

1TEL Structure Solution Example

The von Willebrand factor A domain of human capillary morphogenesis gene II, flexibly fused to the 1TEL crystallization chaperone, Thr-Val linker variant, at 1.2 Angstrom resolution

9doc - 11/06/2024

(Samarawickrama, P., Probst, R., Ludlow, K., Doukov, T., Moody, J.D.)

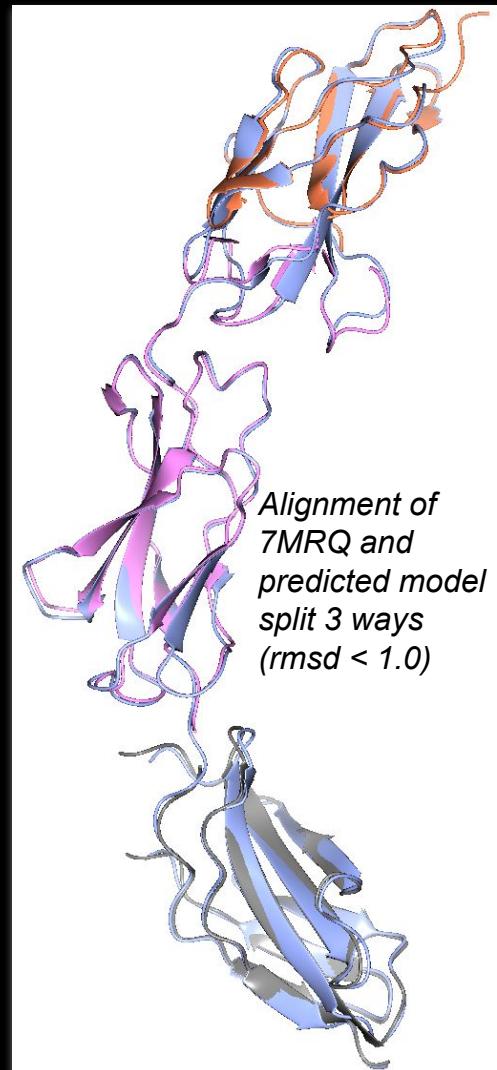
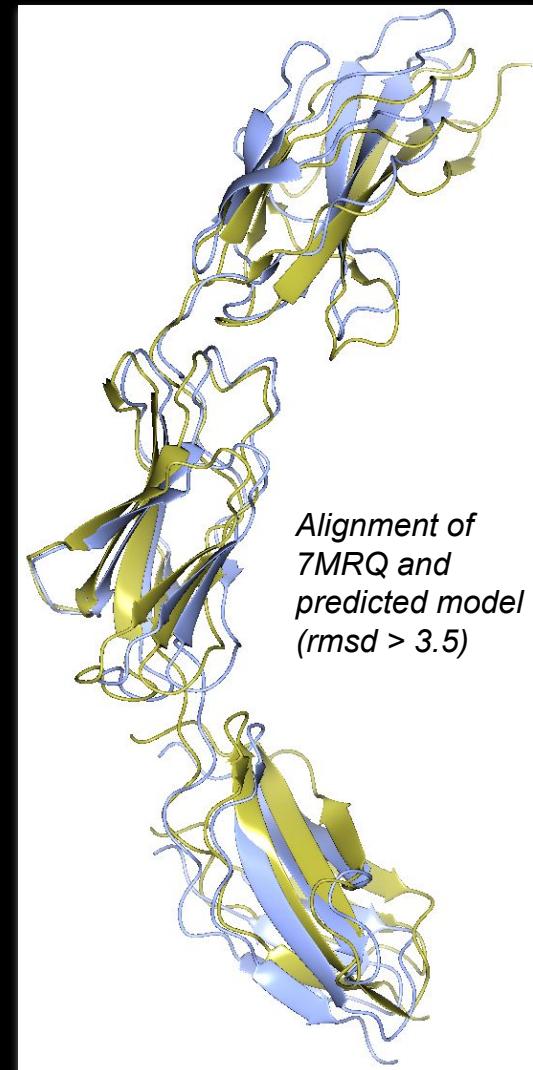
Search model splitting in Molecular Replacement

CCP4 Tutorial

Multi-domain structures

Multi-domain structures

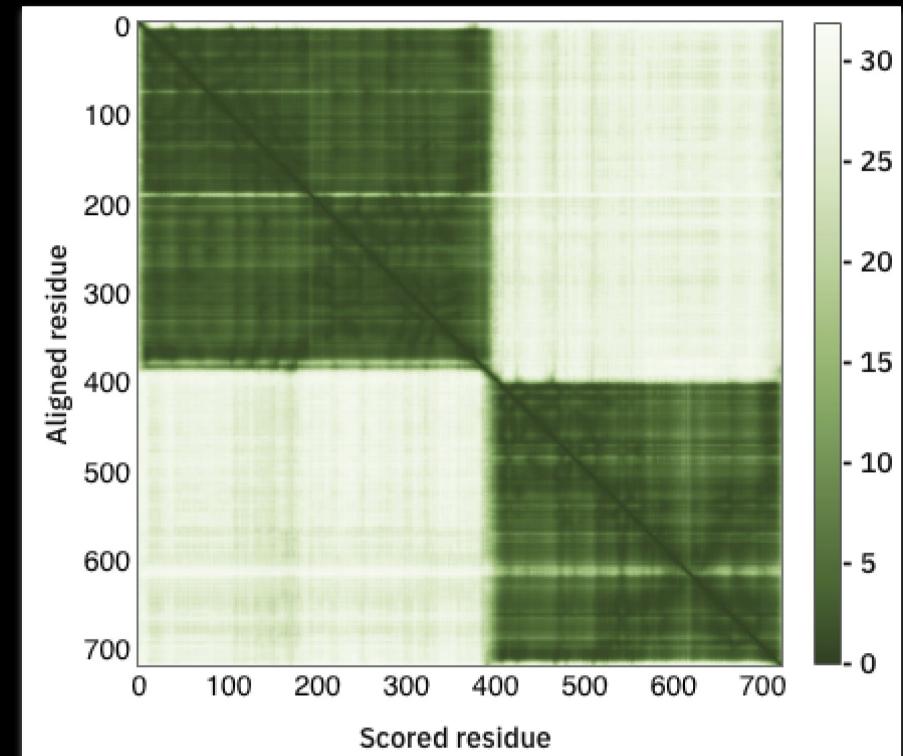
- The larger the target structure, the more likely there will be multiple domains
- A predicted model will often differ in the relative orientation of these domains when compared to the crystal structure
- Splitting a predicted search model into these domains can be a good strategy in MR



*7mrq - Chicken CNTN4 FN1-FN3
domains with T₇₅₁A, V₇₅₂A, Y₇₈₁A,
E₇₈₆A mutations*

Predicted Alignment Error (PAE)

