

SLIDE Tutorial

Setting environmental variables before your first use of SLIDE

Make a directory where all your SLIDE output will be written:

```
mkdir slide_data
```

First some environmental variables have to be set to let the computer know where to find the SLIDE-specific executables/commands and where to put the output files generated throughout your SLIDE runs. Using a text editor, add the following lines to your `.personal` file found in your home directory. If you do not have a `.personal` file, create one. (Alternatively, these lines could be added to your `.cshrc` file.)

```
setenv SLIDE_DIR /users/kuhn/slide_v3.0
setenv PATH ${SLIDE_DIR}/bin:${PATH} (to add $SLIDE_DIR/bin to the path)
setenv SLIDE_DATA_DIR ${HOME}/slide_data
```

Then, in the window(s) in which you'll be working, type:

```
source .personal (or source .cshrc)
```

Docking using a biased (ligand-based) template

Target protein : thrombin with ligand-free active site (PDB entry 1vr1)
Template : biased (ligand based) template
Ligand database : a "mixed sample" of 8 known thrombin ligands and 5 organic molecules from the ZINC database (<http://blaster.docking.org/zinc/>) not specific for thrombin.

1. First of all, you need the structure of your target protein. You can download 3D structures of proteins from the Protein Data Bank (<http://pdbeta.rcsb.org/pdb/home/home.do>). For this exercise, copy the target protein file (1vr1.pdb) and the ligand files (*.mol2) used to identify the binding site from `~kuhn/slide_tutorial/` and save it into the `slide_data` directory you created earlier:

```
cp ~kuhn/slide_tutorial/1vr1.pdb ~/slide_data/
cp ~kuhn/slide_tutorial/*.mol2 ~/slide_data/
```

2. To get the ligands that will be docked, copy the file `ligands.tar.gz` from `~kuhn/slide_tutorial/` into your home directory, then unpack it using the commands:

```
cp ~kuhn/slide_tutorial/ligands.tar.gz .
gunzip ligands.tar.gz
tar xvf ligands.tar
```

This will generate a new folder/directory called `ligands_to_be_docked` that will contain the 13 ligands we will dock into the active site of thrombin using SLIDE. (In general, you will want to generate all the low-energy 3D conformers of each ligand candidate before docking them. You can do this by using OpenEye's Omega software, as described in section 2 of `Slide_Quick_Guide_v3.pdf`. For simplicity, in this lab we will just dock one low-energy conformer of each ligand candidate.)

3. The next command performs two tasks:

(I) Sets up the "mixed_sample" ligand database directory and computes the interaction points of the ligands to be docked (.pts files). The database directory will also contain the database files (.db) pointing to the location of the ligand structure files (.mol2).

(II) Sets up the SLIDE data directory structure for the target 1vr1, the template named biased, and the database called mixed_sample. The entire path for the database directory containing the mol2 file has to be given.

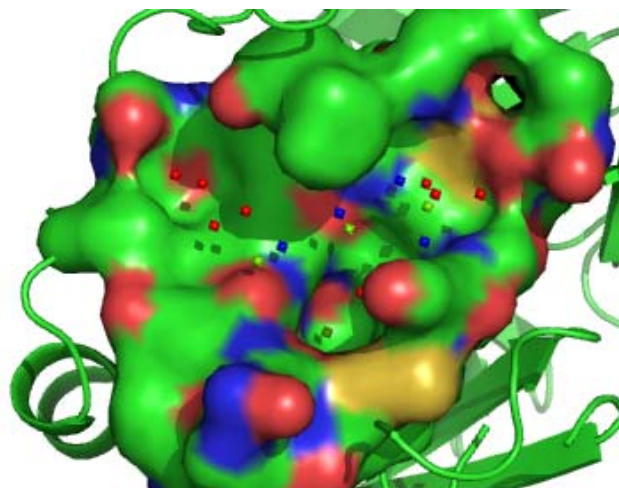
```
setup_dbase 1vr1 biased mixed_sample $HOME/ligands_to_be_docked
$SLIDE_DATA_DIR/1vr1.pdb
```

4. Generate a biased template using two known ligands in their bound orientations. The template points in this case will be ligand interaction points – hydrogen-bonding atoms or hydrophobic interaction centers of the two specified ligands.

```
temp_gen -g 1vr1 biased 2.0 ~/slide_data/1a46_ligand.mol2
~/slide_data/1a61_ligand.mol2
```

This will create the template file in slide_data/1vr1/biased/in with the following numbers of acceptor, donor, and hydrophobic points: 4A, 6D, and 9H. You can look at your template using Pymol. Go into the slide_data/1vr1/ directory, start Pymol and open 1vr1.pdb and template.pdb from the biased/in directory. Color the template points according to their temperature factors and show them as nb_spheres (non-bonded spheres). The temperature-factor (b-factor) coloring scheme of Pymol will color the points the following way:

- hydrophobic - red
- donor – green
- acceptor – blue
- donor/acceptor – cyan



The binding site of thrombin filled with biased template points colored according to interaction type

5. To run SLIDE, type (general form followed by the specific command for this case):

```
run_slide <target> <template> <database> <sphere radius> [overwrite existing
files? y/n] [<outfile>]
```

```
run_slide 1vr1 biased mixed_sample 9.0 y result_1vr1_biased
```

Only those residues of the target protein within a distance of 9 Å (<sphere radius>) of any template point are considered during docking.

