# Tutorial on batch format conversion of receptors and ligands, and continuing run of virtual screening by MolAICal

Qifeng Bai
Email: molaical@yeah.net
Homepage: https://molaical.github.io/baiqf.html
School of Basic Medical Sciences
Lanzhou University
Lanzhou, Gansu 730000, P. R. China

# 1. Introduction

Virtual screening and evaluation tasks require batch processing of receptor and protein formats. Besides, Some operations and unexpected events may cause the virtual screening task of drugs to be interrupted. Therefore, this tutorial introduces the way to use MolAICal for batch conversion of receptor and ligand formats and continuing run of drug virtual screening jobs. For more detailed MolAICal, please visit https://molaical.github.io or https://molaical.gitee.io.

# 2. Materials

### 2.1. Software requirement

1) MolAICal: https://molaical.github.io or https://molaical.gitee.io
2) UCSF Chimera: https://www.cgl.ucsf.edu/chimera
**Notice:** make sure the installation of MolAICal and Chimera is correct!

### 2.2. Example files

1) All the necessary tutorial files are downloaded from:
https://github.com/MolAICal/tutorials/tree/master/023-convert_rec_lig_batch

# Part I. Batch converting of receptors and ligands

In this part, MolAICal supplies a script for converting the files of receptors and ligands in batches.

### 1. Batch converting of receptor files

### 1.1 Go to the virtual environment

Download material files: “023-convert_rec_lig_batch” from Github or Gitee, and go into the directory “023-convert_rec_lig_batch” as follows:
#> cd 023-convert_rec_lig_batch

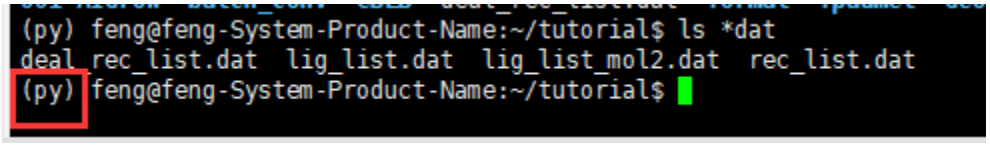
**And then, go to the virtual environment:**

In Windows system:
#> D:\MolAICal-win64\mtools\py\Scripts\activate.bat

In Linux system:
#> source /home/feng/tutorial/MolAICal-linux64/mtools/py/bin/activate

**Note:** Please replace "D:\MolAICal-win64" or "/home/feng/tutorial/MolAICal-linux64" with your **real path** of MolAICal.

If you go into the virtual environment, you will see a graph similar to the following:




### 1.2 Getting receptors' and ligands' list

Here, the below command will get the ligands path list:

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "batch_conv" -ds -ik "_ligand.sdf" > lig_list.dat

Here, the below command will get the receptors path list:

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "batch_conv" -ds -ik "_protein.pdb" -nk "_protein.pdbqt" > rec_list.dat

**Note:** Please replace "MolAICal-xxx" with your **real path** of MolAICal. It should set your real path for "batch_conv" after "-sd".

**-ds**: Whether search files with appointed characters, if input "-ds" only, it means True to search files with appointed characters.

**-sd**: The directory for searching.

**-ik**: The keyword is for searching.

**-nk**: The keyword is excluded from the search.

For more detailed input parameters for commands, please see 'Appendix' in this document.

### 1.3 Converting in the batch of receptor files

################################################################################

**Option:**

If you think your receptor files are not necessary to be further handled, please skip this step and go into next step. Sometimes, receptors contain unnecessary ions and water or the wrong hydrogens. In this option, MolAICal will be used to deal with receptor files in batches.

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -m -dr -cp "D:\Chimera 1.15rc\bin\chimera.exe" -ir "rec_list.dat"

**Note:** Please replace "MolAICal-xxx" or "D:\Chimera 1.15rc\bin\chimera.exe" with your **real path** of MolAICal or Chimera. "rec_list.dat" should be replaced with your real path.