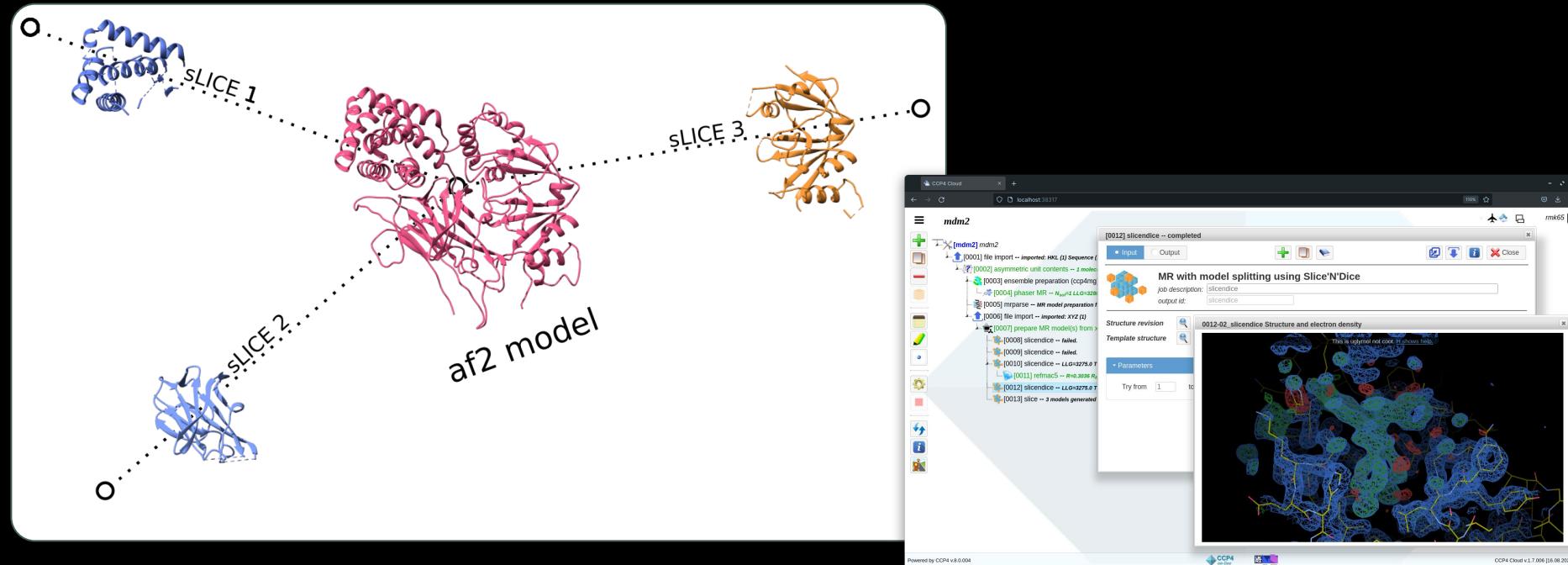
- PAE estimates the expected positional error for each residue in a predicted protein structure if it were aligned to a corresponding residue in the true protein structure.
- The PAE can be used to identify structural units or domains within the prediction
- Low error regions represent structural domains
- High error regions represent the inter-domain regions
- The relative orientation of structural domains in a prediction often differs from that of a crystal structure



The PAE plot for a prediction in the AlphaFold Database

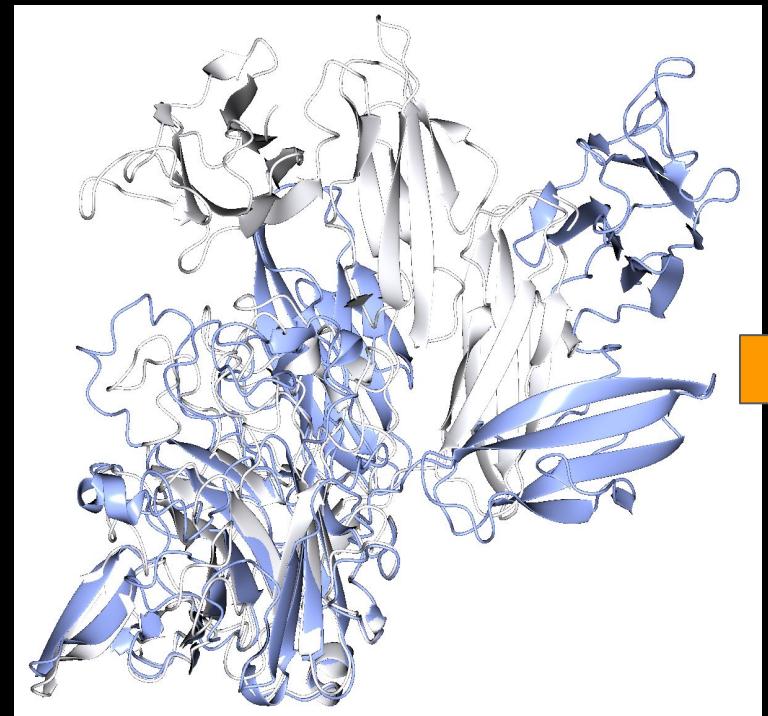
Multi-domain structures

- Splitting search models can be done with the “Split MR model with SliceNDice” task in CCP4Cloud
- There is also a “slicendice” task which uses the split models in MR
- Uses atom clustering to decide where to split the domains. The PAE can be used as a guide to how many splits are needed or it can be informed by a visual inspection of the predicted model



ESM-Fold Prediction

- Prediction (blue) aligned to target structure (white) on largest domain



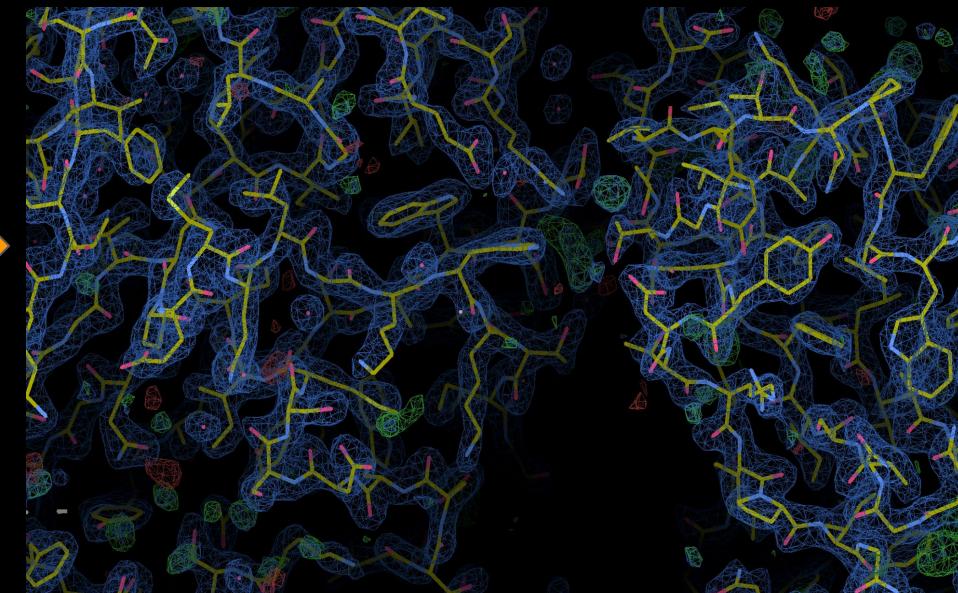
Processing with *Slice’N’Dice*

- 4-way Birch split of prediction
- Truncation to residues with a pLDDT > 70
- pLDDT converted to B-factor



Automated X-ray structure solution

- *Slice’N’Dice* places 7 of the 8 domain models (2 copies of target in asymmetric unit of crystal) using *Phaser* (LLG=755 TFZ=16.1)
- Automated model building with *Modelcraft* brings model close to completion (R/Rfree 0.26/0.3)



