LA-UR- 10-00243

Approved for public release; distribution is unlimited.

Title:	phenix.mr_rosetta: A new tool for difficult molecular replacement problems
Author(s):	Thomas Terwilliger, INST-OFF, LANL Randy Read, University of Cambridge Frank DiMaio, University of Washington David Baker, University of Washington
Intended for:	Web-only article: "Phenix Newsletter"

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

phenix.mr_rosetta: A new tool for difficult molecular replacement problems

Tom Terwilliger[a], Randy Read[b], Frank DiMaio[c] and David Baker[c]
[a]*Los Alamos National Laboratory, Los Alamos NM 87545*
[b]*University of Cambridge, Department of Haematology , Cambridge, CB2 0XY, UK*
[c]*University of Washington, Department of Biochemistry, Seattle, WA, 98195, USA*

Correspondence email: terwilliger@lanl.gov

What is phenix.mr_rosetta?

The *PHENIX* development team is working with the Baker laboratory at the University of Washington to combine the power of Rosetta structure modeling with *PHENIX* automated molecular replacement (MR), model-building, density modification, and refinement. The basic idea is to find MR solutions with `phenix.automr`, rebuild them with Rosetta, including electron density map information, then rebuild those models with `phenix.autobuild`. The combination of Rosetta rebuilding and phenix rebuilding is the key part of this method. MR solutions are found with `phenix.automr` (Phaser), scored with LLG (optionally following Rosetta relaxation), the best solutions are picked and rebuilt with Rosetta including map information, the resulting models are scored with Rosetta, and then rescored with LLG, and the top models are rebuilt with `phenix.autobuild`.

What is phenix.mr_rosetta good for?

`phenix.mr_rosetta` can be very useful for cases where the search model used in molecular replacement is slightly too distant to rebuild successfully with `phenix.autobuild`. It can also be useful in cases where the model is too distant to even find a molecular replacement solution, and pre-refinement with Rosetta can yield an improved search model.

How do I run mr_rosetta?

You can run `phenix.mr_rosetta` in a very automated way, or as a tool to find molecular replacement solutions and to systematically improve them. To run `phenix.mr_rosetta` you need to have both *PHENIX* (any recent version) and Rosetta (development version as of this writing, or version 3.2 or later once available), installed.

The basic inputs for `phenix.mr_rosetta` are pretty simple: (1) a data file with F, SIGF, and freeR flags, (2a) a search model and an alignment file, or (2b) an hhpred file with a list of alignments and PDB file names, and (3) a pair of fragments files that you create and download from the Robetta server.

You can get the `hhpred` file with alignments by pasting your sequence into the server at toolkit.tuebingen.mpg.de/hhpred. This takes about 10 minutes. You can get the fragments files by pasting your sequence into the Robetta server at robetta.bakerlab.org/fragmentsubmit.jsp. This takes a few hours to run, depending on the length of your sequence.

Once you have these files, you simply edit a simple script file for `phenix.mr_rosetta` that specifies these files (and other parameters if you wish). A typical command-line run of `phenix.mr_rosetta` is shown at the right.

```
phenix.mr_rosetta \
   seq_file=seq.dat \
   data=coords1.mtz \
   search_models=coords1.pdb \
   already_placed=True \
   fragment_files = test3.gz \
   fragment_files = test9.gz \
   rosetta_models=20 \
   ncs_copies=2 \
   space_group=p212121  \
   use_all_plausible_sg=False \
   nproc=200 \
   group_run_command=qsub
```

Does phenix.mr_rosetta require a cluster to run?

`phenix.mr_rosetta` does require building a number of Rosetta models, and each model can take from 10-60 minutes to build with a single processor. In

many cases, `phenix.mr_rosetta` can succeed with as few as 20 models in each cycle. This means that a computer with 4 processors can be quite sufficient to run `phenix.mr_rosetta` and can finish in a day or so. In very challenging cases, as many as 2000 models may need to be built (the best models are picked and the density from the top 20% of models is averaged), and this is best carried out on a cluster. `phenix.mr_rosetta` can run on a Sun Grid Engine (SGE) cluster and on a Condor cluster. It may also run on other clusters. To run on a cluster you simply specify the command that you use to submit jobs ("`qsub`" for a SGE cluster for example).

Advanced uses of `phenix.mr_rosetta`

Once you have used `phenix.mr_rosetta` a few times, you will find that you can control where it starts and what it does in quite some detail. You can choose a particular solution that it is working on and have it build Rosetta models for that solution, then write out a table of results that you can examine. This way you can combine your intuition with the scoring that `phenix.mr_rosetta` uses to optimize your search.

You can also pre-refine your search model. This just means running Rosetta modeling on your search model without including any information from the crystallographic experiment. This can be very useful because Rosetta modeling can improve your search model and allow molecular replacement to succeed in cases where it might otherwise fail completely.

Where can I read more?

You can see all about running `phenix.mr_rosetta` in the *PHENIX* documentation.