

## **Protocol Capture**

Rosetta can be obtained through [www.rosettacommons.org](http://www.rosettacommons.org)

All files associated with this protocol capture is provided in the demos/protocol\_capture/rosettaligand\_ensemble/ directory of the Rosetta distribution. This protocol has been tested to work with Rosetta version d978e6f, released August 22<sup>nd</sup>, 2017.

Examples commands for this protocol are numbered in the *commands* file of the protocol capture folder and referenced as (1), (2), (3)...etc.

## **Starting Files:**

The raw starting files are a single target protein receptor structure in PDB format, and a series of ligands in SDF format. The protein structure can be prepared from an experimentally determined structure or from homology modeling. The receptor structure used in this example is the p53 core domain bound to a stabilizing small molecule (PDB: 4AGQ). This file can be found in /inputs/ as protein.pdb. The ligand series should share a core scaffold by which the ligands can be aligned. This example contains five congeneric ligands, but any number between three and eight is a reasonable use case.

## **Ligand Preparation:**

PyMol pair fitting is an easy way to manually align ligands by minimizing distance between core scaffold atoms. Automated ligand alignment tools may also be used but generally do not perform as well compared to manual inspection. Examples of aligned ligands can be found in the /prep/aligned\_ligands/ directory.

Each ligand must have its own conformational library generated prior to using RosettaLigandEnsemble. Conformer generation was done using the BioChemicalLibrary (BCL) <http://www.meilerlab.org/servers/bcl-academic-license> is a suite of software tools readily available for academic users. Other software for conformer generation may also be used but the outputs need to be converted to SDF files.

To generate conformers using the default settings, use command (1). Conformer generation can also be customized to use the PDB or CSD libraries, a greater range of rotamers, or a structural comparison filter to remove similar conformers. For a full list of these options, see the help menu with command (2). Examples of generated conformers files are in /prep/conformers/

### **Params File Preparation**

Rosetta requires params file to properly handle small molecule ligands. Prior to this step, join the aligned ligand structure with the corresponding conformers into a single SDF file such that the aligned structure is first in the file. This will insure that the inputs will maintain the core scaffold alignment when generating the conformers. Examples of these joined files are in /prep/make\_params/

Since PDB files use three digit residue codes and single digit chain designations, it's helpful to assign a code for each ligand file. The example uses the ligands.list file to label each ligand as residues 00B through 00F and corresponding chains B through F. This file also contains pK values for each ligand binding to the target receptor.

To make params files for each ligand, use command (3). The `molfile_to_params` python script is included in the `/main/source/scripts/python/public/` directory of the Rosetta distribution. Running the script without any input prints out the help menu.

For each ligand, command (3) generates a PDB file containing the single aligned ligand structure, a PDB file containing the remaining ligand conformers, and a params file containing connectivity and charge information for the ligand. Examples of these files can be found in the `/prep/rosetta_inputs/` directory using the previously discussed letter designations. If you wish to incorporate SAR during docking, then use a text editor to add `NUMERIC_PROPERTY AFFINITY #####` to the end of the params file, where `#####` represents an SAR measurement of the user's choice. Rank correlation is used in SAR mode and hence units only need to be self-consistent.

### **Input file organization**

For RLE runs, it is convenient to prepare a single PDB file containing the aligned ligands but concatenating the individual ligand PDB files. The conformer and params do not need to be joined. This is done as `ligands.pdb` in the `/inputs/` folder. You'll also find the previously prepared protein receptor PDB in the same directory.

In addition to structural files, a RosettaScripts XML file and a Rosetta options file. The XML file describes the custom protocol to be used by Rosetta. Details of how to setup an XML file and the meaning of the individual tags can be found by searching the documentation website <https://www.rosettacommons.org/docs/latest/>. The example `dock.xml` provided uses the settings

from the benchmark. Actual application use may require the user to alter these values according to biological context. The defined scoring function is based on the existing RosettaLigand scoring function, but may be substituted in the XML script. The provided options file defines Rosetta input and output directories along with a number of sampling parameters. A full options list is available on the documentation website. The `ligand_ensemble` option is necessary to use RLE; a weight of 0 can be used to run RLE without taking SAR data into consideration.

Run command (4) to perform a single simulation and generate a set of RLE models. Each simulation will produce X models, where X is the number of input ligands. These example output models are in the `/outputs/` directory along with a `score.sc` scorefile.

### **Output and analysis:**

Individual protein-ligand predicted structures are labeled by a chain and a number designation, B\_1.pdb through F\_1.pdb. Structures with the same numeric label are based on the same docking simulation and have a common binding pose. The protein interface contacting each ligand are optimized independently. The `score.sc` file contains all score terms for each simulation across a single row. Generally, individual ligand interface scores are used to rank models, with a negative score indicating a better model. These ligand interface scores are listed as `interface_delta_*`, where \* is the single letter ligand chain ID. The values are appended at the end of each output PDB, and also in the scorefile for each protein-ligand pair. One suggestion is for the end user to examine the top ten percent of models for each pair.