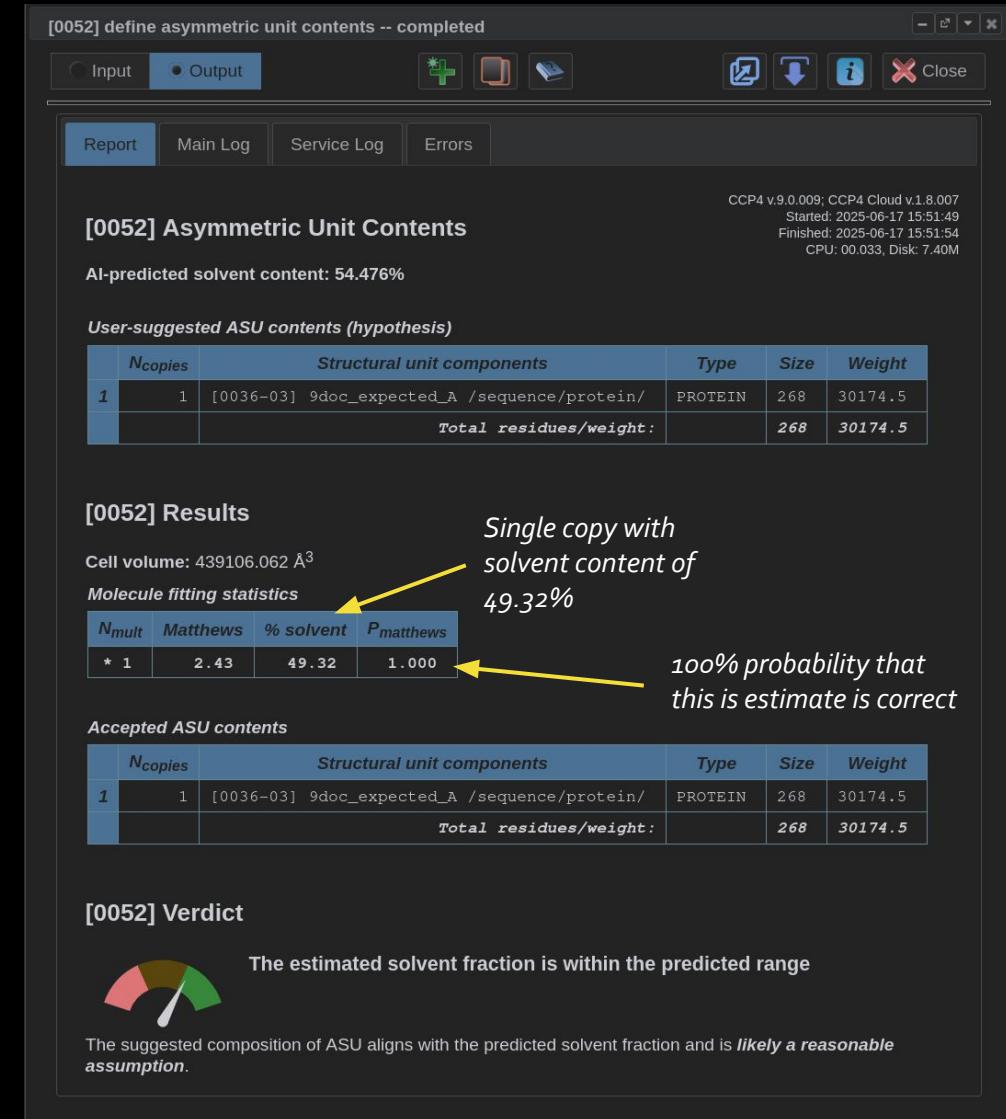
Tutorial

Data and Setup

- We will determine the structure of *the von Willebrand factor A domain of human capillary morphogenesis gene II, flexibly fused to the 1TEL crystallization chaperone, Thr-Val linker variant, at 1.2 Angstrom resolution* (PDB entry 9DOC)
- We will start by creating a CCP4Cloud project and importing the tutorial files. If not already done so, create a folder “MR tutorials”. Within this folder add a new project called “MR-example-2”.
- The data comes from a 900 image dataset collected using a Dectris Pilatus 6M detector on the BL-2 beamline at the SSRL Synchrotron at a wavelength of 0.97946 Angstroms. We will import a merged reflection mtz file already integrated, scaled and merged. The data has been processed in P61. Note that this spacegroup has an enantiomorph, P65. We can only determine the correct option during the molecular replacement step.
- The expected sequence of the structure contains 268 residues.
- Within the project add an “Import from Cloud” task option from the “Data import” menu. Select “Tutorials” -> “Data” -> “2_phasing” -> “mr-2-split_model”. Select all of the files using the Shift key. The import includes sequence, mtz (reflection data) and this pdf document.

Asymmetric unit contents

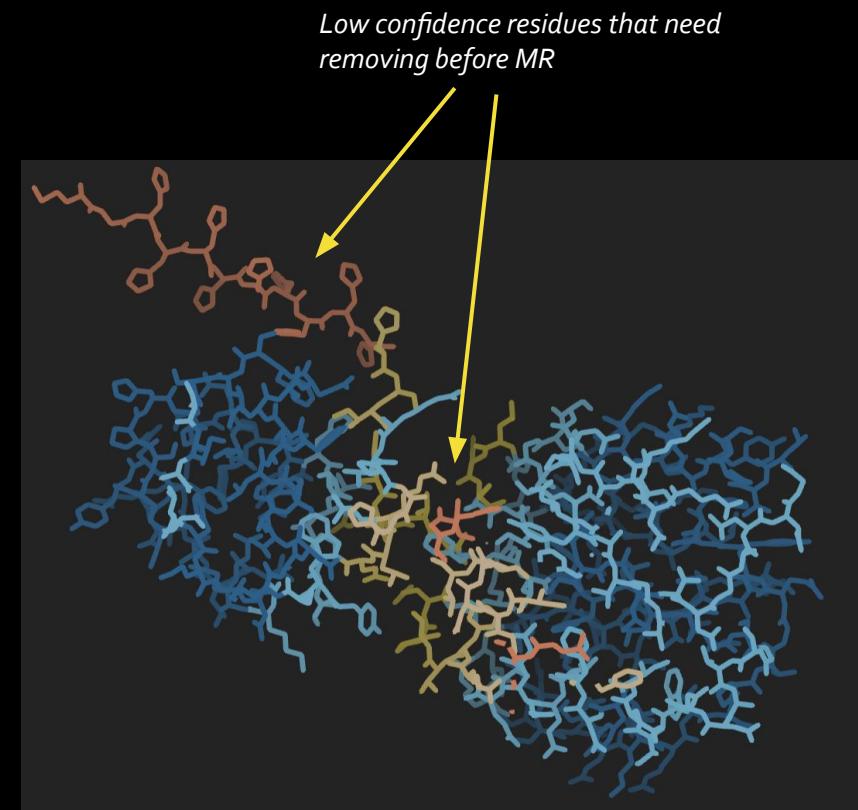
- The first step is to predict the asymmetric unit cell contents so we know how many copies of the model to search for. Add an “Asymmetric Unit Contents” task from the “Asymmetric Unit and Structure Revision” menu. By default it should load the reflection and sequence information for the target. Press the run button.
- This task estimates how many copies of the target molecule(s) can fit in the asymmetric unit based on an AI estimated solvent content and the number of residues in the target sequence. In this case it predicts a single copy with a solvent content of 49.32% with a probability of 100%.



