**Ligand Function Prediction Guide**

**9/14/2022**

**Author: Gregory L. Szwabowski**

**Contact: gszwbwsk@memphis.edu**

The following guide outlines the steps required to take docking results, extract interaction energies and types, and create a text file that can be used as input for ligand function prediction with a random forest model. Files necessary for this tutorial are located in a .zip file available at:

<https://github.com/gszwabowski/guides/blob/master/score-based_ph4_tutorial_files.zip>

Prerequisites

|  |  |  |
| --- | --- | --- |
| **Step** | **Description** | **Figure** |
| 1 | After docking has concluded, the top scoring docked poses for each ligand must be extracted using the *get\_topscored\_pose\_by\_mseq.svl* script. First, open the script (located in the “scripts” folder) using the *Edit…* button and then save and load the script by clicking *SVL 🡪 Save and Load*. Set your CWD to a the directory containing docking results and ensure that databases containing docking results are the only .mdb files located in the directory. This script will work if only one or multiple .mdb files containing docking results are present in a directory. Use the command:  get\_topscored\_pose\_by\_mseq [prefix]  where prefix is the name you wish to suffix the database containing top scoring poses for each docked ligand in each database. Top scored poses for each docking database will be located in .mdb files titled ‘prefix\_topscored\_poses.mdb’. |  |
| 2 | Once top scored poses per mseq are extracted from each database, they will need to be imported into a single database. After setting your CWD to where you wish to aggregate the top scored poses from each database, click *File* 🡪 *New* 🡪 *Database…* and name the database ‘XXX\_topscored\_poses.mdb’, where ‘XXX’ is the name of the target that was docked into. In the new database’s viewer, click *File* 🡪 *Import…* and then click the + icon to select files for database import. For each database that was generated in step 1, select the database and then click *Add*. Once all databases generated in step 1 are added, click *OK and* then *OK* to import them to the ‘XXX\_topscored\_poses.mdb’ database. | *Database with top scored poses imported* |
| 3 | Once the ‘XXX\_topscored\_poses.mdb’ database is filled with the entries from databases generated in step 1, a field denoting the target needs to be added to the database. From the ‘XXX\_topscored\_poses.mdb’ database viewer, click *Edit* 🡪 *New* 🡪 *Field*. Set the new field type to ‘char’, the Name to ‘Target’, and the Value to the name of your target. Click *OK* to create the field in the database. |  |
| 4 | Next, each entry in the database will need to be numbered with the *loopnumber.svl* script. Save and load the *loopnumber.svl* script and use the command  loopnum [mdb]  where mdb is the ‘XXX\_topscored\_poses.mdb’ file. This script will create an index field that numbers each of the entries in the database of aggregated top scoring poses. |  |
| 5 | In the ‘XXX\_topscored\_poses.mdb’ database viewer, select the *Target* and *Index* fields and then click *File* 🡪 Save As… and ensure that the *Selected Fields Only* box is checked. Save this file as ‘XXX\_tm\_indexing.mdb’, where XXX is the name of your target. This |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |