Manual

Before using our decoy maker,

1. Please make sure you have installed Pipeline Pilot and Matlab on your workstation.
2. Please beware that the copyright belongs to authors of Howard University and Peking University. If you intend to modify or update, contact the authors via jie.william.xia@gmail.com or x.simon.wang@gmail.com.

The easy-to-follow instructions are as follows.

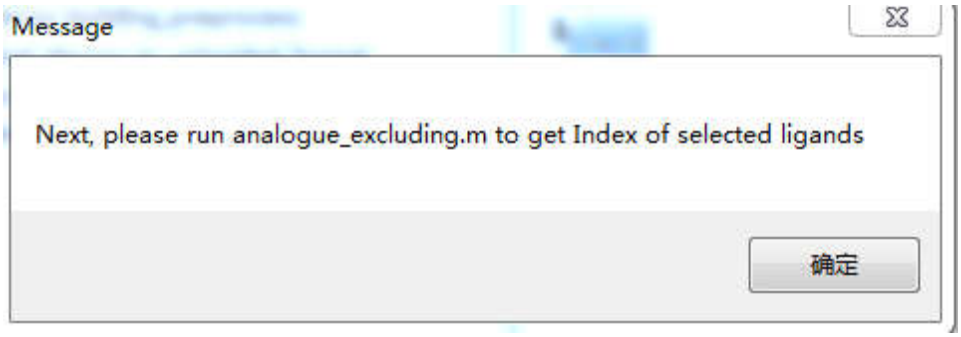
1. *Name user's ligand file as "Raw_Ligands.sdf".*

   Keep in mind that the default working directory in our workflow is **D:\sample-files**, so please make a new folder named "sample-files" in the root directory of D disk. Put "Raw_Ligands.sdf" is in this folder.

2. *Load protocol "Ligand_Data_Curation.xml" into Pipeline Pilot.*

   In this protocol, first you need to input the directory of the "Raw_Ligands.sdf" into the first component following the instruction in the red sticky note.

   Once the protocol is done, you will see a pop-up dialog box (below), click OK to stop running.



   You will get three output files:

   1) Filtered_Ligands.sdf;

2) Ligands_after_curation.sdf;

3) Ligands_after_curation_Y.txt

3. *Load "analogue_excluding.m" in the folder "sample-files" into MatLab.*

Click "run"  in the menu, you will find a matrix named "output", where the indexes of diverse ligands are shown.

Variable Editor - output

output <11x1 double>

	1	2	3
1	1		
2	5		
3	10		
4	14		
5	26		
6	54		
7	82		
8	101		
9	113		
10	114		
11	115		
12			
13			

Open the excel file named "format_for_compound_selection.xlsx" in "d:\sample files", and paste above indexes into the file (see below)

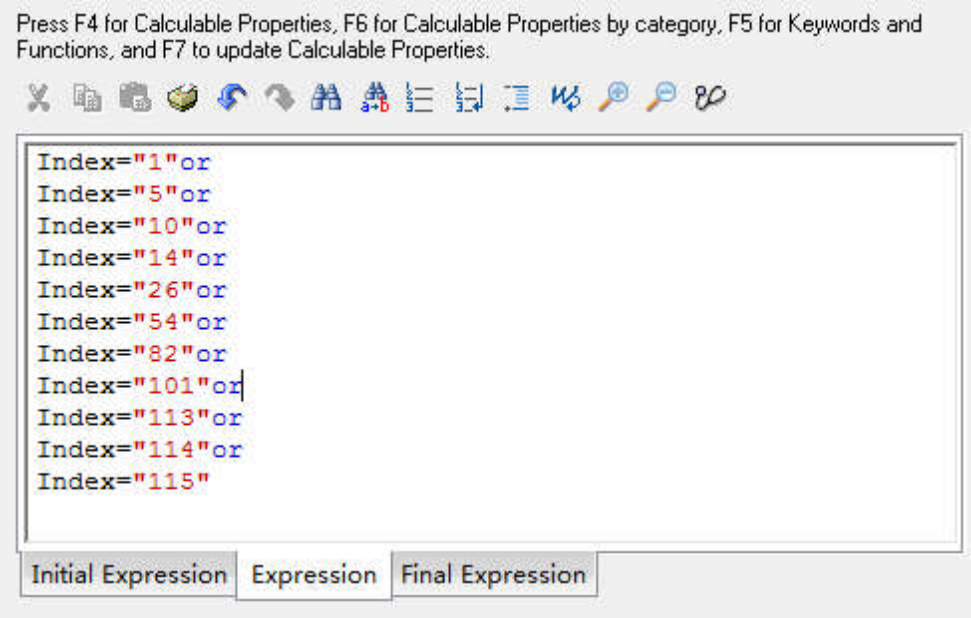
	A	B	C	D	E
1	Index=	"	1	"	or
2	Index=	"	5	"	or
3	Index=	"	10	"	or
4	Index=	"	14	"	or
5	Index=	"	26	"	or
6	Index=	"	54	"	or
7	Index=	"	82	"	or
8	Index=	"	101	"	or
9	Index=	"	113	"	or
10	Index=	"	114	"	or
11	Index=	"	115	"	
12					
13					

4. *Load "Ligand_Property_Calculation_and_scalin.xml" into Pipeline Pilot.*

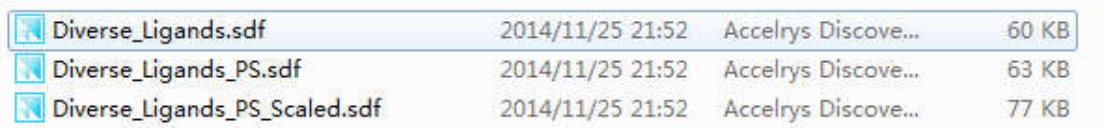
   According to the hints in the red sticky notes,

   1) Input the "Diverse_Ligands.sdf" obtained from the first step;

   2) Paste the index into the component "Set the indexes of selected compounds". Please note there is no SPACE between characters.



   Run this protocol, then you will find three new files in folder "sample-files"



5. *Load "Preliminary_Filter_Part1.xml" into Pipeline Pilot.*

   1) Set the source of decoys "ZINC" in the component with a red sticky note.

   2) Run the protocol. Time varies for different ligand sets. But it usually takes hours.

   Note: Please use "prepare_ZINC.xml" to prepare your original database "ZINC" (download the recent all-purchasable subset from http://zinc.docking.org/browse/subsets/)

6. *Load "Preliminary_Filter_Part2.xml" into Pipeline Pilot.*

   Directly run the protocol and wait till it automatically stops. This step may

also take various hours, depending on the size of the diverse ligand set. You will find “Diverse_Ligands_PF1_TF1_Decoys_final.sdf” in the working directory.

7. *Load “Precise_Filter.xml” into Pipeline Pilot.*

In this step, set the component called “looper” to the size of diverse ligand set. In the sample case, the size is 11. After that, run the protocol.

Note: this step may also take hours to finish.

Congratulations! The benchmarking decoys are successfully built.

You will finally get:

D:\sample-files\benchmarking_decoys.sdf

D:\sample-files\diverse_ligands.sdf

If you find it useful, please cite:

Xia, J.; Jin, H.; Liu, Z.; Zhang, L.; Wang, X. S., An Unbiased Method to Build Benchmarking Sets for Ligand-Based Virtual Screening and Its Application to Gpcrs. J. Chem. Inf. Model. 2014, 54, 1433-1450.

Also, any question or feedback is welcome. Please send emails to jie.william.xia@gmail.com or x.simon.wang@gmail.com.