Asymmetric unit contents report

Structure Prediction

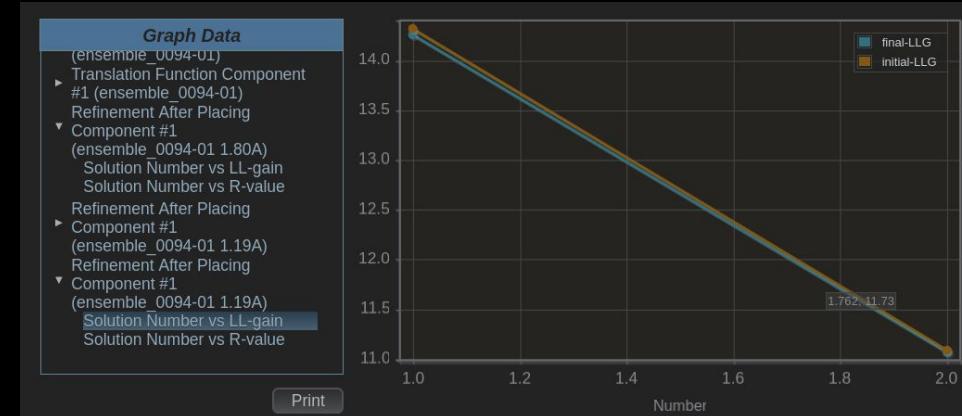
- We will use a predicted model as our search model in MR. To generate this, follow the asymmetric unit prediction task with a structure prediction task. This can be accessed from “Structure Prediction” menu.
- This task uses **AlphaFold2** (AF2) to predict the structure. By default it will load the sequence of the target structure. We want a single copy to be predicted. This task can also predict multimeric and complex forms for targets, but in this case we expect a single monomer in the asymmetric unit.
- When the prediction completes examine the resulting model using Uglymol. Adjust the colouring to show the residue pLDDT confidence scores. Here blue colours indicate high confidence that the positioning of the residues is correct, red/orange indicates low confidence. We will need to remove the low confidence residues before we perform MR.
- Run the “Split MR model with **Slice-n-Dice**” task from the “Molecular replacement” -> “MR model preparation” menu. By default it will convert the pLDDT scores into B-factors (“Convert B-factors” option). Leave the “Number of splits” as 1 to create a single search model. The task will also remove residues with a pLDDT < 70.



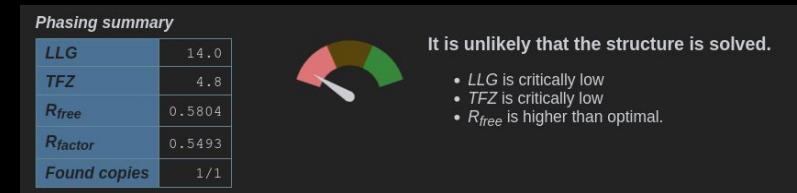
AF2 structure prediction viewed in Uglymol and coloured by pLDDT

Molecular replacement

- View the output model from **Slice-n-Dice** in Uglymol. Notice that the removal of low confidence residues has resulted in a break in the chain, leaving two globular domain parts. This may be significant and if we have difficulty solving the structure in MR, we may have to consider that the relative orientation of these two domains may differ in the crystal form.
- For now, follow the **Slice-n-Dice** task with a **Phaser** task from the “Molecular Replacement” -> “MR Solvers” menu. Leave all options as default. The data and the model from the previous task will have been set as input. **Phaser** will also search all possible spacegroups, P6₁ and P6₅ in this case. Run the task.
- As the task runs, look at the plots that are produced. **Phaser** performs three key steps, the rotation function, translation function and a rigid body refinement step for each copy (component) of the search model that needs to be placed. After each step, observe the LLG plot. We expect a value of 60 or more after the refinement step to be indicative of a solution.
- The final LLG will be much lower than this, most likely indicating that the model could not be placed correctly. We will need to revisit the model preparation step and consider why it doesn’t work in MR.



Refinement LLG plot after placement of search model

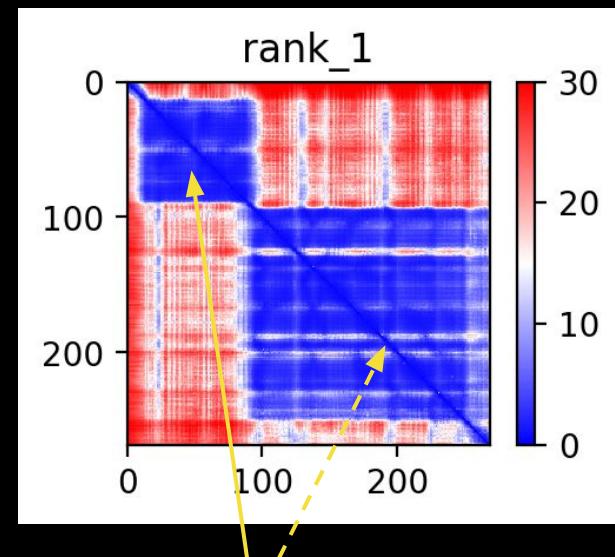


Summary of Phaser solution in CCP4Cloud report

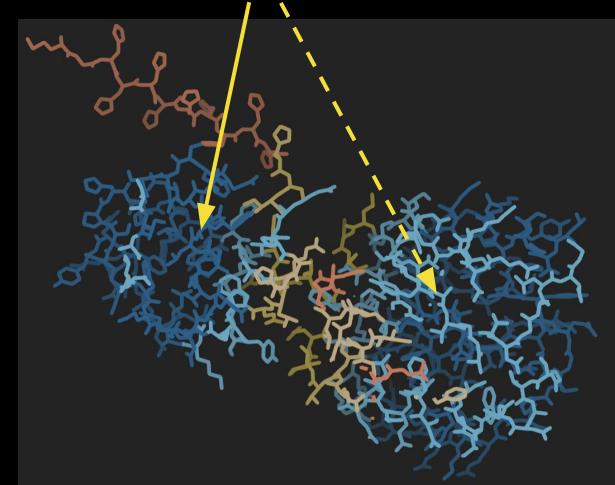
Structure Prediction - domains

- Return to the structure prediction task and look in more details at the report. Several plots are available as well as the option to view the prediction using Uglymol. Examining the Predicted Alignment Error (PAE) plot (top right), we can see that there is evidence for two distinct domains (blue regions) which we can also see in the Uglymol view of the model (bottom right). In this viewer, press “c” to adjust the colouring of the model to pLDDT colouring. We can see that the two domains have high confidence residues (blue) while the linking residues are low confidence (orange/red).
- In cases where there is more than one domain present, it is unlikely that the crystal structure will have these domains in the same orientation relative to each other. It is therefore a good idea to split these up into separate search model for use in MR.

The PAE plot for the AF2 prediction



The predicted alignment error (PAE) shows two distinct structural domains



AF2 prediction viewed in
Uglymol and coloured by pLDDT

Splitting the predicted model

- We will need to split the domains apart into separate search models. As before, we will want to remove the low confidence ($p\text{LDDT} < 70$) residues as these are not likely to be in the same position in the crystal structure. The other preparation step required is to convert the $p\text{LDDT}$ scores for the remaining residues into B-factor estimates. This is important for **Phaser's** scoring system when it comes to performing the MR step.

