Creating Molecular Replacement Search Ensembles with CCP4MG/MrBUMP in CCP4i2

Introduction

This tutorial shows how to interactively use the model preparation steps of the MrBUMP molecular replacement pipeline using CCP4MG. MrBUMP requires only a sequence to prepare a search ensemble and therefore a sequence is the only input required by CCP4i2 when using the CCP4MG/MrBUMP task.

The task works in the following way:

1. CCP4MG is opened with the specified sequence as input. CCP4MG displays this sequence.
2. CCP4MG immediately starts running the model preparation steps of MrBUMP, which does the following:
   The program *phmmer* is used to find a set of structures with similar sequences to the input, but which have no more than specific value of sequence similarity with each other.
   The structures are then pruned using *Scupltor* and aligned using *Gesamt*.
3. The pruned and aligned structures are then loaded into CCP4MG as ribbon models.
4. The user may then change which atoms are displayed in CCP4MG. The displayed atoms can then be saved in CCP4i2 as a molecular replacent ensemble. Any number of different ensembles may be saved.

Example 1 - *gamma* test data

1. Create a new project named e.g. *"mrbump_gamma"*
2. Copy the CCP4 *gamma* test data to this project with the *Utilities -> Copy demo data to project -> Gamma* menu item.
   This will place the test data in *$HOME/CCP4I2_PROJECTS /mrbump_gamma/gamma* (Mac and Linux) or *C:\Users\<your username>\CCP4I2_PROJECTS\mrbump_gamma\gamma* (Windows).
3. Click the task menu button in CCP4i2.
   Expand the ***Bioinformatics including model preparation for Molecular Replacement*** section.
   Create a new ***Interactive model preparation - CCP4MG and MrBUMP*** task.
4. As stated above, the only required input is a sequence. However, this task - like many others - uses the CCP4i2 concept of *Crystal contents* as its input. (In fact, CCP4i2 in the end just passes a single sequence to CCP4MG, but this detail is not really important).

   To input a sequence:
   - Click on the ***Specify crystal contents*** button.
     Click on ***Browse for sequence file button***, load ***gamma.pir***
     Click the ***Save*** button.
   - (Alternatively the drop-down menu may be used to select a previously

defined crystal contents. This is relevant in real-life but not for the purposes of this tutorial.)
    - There are two options in the CCP4MG/MrBUMP task: ***Non-redundancy level for homologue search***, this is the value which determines how similar in sequence the search models can be; and ***Cutoff threshold for phmmer results***, this number is a score threshold specific to the homelogue search - it may be reduced in cases where there a very few homologues. We should leave both these at their default values of 95 and 20 for this example.
5. Click CCP4i2's *Run* button. CCP4MG will open with the *gamma* sequence displayed. Another window opens which shows, for the moment, the MrBUMP log file. After a while this will change to a view showing how the sequences of the search models match the input sequence. They are ordered by *phmmer* score with the best at the top.
6. Move the MrBUMP results window so that you can see it and the main CCP4MG window clearly. Move the ***Show residues with gesamt variance*** slider to the left and see what happens to the displayed atoms. The more the slider is moved to the left, the fewer the displayed atoms. Lowering the slider value requests CCP4MG to display atoms which are more conserved in 3D position between all the structures. Moving the slider to the left will remove features such as loops with high positional variability. The far right value of 110 means show everything.
7. Set the slider to approximately 3/4 the way to the right (about 75). Click on the *File* and select ***Save all visible to CC4i2 database***. Repeat this with the slider about 1/2 way (50).
8. Quit CCP4MG. The CCP4i2 report for the task will show the search hits found by MrBUMP and pictures of any ensembles saved by CCP4MG. These ensembles will also be listed in the *Output data* section of the report.
9. This output file of this task can now be used as a molecular replacement search model.
10. Run a **"Data reduction - AIMLESS"** using ***gamma_native.mtz***
11. Create a **"Basic MR - PHASER"** task. Use the output of the AIMLESS job above as the input Reflections; use the output of the CCP4MG/MrBUMP model as the Search Model. Set the *Composition of asymmetric unit* to ***"Provided as full specification by sequence"*** and then set *Crystal contents* to the same as in the CCP4MG/MrBUMP model.
12. Run the Phaser task. At the end of this task you should have a model suitable as input for auto-building/refinement.

Example 2 - *5yca*

1. Run an ***Import sequence*** job with the following in the ***Or enter the raw text of the sequence..*** box.

```
>5yca 5yca
GPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIEADQTPEDLDMEDNDIIEAHRSLPAE
RNPLYKDDTLDHTPLIPKCRAQVIEFPDGPATFVRLKCTNPESKVPHFLMRMAKDSSISATSMFRSAFPKATQEEEDLEM
RWIRDNLNPIEDKRVAGLWVPPADALALAKDYSMTPFINALLEASS
```

2. Create a new CCP4MG/MrBump task.
   Set the *Crystal contents* to the input from job 1.
   Set the ***Non-redundancy level for homologue search*** to **100**
3. Run the MG/MrBump task. This time when the MrBUMP completes you will notice that only one model is loaded. This example requires more careful consideration than the previous one. In this case their are 2 domains identified by MrBUMP. If you look at the MrBUMP results window you will see that the white bar at the top has "Domain 1" and "Domain 2" written in it and in the structures listed below, there is one which overlaps strongly with Domain 1 and several which overlap with Domain 2. We are in fact interested in Domain 2.

Uncheck ***"Show MrBump models"***, all models should disappear.
Change the drop down menu from "Domain 1" to "Domain 2".
Re-check ***"Show MrBump models"***, many models should appear.

4. Set the slider to approximately 3/4 the way to the right (about 75). Click on the *File* and select ***Save all visible to CC4i2 database***. Quit CCP4MG.
5. This output file of this task can now be used as a molecular replacement search model. This will not be covered in this tutorial.