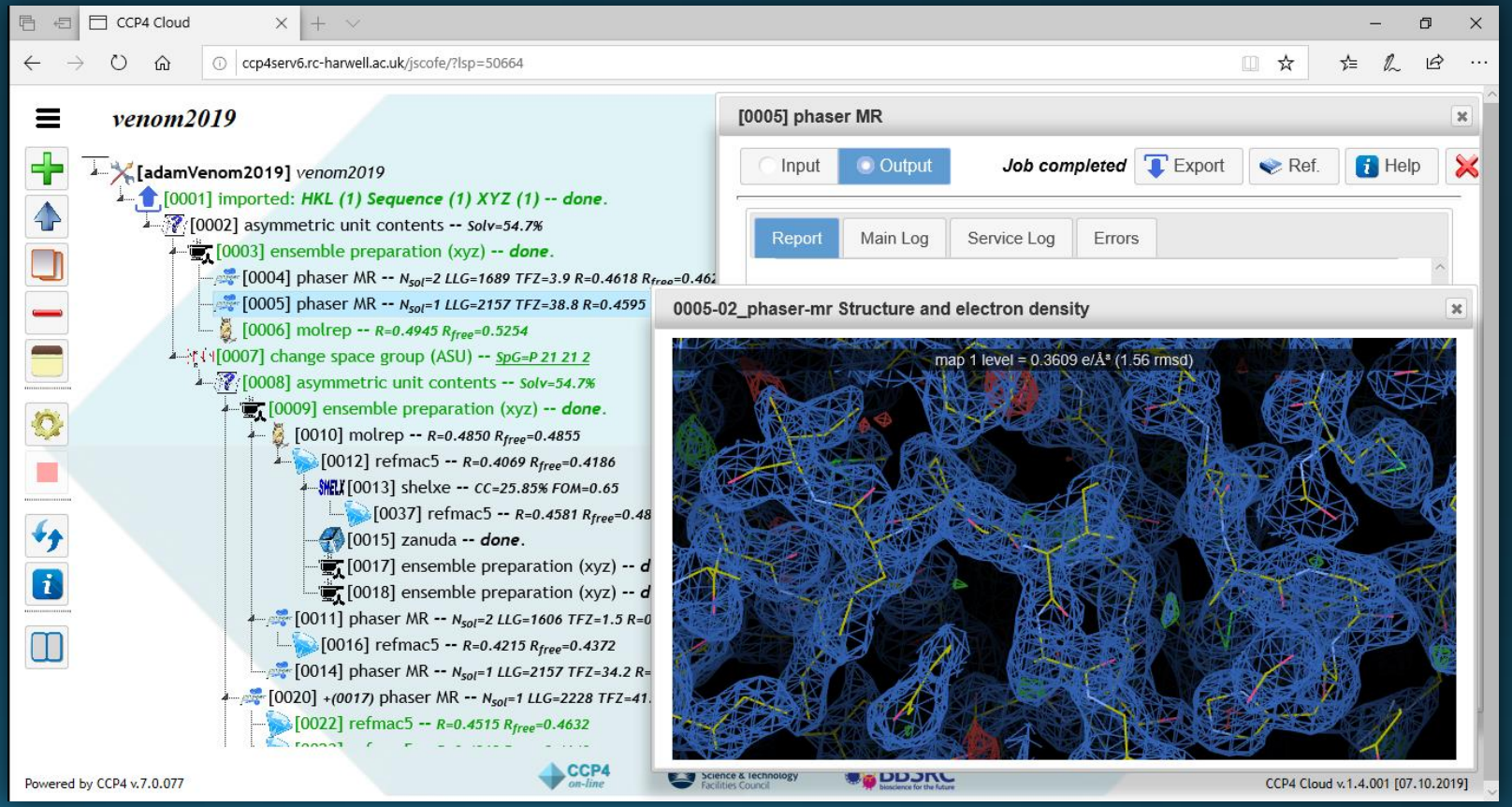


http://www.ccp4.ac.uk/ccp4online

CCP4

MR on the CCP4 Cloud

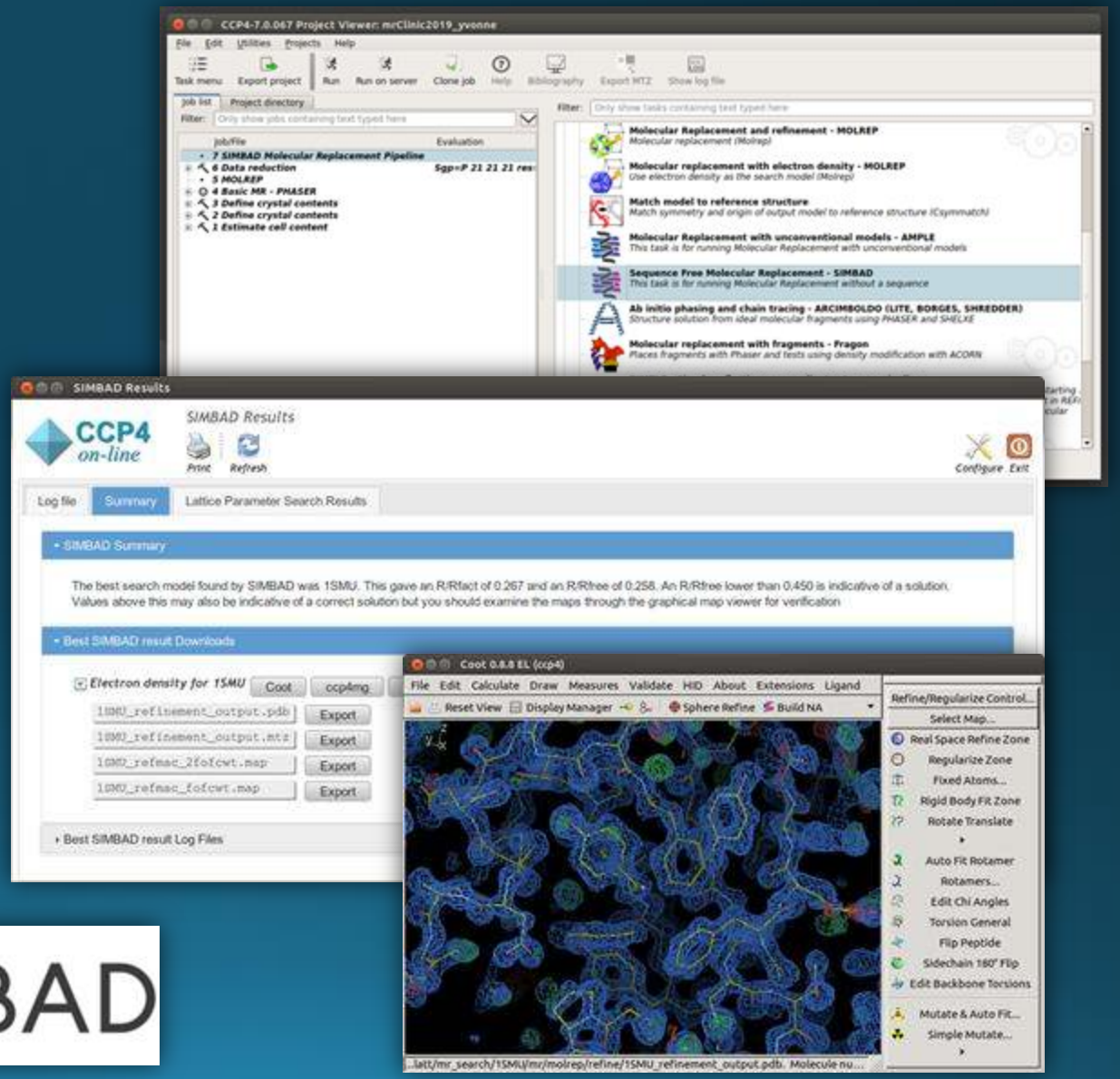
- Excellent interface for tackling difficult MR problems
- Explore broad range of search models, preparation protocols and refinements
- Exploit large computational resources



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



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***
- 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

CCP4