The default dealing receptor name is "rec_deal.pdb". If users want to use the handled results, please use the below command to obtain the paths of the handled receptor files:

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "batch_conv" -ds -ik

"rec_deal.pdb" > deal_rec_list.dat

**Note:** If you use this step, please substitute "deal_rec_list.dat" for "rec_list.dat" in the following steps.
##########################################################################

If you **omit** the **above option**, convert in the batch of receptors as follows:
#> python D:\MolAICal-xxx\scripts\molaical_batch.py -rc -ir "rec_list.dat" -irf "pdb" -orf "pdbqt"

### 1.4 Converting in the batch of ligand files

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -lc -il "lig_list.dat" -ilf "sdf" -olf "pdbqt"

## Part II. Batch converting of receptors and ligands by MolAICal command

In this part, It calls the MolAICal command to convert the files of receptors and ligands in batch. However, this part tutorial has the limitation in the conversion formats for the files of ligands and receptors. For example, It cannot support the SDF format of ligands and receptors. You can choose to study part I or part II for your studies.

### 2.1 Converting in the batch of receptor files with the MolAICal command

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -m -rc -ir "rec_list.dat" -mp "D:\MolAICal-xxx\molaical.exe"

**Note:** Please replace "MolAICal-xxx" with your **real path** of MolAICal. "rec_list.dat" should be replaced with your real path.

### 2.2 Converting in the batch of ligand files with the MolAICal command

In this part, the MolAICal command does not support the SDF format, so the mol2 or PDB format is selected for conversion. Here, Getting the list of ligand files in mol2 format as an example:

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "batch_conv" -ds -ik "_ligand.mol2" > lig_list_mol2.dat

Then, converting in batch of ligand files:

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -m -lc -il "lig_list_mol2.dat" -mp "D:\MolAICal-xxx\molaical.exe"

**Note:** Please replace "MolAICal-xxx" with your **real path** of MolAICal. "lig_list_mol2.dat" should be replaced with your real path. It should set your real path for "batch_conv" after "-sd".

## Part III. Run-flat of virtual screening by MolAICal

Please ensure that in the virtual environment. **See part 1.1 for going to the virtual environment in this tutorial document**.

Go into the material directory as follows:
#> cd 023-convert_rec_lig_batch/vs_continue

Firstly, obtain the screened results from folders, the screened results by MolAICal have the suffix "_out.pdbqt". So get the list of screened results as follows:

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -sd "F:\023-convert_rec_lig_batch\vs_continue\vs_results" -ds -ik "_out.pdbqt" > out_list.dat

**-ds**: Whether search files with appointed characters, if input "-ds" only, it means True to search files with appointed characters.
**-sd**: The directory for searching.
**-ik**: The keyword is for searching.

And then, check file "out_list.dat", if no errors, move results into one folder as follows:

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -cvs -il F:\023-convert_rec_lig_batch\vs_continue\out_list.dat -ol F:\023-convert_rec_lig_batch\vs_continue\target

**-cvs**: Whether prepare for continuing to run virtual screening. If input "-cvs" only, it means True to prepare for continuing to run virtual screening jobs.
**-il**: The file of ligands list.
**-ol**: In the section of continuing virtual screening when using "-cvs", it is folder for storing.

**If the screened molecules are moved into the new place, the remaining molecules can continue to run in the VS way of MolAICal.**

**Option**: sometimes, there may be huge numbers of molecules in one folder. To reduce the number of molecules in one folder, it can assign the number of molecules in one folder as follows:

#> python D:\MolAICal-xxx\scripts\molaical_batch.py -cvs -dp 1 -il F:\023-convert_rec_lig_batch\vs_continue\out_list.dat -ol F:\023-convert_rec_lig_batch\vs_continue\target

**-cvs**: Whether prepare for continuing to run virtual screening. If input "-cvs" only, it means True to prepare for continuing to run virtual screening jobs.
**-il**: The file of ligands list.
**-ol**: In the section of continuing virtual screening when using "-cvs", it is folder for storing.
**-dp**: In the section of continuing virtual screening when using "-cvs", it is the storing number of screened molecules in every folder. The default value is "None" which means to store all screened molecules in one folder. It can be 1, 2, or 3… which represents the storing number of molecules in every folder.
Otherwise, it represents the depth of searching directories, the default value is "None" which represents searching all the directories. It can be 0, 1, or 2… that represents the current, 1, or 2…level directories.