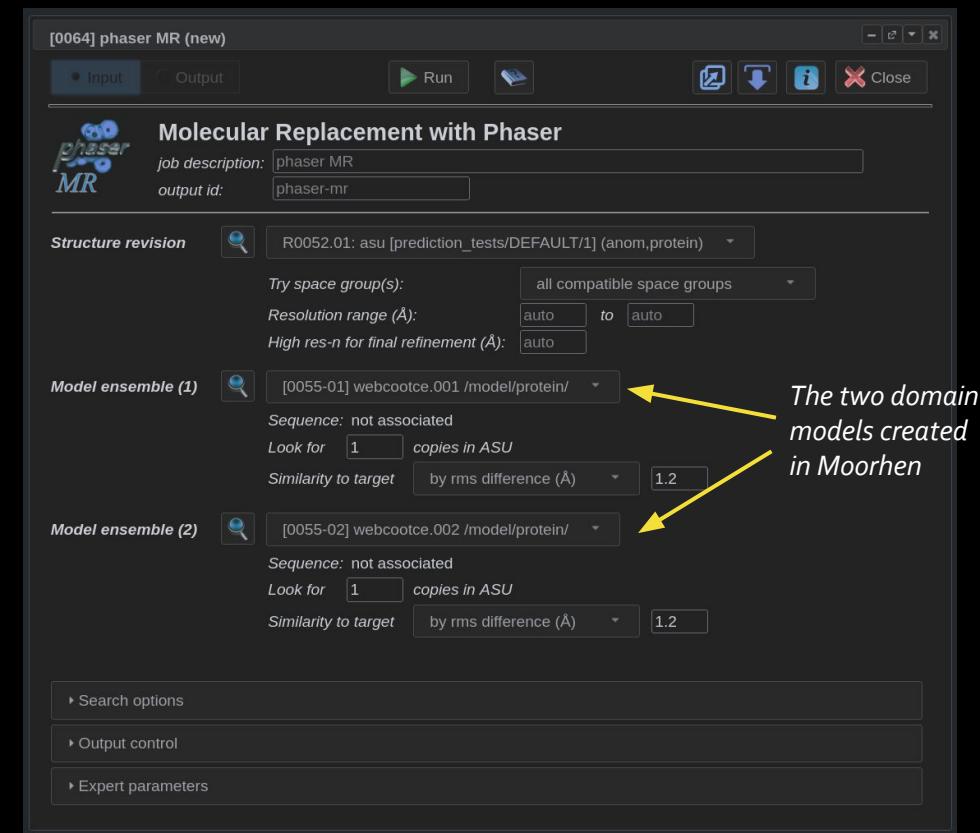


WebCoot/Moorhen Slice-n-dice interface

- To prepare the model we will use the graphical interface "Moorhen". You can find this under the "Coot" menu. Starting a new branch in the CCP4Cloud project, follow the structure prediction task with the "Edit Coordinates with WebCoot/Moorhen" task. This will load the predicted model into the graphical viewer.
- From Moorhen's menus choose "Calculate" -> "Slice-n-dice". The model view will change to ribbons and will be coloured by $p\text{LDDT}$ score. A dialogue box will open which we can use to control the $p\text{LDDT}$ threshold for removing residues and the number of domains to be "sliced" from the initial model. By default, the $p\text{LDDT}$ threshold is set to 70 and the number of slices to 2. This is what we want for this model. You can adjust the slider for $p\text{LDDT}$ to see what residues are being removed/retained.
- With the $p\text{LDDT}$ threshold set to 70 and the number of slices set to 2, press the "Slice" button. You should now see that the model has been converted into two separate domain models (coloured). We will use these two models in MR. Press the "Save & Close" button and then exit Moorhen. The models will be saved to CCP4Cloud.

Molecular Replacement

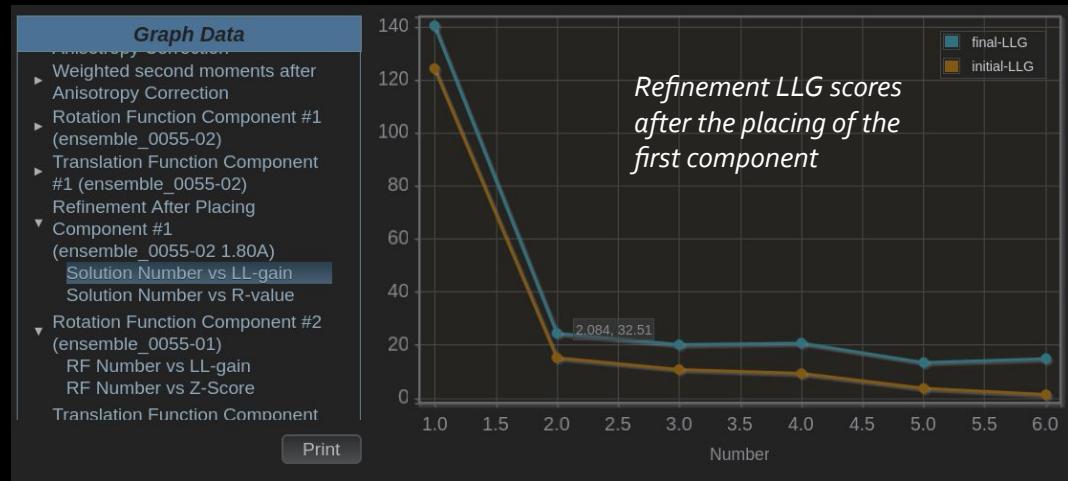
- Add a **Phaser** task to the previous task. This can be found under “Molecular Replacement” -> “Mr Solvers” -> “Molecular replacement with Phaser”.
- By default, the interface will load the structure revision which includes the reflection data intensity (or amplitude) information as well as the estimate for the asymmetric unit contents. Note that it will search all possible spacegroups. In this case that will be P6₁ and P6₅.
- One of the models we have created in Moorhen is loaded by default (Model ensemble (1)). We need to provide the second domain model using the Model ensemble (2) field. By default, the models are set to have an estimated rmsd (root mean square deviation) to the target of 1.2 Angstroms. This is required by **Phaser** to account for possible errors in the model when compared to the true structure. We are searching for 1 copy of each domain as there is 1 copy of the overall molecule present in the asymmetric unit. Run the job.



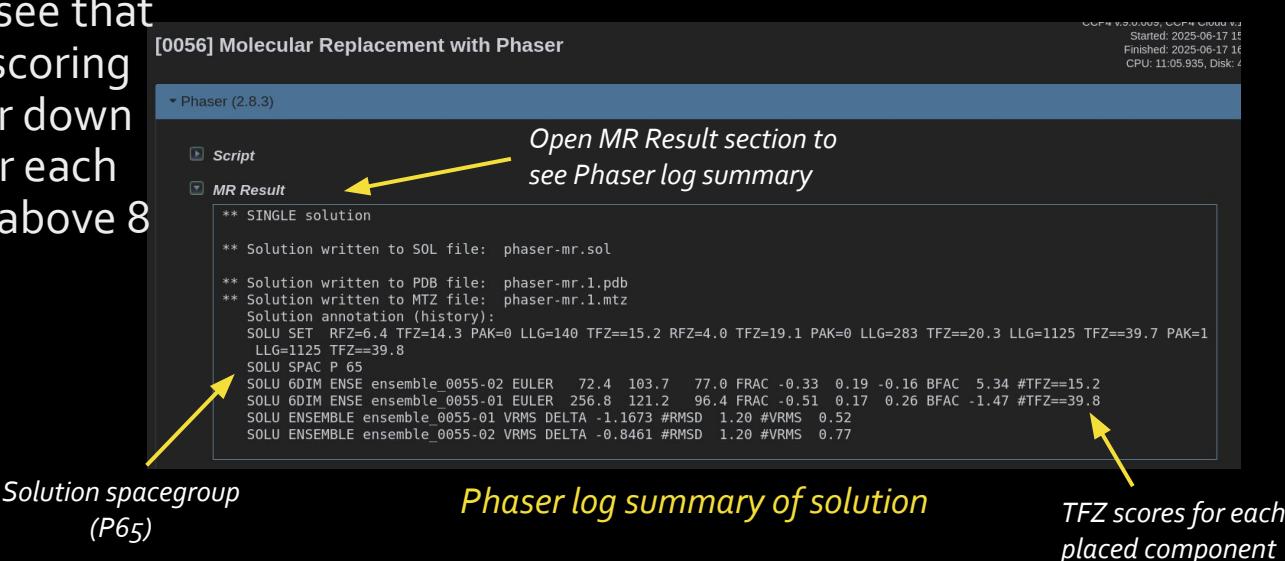
The CCP4Cloud Phaser interface

Molecular Replacement

- Follow the steps that **Phaser** is doing through the plots provided in the report page. For each of the two domain models (components), it will perform a rotation search, translation search and a rigid body refinement step. For each component we expect an increase in the LLG score for a correct solution. In this case, the final LLG should be above 1000 if everything has been placed correctly.
- When the task completes, open the “MR Result” section to see the final summary from Phaser. You should see that there was a single solution, meaning that the top scoring placement for the models is unambiguous. Further down you will see Translation Function Z-scores (TFZ) for each of the placed components. We expect these to be above 8 for a correct placement.
- Note also that the solution spacegroup is in P65.

