Crank2 pipeline tutorial

November 4, 2014

1 Introduction

This practical will demonstrate the use of CRANK2 in CCP4 6.4.0. CRANK has pipelines for SAD, SIRAS and MAD (with or without native) data, while CRANK2 implements a newer and improved method for SAD and SAD with a partial MR model, often referred to as MR-SAD.

First we will use data from a seleno-methionine protein crystal (pdb code: 1FSE; resolution: 2.7A), originally solved by MAD. In this tutorial, we will solve the structure from the inflection point (SAD) data only using CRANK2.

Then, we will use MR-SAD in CRANK2 to build an RNA polymerase (RNAP) associated factor Spt4/5 bound to RNAP clamp domain (pdb code: 3QQC, resolution: 3.3A). We will start from a partial MR solution (pdb code: 1Y1W, resolution: 4A) containing 2 zinc atoms out of total 4. The MR-SAD pipeline will find the remaining 2 atoms and rebuild the model using the 3.3A anomalous data. Although the original authors were not able to obtain an interpretable map using the anomalous signal only, CRANK2 is also able to build majority of the structure starting from the 2 zinc atoms only (without the protein atoms from the partial model).

2 Instructions

2.1 Running CRANK2 from SAD data

Please follow these steps to solve the structure with CRANK-2 from the GerE inflection point (SAD) data:

- From Automated Search & Phasing CCP4 section, choose the Crank-2 pipeline.
- Specify a title in the "Title" field (e.g. "tutorial GerE from SAD infl." or anything you like).
- Select the file /Users/Shared/crank-tutorial/gere-sad/gere.pir for the "SEQin" field (check the box "Input protein sequence" first if not checked).
- In "MTZ in", select the file /Users/Shared/crank-tutorial/gere-sad/gere_infl.mtz
- In the "Crystal #1" part, select "Amplitudes" option for the "Input" field (generally, intensities are preferable but in this case, the MTZ file only contains amplitudes). The F_infl+/- MTZ labels should be picked automatically.
- Specify "Se" in "Substructure field" and in the field "Number of substructure monomers" enter 2.

- The theoretical anomalous scattering coefficients are calculated from the wavelength from the MTZ file. However, the real coefficients are significantly different in this case, thus change f'=-6 and f"=4.
- If needed, in the "FP+" box select "F_infl(+)", the other fields would be set automatically.
- Run by clicking the Run / Run now button.

2.2 Running CRANK2 with SAD data and an MR partial model (MR-SAD) pipeline

Please follow these steps to solve the Spt4/5-clamp structure with CRANK-2 from the SAD data, starting from 2 zinc atoms from a partial MR model (found by MOLREP):

- Define a new "spt45" project.
- From Automated Search&Phasing CCP4 section, choose the Crank-2 pipeline.
- Specify a title in the "Title" field (e.g. "tutorial clamp MRSAD" or anything you like).
- Select the file /Users/Shared/crank-tutorial/spt45/3qqc.pir in the "SEQin" field (check the box "Input protein sequence" first if not checked).
- In "MTZ in", select the file /Users/Shared/crank-tutorial/spt45/spt45.mtz
- Specify "Zn" in "Substructure field". The fields "Number of substructure monomers" and "Atomic anomalous scattering coefficients" will be filled automatically and do not need to be changed.
- Change the field "No partial model to input" into "Input partial model to find sites and (re)build model"
- Select the file /Users/Shared/crank-tutorial/1Y1W_molrep.pdb into the 'XYZ in' field.
- Click on the "Advanced options" and in step #1, change the RMS threshold for atom picking to 4.0
- Run by clicking the Run / Run now button.

You can also try to build the structure from the anomalous signal of the zinc atoms only by selecting the option 'Build from substructure only (will remove input protein atoms, if any)'. However, this job would take a longer time to finish.

3 CRANK/CRANK2 Output

A graphical output is displayed by double clicking the ccp4i Crank2 job. In addition to that, text log files are outputted by Crank2, as discussed below. Files associated with each step in the Crank pipeline are stored in the corresponding subdirectories. These include scripts used to run the corresponding program.

3.1 FA estimation: Afro/Shelxc

The log file from this step contains information related to the FA values generation.

3.2 Substructure detection: Crunch2/Shelxd/Prasa

Substructure detection programs attempt to finds heavy atom positions using E-values from the previous step. Heavy atom substructure is written into a PDB file produced. The log files outputted provide statistics that scores the solutions found.

3.3 Phasing and substructure refinement: BP3/Refmac

The chosen program outputs expected values for structure factor amplitudes and phases ("best" values - FB, PHIB) and associated figure of merit values (FOM), Hendrickson-Lattman (HL) coefficients (HLA, HLB, HLC, HLD) and difference map (FDIFF, PDIFF) for two enantiomorph structures.

3.4 Hand determination & density modification: Solomon/Parrot

Density modification is first used in "quick" mode for hand determination (together with other statistics) - to select one of the enantiomorph structures. Results from these runs are stored in hand1 and hand2 subdirectories. For structure with the best score, optimized structure factor values (FDM,PHIDM), HL coefficients (HLA,HLB,HLC,HLD) and FoM (FOM) are written into the output mtz file.

3.5 Model building: Buccaneer/Arpwarp

Improved map is input into model building. User can monitor number of built and sequenced residues and number of build chains in the log file. Final map and build structure are output in the mtz and pdb files, respectively.

4 Building a better model

After the model building is finished, you can visualize it in Coot and try to assess model quality in various regions. Unless something went wrong, you should see several protein chains built with some of them sequenced. There are various ways to build a better model with Crank by using its options. If you believe that substructure has been determined correctly then you can skip this step and try to build a better model by looking at the other steps and applying changes to them.