PyMine 1.0.1 User's Guide

June 25, 2015

Rajan Chaudhari and Zhijun Li*

University of the Sciences, Philadelphia PA 19104

Introduction

PyMine is a PyMol plugin that lets you integrate and visualize chemical and biological data used for drug discovery.

Scope of PyMine

1) To integrate relevant biological and chemical information related to receptor of interest.
2) To visualize 3D structure of the receptor and annotate it with relevant biological information useful for drug discovery
3) PyMine does not carry out any calculations but the data imported through it could be used for drug discovery purpose.

Installation

Requirements:

1) Ubuntu 11.04 or above OR Mac OS X 10.7 or above
2) Pymol 1.7 or above
3) PyMine 1.0.1

On Linux:

1) Download and install PyMOL. http://sourceforge.net/projects/pymol/
2) Download and unzip PyMine from https://github.com/zhijunlilab/PyMine
3) Open PyMol. Install PyMine: Plugins -> Manage Plugins -> Install -> (locate pymine.py file).
4) Restart PyMOl

On Macintosh using MacPyMOL:

1) Rename the "MacPyMOL.app" to "PyMOLX11Hybrid.app" in Applications folder.
2) Install XQuartz found at http://xquartz.macosforge.org/landing/
3) Follow the installatin procedure of plugin mentioned above.

On Windows:

Though PyMine was designed for Linux and Macintosh, it should work without any issues on windows as well. We might add windows support in future updates.

Usage

1) Start PyMOL and go to Plugins -> PyMine
2) Type the PDB ID and chain ID of the interested target and click SUBMIT OR provide a PDB file and the relevant UniProt ID, and then click SUBMIT.
3) To find similar ligands, copy the smile string from the ligand data panel (command-C) and paste it into the smile text box (control-V). Click on Find Similar Ligands button.

Input Panel

Input using PDB ID and Chain ID
Users can provide the PDB ID and Chain ID of the protein of interest to retrieve relevant information from chemical and biological databases mentioned in our publication in BMC Research Notes (2015).

Input using User's own PDB file and UniProt ID
Users can also provide their own PDB file along with the associated UniProt ID. In order to use this function, the following criteria must be met.

1) PDB file and relevant UniProt ID must be provided. If either one is missing then PyMine would try to gather information based on PDB ID and chain ID input.
2) Amino Acid numbering in the PDB file must match with the UniProt numbering. This will make sure the SAV mappings are correct.

Data Panel

Most of the databases are based on wild type protein sequence. If a mutant protein structure is used then users may not get the information they expected.

Protein Tab:
Based on the available information, the Protein tab displays protein information such as:

1) Single Amino Variations (SAVs) and diseases associated with those variations imported from HUMSAVAR database.
2) Binding site information imported from IBIS database.

IBIS provides binding site information such as protein-protein interaction, protein-small molecule, protein-nucleic acid interaction and protein-ions interaction either experimentally determined or predicted using protein homology information.

Example:
PPI 1RTKA 619 624 660 680 681 683 684 685 686 687 724 728
Thrombin_light 3PO1C_3PO1A

Here the first word indicates the type of interaction. In this case it is protein-protein interaction. Second word is the PDB ID of current protein. Next to that is the residue numbers that are part of binding site. Fourth column indicates name of the domain where similar interaction is found and in last column is the PDB IDs including Chain ID involved in the interaction.

Ligands Tab

Ligands tab displays information regarding small molecules, peptides and ions present in the PDB file including their names and smiles strings if applicable.

If assay information is available in ChEMBL database for receptor of interest, then Ligands tab displays ChEMBL IDs of compounds having EC50, IC50 and Ki values <= 10nM.

If approved drugs are available for the receptor of interest then Ligands tab will also display that information.

UniProt Tab

UniProt tab displays UniProt file of the receptor in text format. It contains important structural and biological information such as function of the receptor, relevant diseases, polymorphisms, secondary structure annotations and mapped IDs of other databases.

PDB Tab

PDB tab displays PDB file in text format. Important information about receptor of interest such as expression system, structure quality, sequence variations with respect to UniProt sequence etc. could be gathered from the PDB text file that could get ignored otherwise.

Pathway Tab

Pathway tab displays name and description of the pathways associates with the receptor activity. It also displays the link to KEGG pathway file where important information such as list of genes involved in the pathway and compounds targeting the pathway is available. In addition, a pathway image button is provided that displays the relevant pathway image in a new window.

Similar Ligands Tab

Similar ligands tab displays smiles strings of the similar ligands along with their similarity score obtained from the ChEMBL database.

PyMOL Graphics Visualizer

In PyMOL graphics window, SAVs are annotated in CPK style. Binding sites could be selected by expanding the relevant binding sites list on the right hand side of the PyMOL window.

Numbering of the binding sites in each list is in accordance with the binding sites shown in Protein Tab.

By toggling among the buttons provided on the right panel, users can selectively display certain information over others for visual study.

Output Directory Structure

An output directory is created on the Desktop where all the relevant files are placed for that receptor.

Contact:

Zhijun Li z.li@usciences.edu

Rajan Chaudhari rchaudhari@mail.usciences.edu