7. Analyze the results

The summary of your run, result_1vr1_biased, will be in: ~/slide_data/1vr1/biased/in/. The score components in this file are explained in section 5 of Slide_Quick_Guide_v3.pdf.

The docked ligand orientations will be in: ~/slide_data/1vr1/biased/mixed_sample_ligands/.

The rotated protein side-chains will be in: ~/slide_data/1vr1/biased/mixed_sample_targets/.

Check out which ligands were successfully docked and look at their docked orientations using Pymol. By default, SLIDE returns the best-scoring orientation and conformation for each ligand candidate. To obtain a list of the scores of all the molecules that sterically fit in the binding site, do the following:

```
cd ~/slide_data/1vr1/biased/mixed_sample_ligands
grep "Affiscore" *.mol2 > scores
```

To also list the “ligand efficiency” scores (the average score per atom in the compound, given each docking), use the command:

```
grep "Efficiency" *.mol2 > lig_effic_scores
```

Questions to help you analyze the results. Proceed as far with this analysis as you can comfortably do, during the lab time:

- (a) Were all the known thrombin ligands docked? They are molecules with names starting with a PDB identifier rather than a ZINC (ligand database) identifier.
- (b) Was the top-scoring known ligand docked by SLIDE in a similar orientation to that found in the crystal structure in complex with this ligand? Assess by using PyMol to compare the crystal structure with the SLIDE-docked structure.
- (c) Did any of the ZINC ligand candidates bind with a similar or greater strength to the known ligands, in terms of Affiscore and ligand efficiency score?
- (e) Was the best scoring ZINC molecule isosteric with (occupying the same shape and volume of space) as any of the known thrombin ligands, or did it bind in a novel way?
- (f) For the top-scoring ZINC candidate, list any of its atoms that do not make good contacts with the protein (based on Pymol analysis of protein-ligand hydrogen bonding, for example, or by comparison of ligand atom positions with favored chemical interaction sites represented in the biased template in slide_data/1vr1/biased/in/*.pdb). See note at end of this page.
- (g) Do you expect that this compound will be water-soluble (needed for assaying for enzymatic inhibition), based on whether there are many (soluble) polar atoms or many (less soluble) hydrophobic atoms in the molecule?
- (h) Considering that affinity scores approximate the true ΔG of binding and provide a rough ranking of molecules, and considering the results of your other analyses above, which ZINC molecule(s) look like the best inhibitor candidate(s), and why?
- (i) Assuming that the five ZINC compounds were randomly chosen from the 3 million drug-like compounds in ZINC, would you expect to find tightly binding inhibitors among this set, and why or why not?

Notes on PyMol functionality useful for addressing part (f) above:

Color the ligand atoms differently from the protein atoms (e.g., C atoms pink, N blue, O red, S yellow, H white) using the color option for the ligand. Use the wizard-measurements-polar neighbors feature to find the potential hydrogen bonding neighbors in the protein for each polar (N or O) ligand atom and its polar H's (attached to N or O). Using the h_add function (syntax: h_add <ligand_name>, in the PyMol command window) can help identify the direction of H-bonds and the donor-H-acceptor angle (see paragraph below). Once an atom near the center of the ligand has been selected, using the A (action) menu's Modify-around function in PyMol will allow you to select and show only those atoms within 6 Angstroms of the selected atom, which may make it easier to visualize H-bonds and the neighborhood of the ligand.

Because PyMol is not checking H-bond or salt bridge angles in the polar neighbor function (it is only using a distance cutoff; 3.5 Å is typically the maximum for an H-bond and 4.5 Å is typically the max for a salt bridge), some of the polar atom neighbors in the protein may not be actual H-bonding or salt-bridge partners of the ligand. Typically, the donor-H-acceptor angle should be 120 degrees or greater (the H should be relatively co-linear with the donor and acceptor) to make a good H-bond. Similarly, PyMol can overestimate the number of H-bonds formed, because the distance criteria might list more H-bonds than there are H's to donate or lone pairs to accept. As a conservative estimate for this exercise, consider that each H attached to a polar atom can participate in one H-bond, and choose its acceptor as the atom with the closest distance and most linear donor-H-acceptor angle. Each lone pair of electrons can accept one H-bond. Thus, an NH group can donate one H-bond; an NH₂ group can typically donate two (one for each H); an OH group can donate one H-bond and accept two (one on each of its lone pairs); an =O or O-group can accept two; and an -N= with no H can accept one. (Note that except for special cases, C atoms are non-polar and should not be considered as being involved in H-bonds of any significant strength, and H-bonds involving the two lone pairs acting as acceptors on S atoms in methionine and cysteine are relatively rare.)