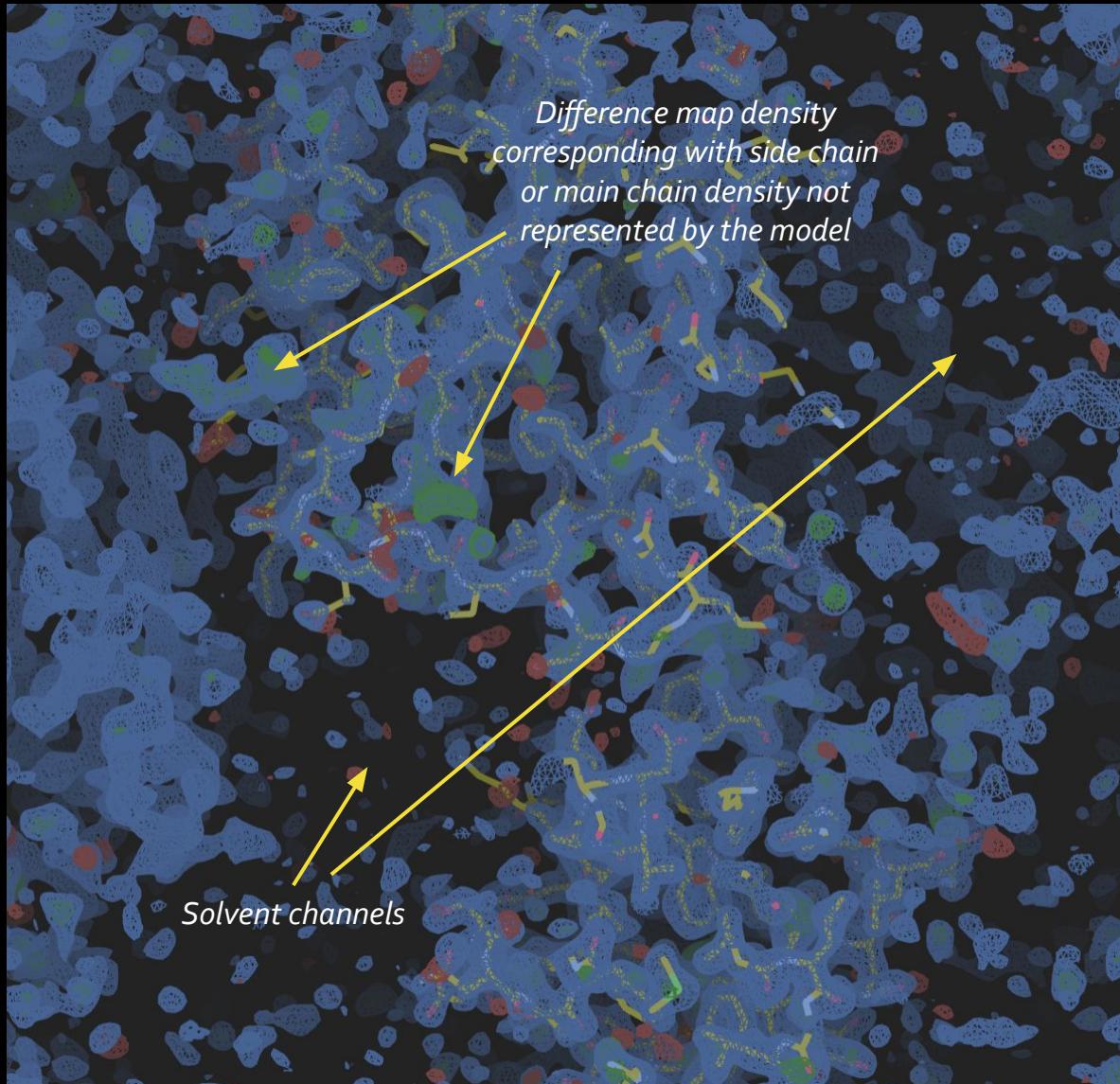


Phaser plots in CCP4Cloud report



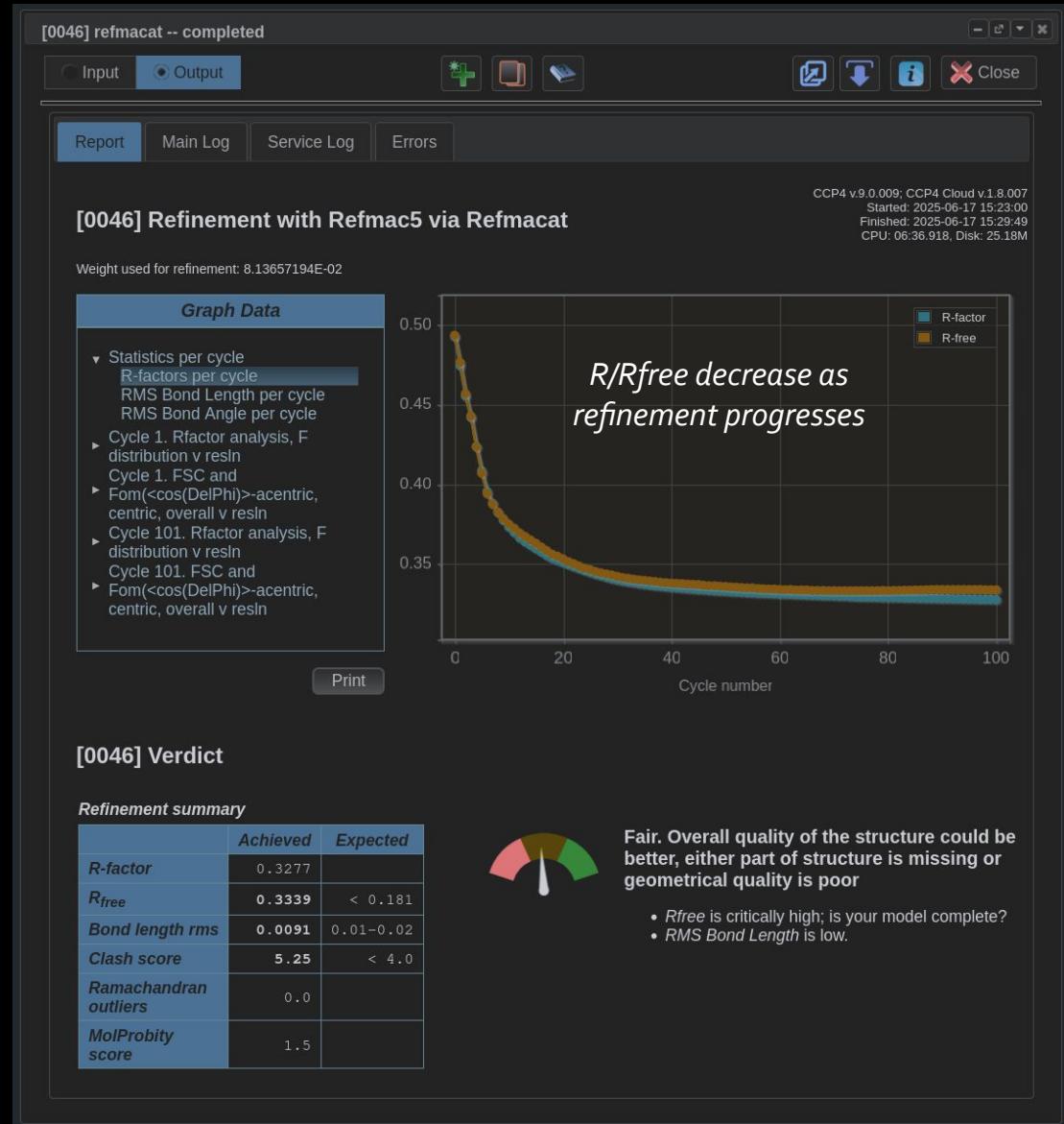
Initial Refinement

- View the resulting model placement and electron density from **Phaser** in Uglymol. There is some evidence in the difference map indicating that there are missing parts of the model. There is also evidence of solvent channels but there remains some noise (electron density peaks) in these areas. This is due to the phase error resulting from the fact that the search models, although correctly placed, differ in their atom positioning when compared to the true structure.
- To improve the model, and subsequently reduce the phase error, we need to refine it. We will do this using 100 cycles of jelly-body refinement in **Refmac**.
- Add a **Refmac** task from the “Refinement” menu and set the number of cycles to 100. Under the “Restraints” tab set “Use jelly-body restraints” to Yes. Run the task.



