Mandeep Mann
2018/11/16

Hit Expansion using Substructure Search & Free Energy Perturbation

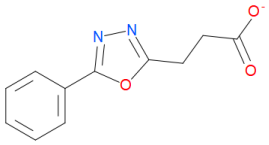
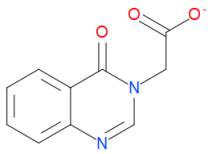
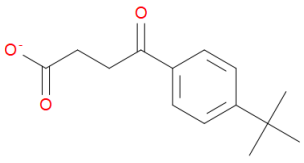
Objective: to identify commercially available chemical analogs of primary hits that I previously crystallized in complex with the zinc finger ubiquitin binding domain (Zf-UBD) of USP5 and to prioritize chemical analogues by free energy perturbation (FEP)

Hit Expansion Method & Results:

A. MolPort/Emolecules Library

1. SMILES of each hit compound (**Table 1**) that was co-crystallized with USP5 Zf-UBD was copied into MarvinSketch (ChemAxon) and changed to SMARTS view. The following constraints were added to each of the compounds:
    a. Constraints on all aromatic rings [a]
    b. Constraints to carboxylate group [OX2H, OX1-1]
2. MolPort→Find Chemicals→SMILES and SMARTS search → search type: substructure→ search options: 1000→SMARTS used for substructure search (similarity=0.8)
3. Emolecules→SMILES used for substructure search (similarity: 0.8)
4. Sdf file of generated analogues was downloaded

Table 1. Compounds used for hit expansion

Name	Structure	SMILES	SMARTS
DAT180		C(Cc1nnc(c2ccccc2)o1)C([O-])=O	C(Ca1aaa(a2aaaaa2)a1)C([OX1-1])=[OX2H]
DAT194		C(C([O-])=O)N1C=Nc2ccccc2C1=O	C(C([OX1-1])=[OX2H])N1C=Na2aaaaa2C1=O
DAT201		CC(C)(C)c1ccc(cc1)C(CCC([O-])=O)=O *c1ccc(cc1)C(CCC([O-])=O)=O Note: *SMILES used for substructure search	CC(C)(C)a1aaa(aa1)C(CCC([OX1-1])=[OX2H])=O *a1aaa(aa1)C(CCC([OX1-1])=[OX2H])=O Note: *SMARTS used for substructure search

B. GLIDE docking

1. PDB files of co-crystal structures uploaded in Molsoft ICM-pro (PDB: 6DXT [DAT180], PDB: 6DXH [DAT201], refmac6 [DAT194]) and missing sidechains were added. The resulting file was saved and opened in Schrodinger Maestro.
2. The protein was prepared using 'Protein Preparation Wizard'. The structure was preprocessed for H-bond assignment. Under the 'Refine' tab, H-bonds are optimized at pH 7.3 and minimized using restrained minimization.
3. Tasks→LigPrep→open sdf file and run job. This converts the ligands to 3D format for docking
4. Tools→Receptor Grid Generation
    a. Receptor: click on ligand to choose it
    b. Site: adjust size of site for docking
    c. Constraints: H-bonds (click to choose H-bonding atoms)
    d. Run job. This will generate a grid file (.zip)
5. Tools→Ligand Dcoking
    a. Open grid file (.zip)
    b. Ligands: upload sdf file from ligprep (.maegz)
    c. Setting: SP (standard precision); ligand sampling: flexible
    d. Check 'enhance planarity of pi groups'
    e. Core: check core comparison
    f. Tolerance: 1.0
    g. Choose core molecule: select part of the ligand ex. The carboxylate group. This will ensure docking occurs at expected site
    h. Check 'skip ligands'
    i. Constraints: check H-bond constraints
    j. Output: write out at most 3 poses/ligand
    k. Run job
6. Docked compound data was exported to a sdf file, opened in ICM-pro and clustered. Compounds were selected based on low complexity (i.e. low molecular weight), docking scores, visual inspection and chemical intuition for each cluster.
7. Selected compounds were then prioritized for ordering using FEP

<u>FEP Method</u>:

1. PDB files of co-crystal structures uploaded in Molsoft ICM-pro (6DXT-DAT180, 6DXH-DAT201, refmac6-DAT194) and missing sidechains were added The resulting file was saved and opened in Schrodinger Maestro.
2. The protein was prepared using 'Protein Preparation Wizard'. The structure was preprocessed for H-bond assignment. Under the 'Refine' tab, H-bonds are optimized at pH 7.3 and minimized using restrained minimization.
3. The crystal structure was duplicated in the workspace, where one entry has only the receptor and solvent and the other has only the ligand.