# Appendix

**1. Advance (Option):**

If users want to improve the conversion speed, users can split the input file by the "split" command in Linux, and then submit the the splitted output files. The split command is as follows:

#> split -a 3 -d -l 10 --additional-suffix=.txt lig_list.dat test-

**-a 3**: use "000" for nomenclature, users can modify it according to the requirement.
**-d**: use digit for nomenclature.
**-l 2**: split file every 2 lines, users can modify it according to the requirement.
**--additional-suffix=.txt**: the suffix for the output file.
**lig_list.dat**: input file.

**test-**: the prefix of the output file.

**For Windows:**
Users can split files in Linux, and download the output files in the Windows OS System.

Besides, users can use the "split" command on Windows OS system by installing Gnuwin32 or Git: https://getgnuwin32.sourceforge.net or https://git-scm.com

2. **Detailed parameters:**
The parameters in MolAICal for batch format conversion of receptor and ligand files are as follows:

**-ir**: The file of receptors list.
**-or**: Output names of receptors.
**-irf**: The format for inputted receptor. The format name is the suffix of the molecule, for example, "test.mol", "test.pdb", or "test.sdf" should use the letters "mol", "pdb" or "sdf" after the parameter "**-irf**".
**-orf**: The format for outputted receptor. The format name is the suffix of the molecule, for example, "test.mol", "test.pdb", or "test.sdf" should use the letters "mol", "pdb" or "sdf" after the parameter "**-orf**".
**-il**: The file of ligands list.
**-ol**: In the section of continuing virtual screening when using "-cvs", it is folder for storing. Otherwise, output names of ligands.
**-ilf**: The format of inputted ligand.
**-olf**: The format of outputted ligand.
**-rc**: Whether convert receptor format, if input "-rc" only, it means True to convert receptor format.
**-lc**: Whether convert ligand format, if input "-lc" only, it means True to convert ligand format.
**-lc3**: Whether convert 3D ligand format, if input "-lc3" only, it means True to convert 3D ligand format.
**-ds**: Whether search files with appointed characters, if input "-ds" only, it means True to search files with appointed characters.
**-sd**: The directory for searching.
**-ik**: The keyword is for searching.
**-nk**: The keyword is excluded from the search.
**-fn**: The file contains different keywords for searching.
**-dp**: In the section of continuing virtual screening when using "-cvs", it is the storing number of screened molecules in every folder. The default value is "None" which means to store all screened molecules in one folder. It can be 1, 2, or 3… which represents the storing number of molecules in every folder.
Otherwise, it represents the depth of searching directories, the default value is "None" which represents searching all the directories. It can be 0, 1, or 2… that represents the current, 1, or 2…level directories.
**-mp**: The absolute path of MolAICal executable file.
**-cp**: The absolute path of UCSF Chimera executable file.

**-m**: Whether run the MolAICal command for molecular format conversion, if input "-m" only, it means True to run the MolAICal command for molecular format conversion.
**-dr**: Whether run MolAICAl for dealing with the receptor files. For example, delete ions, and water, and modify hydrogen. If input "-dr" only, it means True to run MolAICAl for dealing with the receptor files.
**-rn**: The writing receptor name.
**-ah**: Whether delete hydrogen. The value is "none", "add", "dah", or "del":
"none" is no action;
"add" means to add hydrogen;
"dah" means to add hydrogen after deleting hydrogen;
"del" means to delete hydrogen, it is the default value.
**-cvs**: Whether prepare for continuing to run virtual screening. If input "-cvs" only, it means True to prepare for continuing to run virtual screening jobs.