Initial Refinement

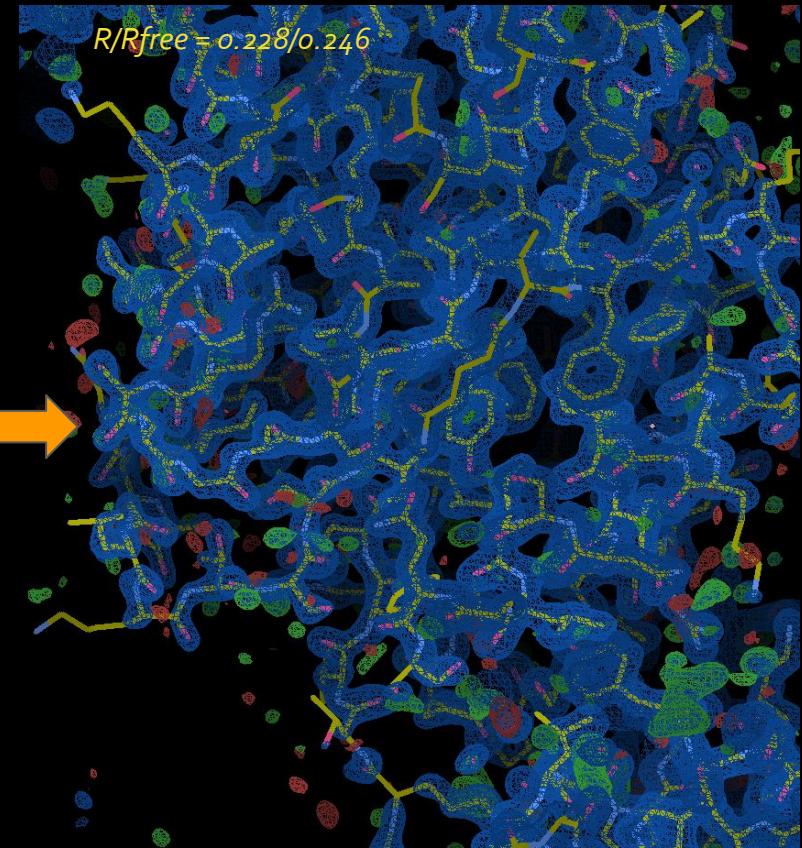
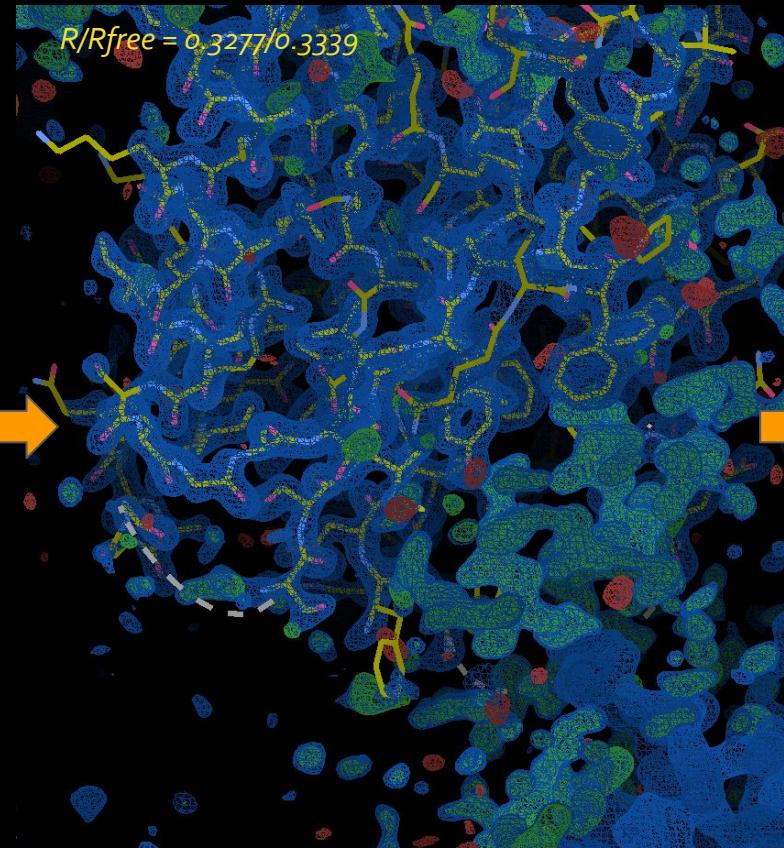
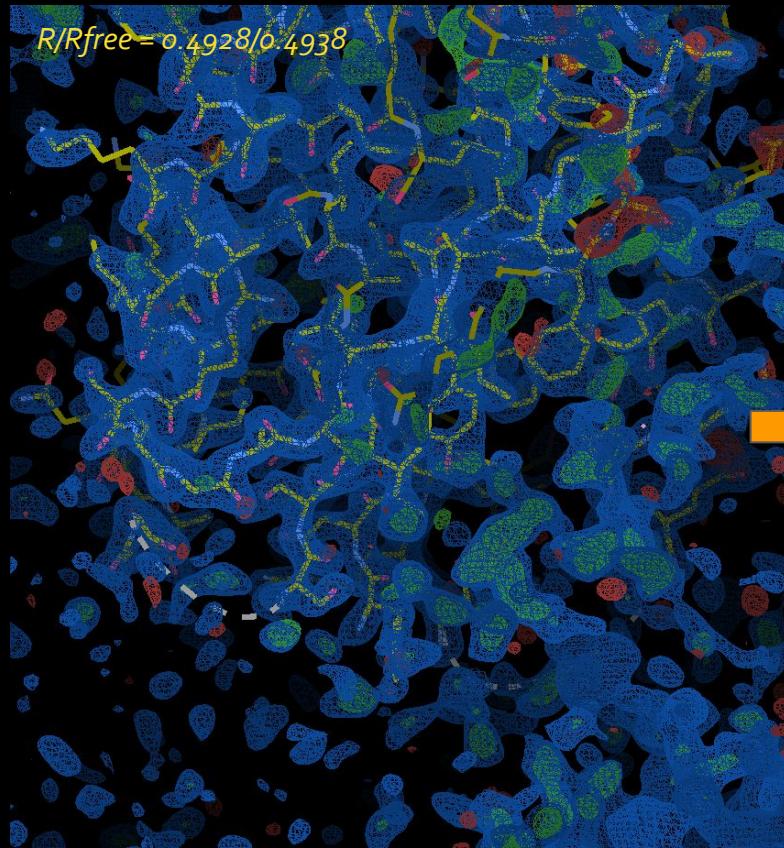
- As refinement proceeds, observe as the R/Rfree values drop. R/Rfree measure the agreement between the model and the experimental data with lower values showing greater agreement.
- View the resulting model and map in Uglymol, notice how the solvent channels are “flatter” with fewer peaks and the green difference map density will show more clearly parts of the data where the model needs further building.
- To complete these missing parts of the model we can use both automated and manual model building.
- Follow the refinement task with a run of “Modelcraft”. This can be found under the “Automatic Model Building” menu. Leave all options as default, the structure revision output from refinement will be used. Set the task running.