4. The ligand was mutated using the “Build” function in the workspace. Ligand atoms were selected and minimized. Rapid Torsion Scan was done to choose the best orientation of the bond with an angle/rotation of 72/5.
5. The Schrodinger project file was then opened in Schrodinger Desmond. The Ligand FEP tool was used to import the mutated ligands using the project table. Mutation was selected under atom property. In the Tools section the following job settings were selected:
   a) host: torque-fep master (8)
   b) subjob: torque-g03/g04 (8,4)
   c) ‘Write’ job for each ligand. Everytime a ligand is ‘written’ a new folder is created under the project file.
6. The following script was saved as a file called fep_parallel.csh

```
#################################################################################
##
#!/bin/tcsh
#PBS -N fep_parallel.csh
#PBS -o qsub.out
#PBS -e qsub.err
#PBS -V
#PBS -q qschrod
#PBS -l nodes=1:ppn=1:core2:sgc

foreach d (`cat inp.list`)
cd $d

/opt/schrodinger-desmond/utilities/multisim -HOST fe01 -SUBHOST torque-g03 -m $d.msj -
JOBNAME $d $d.mae -WAIT

cd ../
end
#################################################################################
##
```

7. The fep_parallel.csh file was made executable in the terminal
8. A list of the written files was saved as inp.list
9. The job was started using the following command: nohup ./fep_parallel.csh &

Results:

Please find attached the project files for MolPort and Emolecules substructure hits for each of the compounds (.icb), 3D docked poses of selected compounds after docking (.icb) and sdf files containing final selections with docking scores and FEP free energy difference calculations (.sdf). Summary of the hit expansion for each compounds is shown below in **Figure 1.**

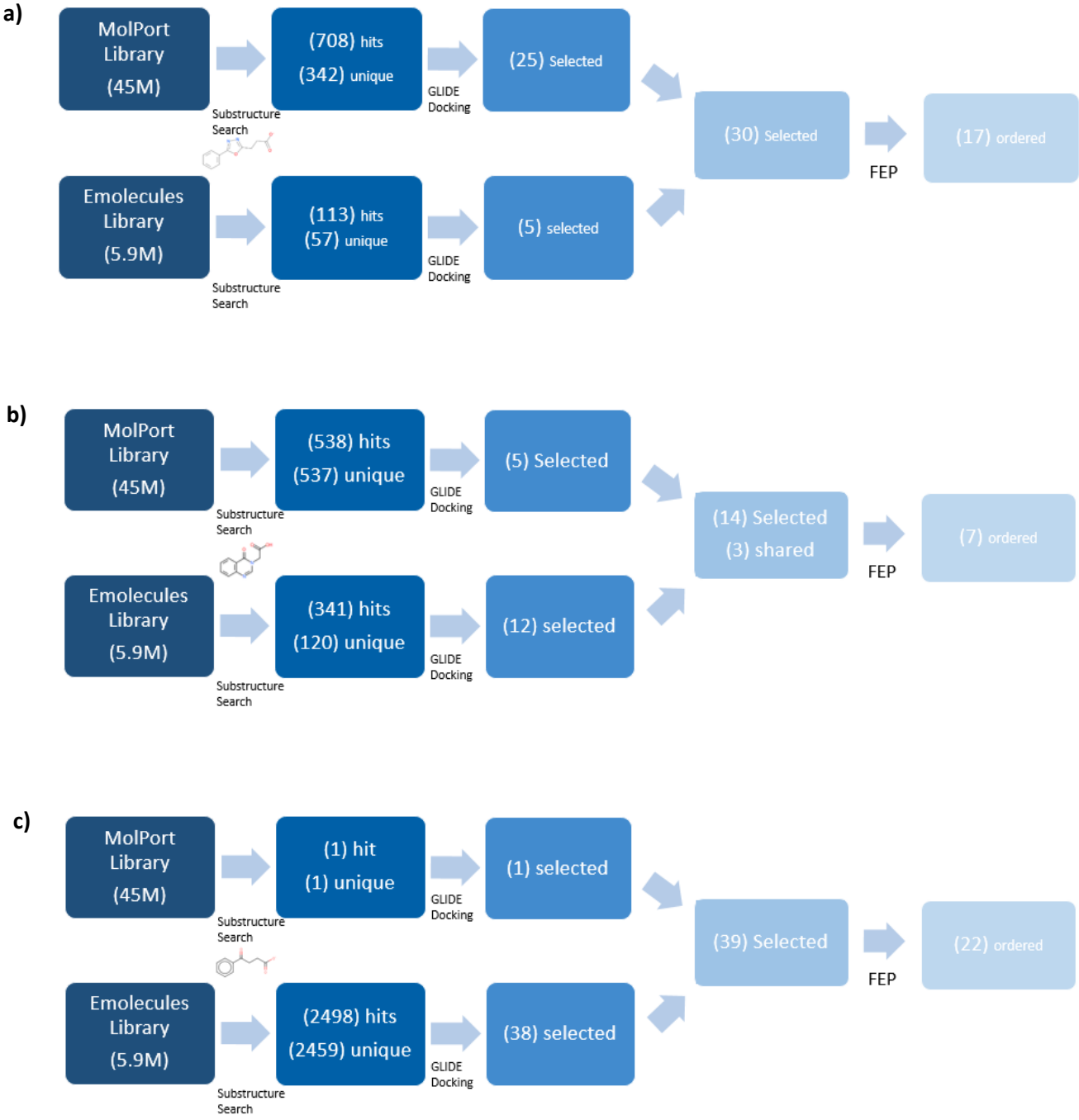


Figure 1. Summary of hit expansion for a) DAT180 b) DAT194 c) DAT201

Mandeep Mann
2018/11/16

Future Directions:

Selected compounds have been ordered through commercial sources (Molport/Emolecules). I'll be testing these compounds against USP5 Zf-UBD using a surface plasmon resonance (SPR) in the near future.