RockDock – How to use it

Rockdock is a pipeline wrapper that controls submissions of ultra-large libraries for SLURM job managers (i.e. Rocket HPC) that submits, ranks and sort molecules. The application uses openbabel, Seesar and Dock6. The flowchart depicted in figure 1 shows how the docking procedure is done within the rockdock scaffold.

1 – Preparation

Dock6 is used for pose generation. The preparation for setup is done via the default flexible docking protocols explained on the dock6 website. All these steps use the dock6 tools. All this process should be done inside the folder named input\_location (in bold are the names required by rockdock – so when preparing the input use the names in bold):

* Preparation of the surface file (surface.dms), receptor files with charges (receptor.mol2), a receptor in pdb format for Hyde (receptor.pdb) and a ligand center to define the binding site (center.mol2)
* Creating the spheres for docking suing sphgen: an example input file (INSPH) using the default naming.
* Selection of binding site spheres – using sphere\_selector: this step should generate a selected sphere set that will guide your docking (the selected spheres will be named selected\_spheres.sph)
* Generation of the docking box using showbox. Using the previous files, this step will generate a box named box.pdb
* Generate the grid: generate the grid using the box that surrounds the binding site. Please use the grid.in input file located in the input\_location folder – this will generate a set of files with the prefix grid (grid.nrg and grid.bmp)

If you want more control over each tranche submission, add your email to the lines “#SBATCH --mail-user=” inside the rockdock\_prep.

With he generated input files located in the input\_location folder and a selected ligand library in a mol2 format, you can prepare the submission by running :

./rockdock\_prep <lib.mol2> < the number of molecules per thread>

This step will generate a new folder (dock\_output). This folder will have a series of folders with separated sets for docking – each will be used to a single thread (called a submission tranche)

In sequence, for submission of the pose generation via Dock6, run

./rockdock\_submitter.sh

This script will submit n tranches ( n = (number of molecules in your library)/(size of the tranche)), each to a single thread.

With this stage finished, you have two options: 1) rescore using Hyde a subset of the generated poses 2) rescore all generated poses using Hyde.

Option 1) Subset from dock6

To generate a subset of dock6 output molecules, you first need to generate a concatenated set of all molecules generated within the dock\_output folder. This can be done via:

cat “submission\_folder”/dock\_output/set\_\*/flex\*mol2 > dock6\_fullposes.mol2

Inside the tools folder, a script called “ dock6filter” can generate a sorted subset of the generated poses. To use this script, do:

dock6\_filter -i dock6\_fullposes.mol2 -o dock6\_top\_subset.mol2 -r (gscore or cluster) -n (number of molecules in the subset – usually 1000)

If you use gscore, you will sort molecules by the lowest AMBER interaction energies calculated by dock6. If you use cluster, it will sort by the cluster size of the pose.

Option 2) Rescore all generated poses using Hyde.

If you want to rescore and sort all poses using Hyde (which is time consuming per molecule), you can run.

./rockdock\_rescorer

This command will submit a hyde scorer job per generated tranche.

After all, tranches are finished, you may want to sort the generated sdf files to evaluate the top n-th molecules. Similar to dock6filter, there is a script inside the tools folder named seesar\_filter.

To use it, first generate the concatenated lib for all molecules after hyde scoring using

cat “submission\_folder”/dock\_output/set\_\*/flex\*sdf > seesar\_rescore\_fullposes.sdf

seesar\_filter -i seesar\_rescore\_fullposes.sdf -o seesar\_rescore\_subset.sdf -n (number of molecules in the subset – usually 1000)