The Refmac refinement report page in CCP4Cloud

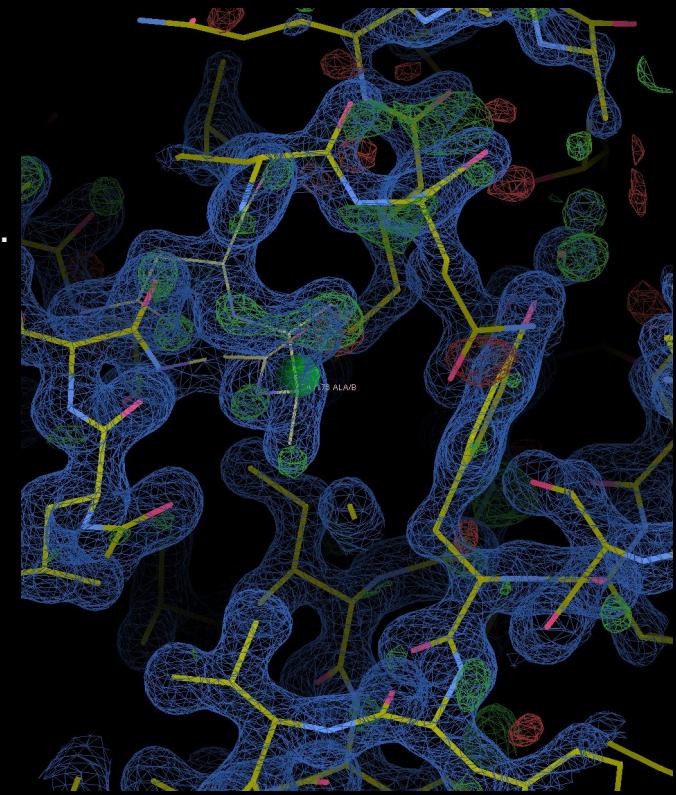
Automated model building

- Automated model building should reduce the R/Rfree further and bring the model close to completion. Look at the output map and model again in Uglymol. Does it appear more complete than it was after refinement?



Manual model building

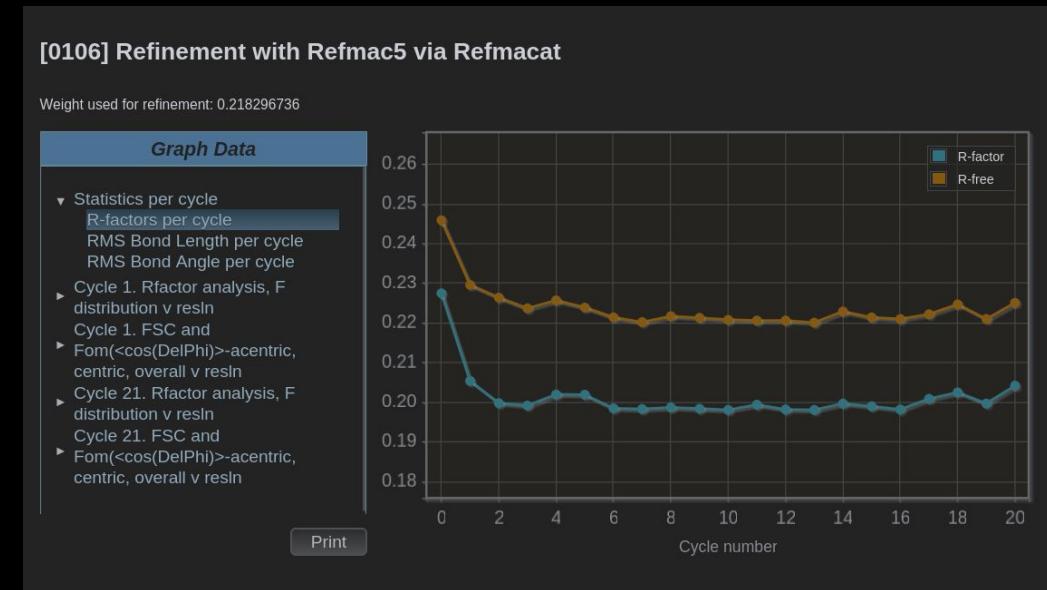
- At this point not much more can be done automatically to complete the structure. Completion and validation of the structure need to be done manually in a graphical interface such as **Coot**. Full completion of the structure can be done outside the scope of this tutorial but we will go through a few useful functions to improve the model from this point.
- Add a Coot task to Modelcraft task. **Coot** will by default take the structure revision from Modelcraft which includes the current model and the density maps. Run the task to start the Coot interface. Click through the
- You may be asked to fix some nomenclature errors in the model, click Yes to proceed. You may also be informed in a dialogue box that it is necessary to save the coordinates from Coot in order for them to be saved into CCP4Cloud. Click "OK" if you see this box.
- **Coot** should also open up a "Molprobity to-do list". This will highlight part of the model that may need your attention, such as rotamer outliers, C-beta outliers and parts of the model that are causing clashes. These should be examined and fixed if possible.
- A useful first place to start is to run the "stepped refine" option. This option can be accessed from "Calculate" -> "All Molecule" -> "Stepped refine". This will automatically step through each residue and perform real space refinement. This is a quick way to fix some of the easier to fix side chain problems and gives a good view of the whole structure.



Stepped refinement in Coot

Manual model building

- When stepped refinement has completed save the coordinate file. To do this go to “File” -> “Save Coordinates”. In the dialogue that opens don’t adjust the file name or the location where it will be saved. Simply press “Save”. This will enable CCP4Cloud to import the model into your project for further processing.
- Exit Coot and return to your CCP4Cloud project.



R/Rfree in Refmac with anisotropic B-factor refinement enabled

- The next step is to refine the output model from Coot. This should allow us to generate an improved electron density map which we will again use in Coot to aid us in further manual model building. Add a refinement with Refmac task to the Coot task. Set the number of cycles to 20 and ensure that the jelly-body option is turned off. At this point we will start using restrained refinement, which uses both the experimental data and chemical knowledge as restraints for the refinement of the molecule.
- Given that the resolution of the data is better than 1.5 Angstroms we can invoke anisotropic refinement of B-factors to account for the fact that atoms are more restrained in their movement in certain directions e.g. where an atom is bonded to a neighbouring atom. This should improve the fit of the model to the experimental data and reduce the R/Rfree values further. This option can be found under the “Model parameterisation” tab. Set the Atomic B-factors to “Anisotropic”. Run the task.

Manual model building

- From this stage it should be straightforward to complete the model and make it ready for deposition. The validation tools in Coot can be used to improve the model in various places. The addition of further water molecules and solvent will all help to lower the R/Rfree reflecting the fact that the model is improving in its agreement with the experimental data. After each adjustment to the model, a round of restrained refinement can be performed in Refmac which subsequently produces an improved map.