SLIDE Tutorial (v. 3.3.5 and 3.4)

Maria I. Zavodszky, Matthew E. Tonero, and Leslie A. Kuhn
Protein Structural Analysis and Design Lab, Michigan State University

Before your first use of SLIDE

See section 1 of the Quick Guide to SLIDE to set environment variables before running SLIDE. Sections 2-6 of the Quick Guide describe SLIDE input preparation, different ways of running SLIDE (and optimizing its parameters for special cases), and interpreting the results. Reading the Quick Guide is important for using SLIDE effectively on your own. This tutorial is designed to take you through common ways of running SLIDE, with the input files already prepared for you. Input file preparation is briefly described on the last page of this tutorial, and the file input/output structure for the first two tutorials is shown on the next-to-last page. The tutorial results also allow you to compare your output with pre-computed output to ensure that your SLIDE installation is running correctly.

Tutorial 1. Docking using a biased (ligand-based) template

One way SLIDE can be used to screen and dock ligands is to match interactions made by previously known ligands with your target protein. In this example, the bound coordinates of each known ligand relative to the target protein are provided as input to template generation. The x,y,z coordinates of the interaction centers in the known ligands (positions in the ligand that can form a hydrophobic, hydrogen bonding, or salt bridge interaction) are then clustered in the binding site to create the docking template (see section 3A in the Quick Guide to SLIDE). Each prospective ligand molecule is then docked by matching the chemistry and position of its interaction centers with those in the ligand-based (biased) template.

Target protein : thrombin with ligand-free active site (PDB entry 1vr1)
Template : biased (ligand-based) template
Ligand database : a mixed sample of 7 known thrombin ligands from crystal structures, 8 organic molecules (ligand candidates) and 1 multi-mol2 file, containing 10 organic molecules (additional ligand candidates).

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://www.rcsb.org/pdb). For this tutorial, copy the file 1vr1.pdb from $SLIDE_DIR/examples/slide_data/1vr1/ and save it into the `slide_data` directory you created earlier:

   ```
   cp $SLIDE_DIR/examples/slide_data/1vr1/1vr1.pdb ~/slide_data/
   ```

2. For the database of molecules to be screened, make a directory in your home directory called `ligands_to_be_docked`, and copy into it all the mol2 files from the directory called $SLIDE_DIR/examples/slide_data/databases

   ```
   mkdir $HOME/ligands_to_be_docked

   cp $SLIDE_DIR/examples/slide_data/ligands_to_be_docked/*.mol2 \
   $HOME/ligands_to_be_docked
   ```

(Note that "\" is the Unix continuation character, indicating that the command continues on the following line. Instead of using the continuation character, you can type the entire command on one line and let the window wrap your text.)

3. The next command performs two tasks:

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).

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 should 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 to define the binding site:

```
temp_gen -g 1vr1 biased 2.0 $HOME/ligands_to_be_docked/1a46_ligand.mol2
$HOME/ligands_to_be_docked/1a61_ligand.mol2
```

This will create the `template` file in `$SLIDE_DATA_DIR/1vr1/biased/in` with the following numbers of acceptor, donor, and hydrophobic points: 4A, 6D, and 9H. You can look at your template using PyMOL (or any other molecular graphics program). Go into the `$SLIDE_DATA_DIR/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. The default coloring scheme of PyMOL will color the points the following (rather unintuitive) way: hydrophobic - red, donor – green, acceptor – blue, donor/acceptor – cyan. Alternatively, you can separately read in the hydrophobic, donor, acceptor, and donor/acceptor sets of template points using the hphob.pdb, donor.pdb, acceptor.pdb, and doneptor.pdb files in the same directory, and color each set/type of template point as you wish.
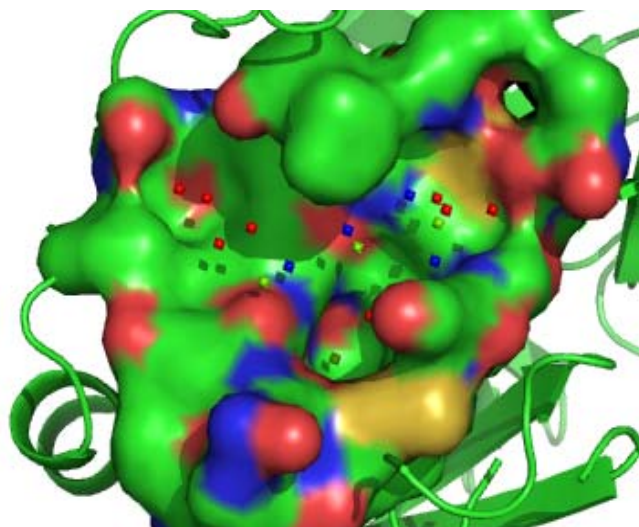
The binding site of thrombin filled with biased template points colored according to interaction type

6. Run SLIDE:

```
run_slide 1vr1 biased mixed_sample 9.0 y result_examples_1vr1_biased
```

Analyze the results. The summary of your run, result_examples_1vr1_biased, will be in the directory:
`~/slide_data/1vr1/biased/in/`

The docked ligand orientations will be in the directory:
`~/slide_data/1vr1/biased/mixed_sample_ligands/`

The rotated protein side–chains will be in the directory:
`~/slide_data/1vr1/biased/mixed_sample_targets/`

By default, SLIDE will output the best predicted orientation of a ligand, which may be viewed using PyMOL. To sort the successfully docked ligands by their predicted binding affinity values, run the following command:

```
results_table.pl 1vr1 biased mixed_sample 25 > 1vr1_biased_example.results_table
```

which will output the molecules with the top 25 predicted binding affinity values. Compare your `result_examples_1vr1_biased` and `1vr1_biased_example.results_table` files to the ones found in the `examples/slide_data/results` directory.

Tutorial 2. Docking using an unbiased template

With the biased template, molecules were docked based on the ability to interact with the protein similarly to two known inhibitors. For cases in which we do not know the bound orientations of known inhibitors, or want to explore new interactions, we can instead use an unbiased template. In this case, we specify only the volume of the ligand binding site in the target protein, and SLIDE's template generator will identify favorable positions for ligand groups to make hydrophobic or hydrogen-bonding interactions with the protein.

Target protein   :   thrombin with ligand-free active site (PDB entry 1vr1)

Template   :   knowledge-based (unbiased) template

Ligand database   :   a mixed sample of 7 known thrombin ligands from crystal structures, 8 organic molecules (ligand candidates) and 1 multi mol2 file, containing 10 organic molecules (additional ligand candidates).

1. Set up the `mixed_sample` ligand database and the SLIDE data directory structure for the target: `1vr1`, template: `unbiased`, database: `mixed_sample` as shown. The entire path for the database directory containing the mol2 file should be given:

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

2. Generate an unbiased template, with the following arguments on a single `temp_gen` command line:

```
temp_gen -l 1vr1 unbiased dense 0.5 3.5 \
$HOME/ligands_to_be_docked/1a46_ligand.mol2 \
$HOME/ligands_to_be_docked/1a61_ligand.mol2
```

Note that the unbiased option uses the flag –l (lowercase letter L, not number one). In this case, the two known ligands are only used to define the volume in which to generate a template, and the template points and their chemistry are defined by favorable interactions from the protein's perspective. Template generation may take a few minutes, depending on the size of the binding site and the sampling parameters. A `template` file will be created with the following number of acceptor, donor, hydrophobic, and donor/acceptor points: 29A, 41D, 41H, and 15N. The program also creates separate files for different template point types (acceptor, donor, doneptor, hydrophobic)
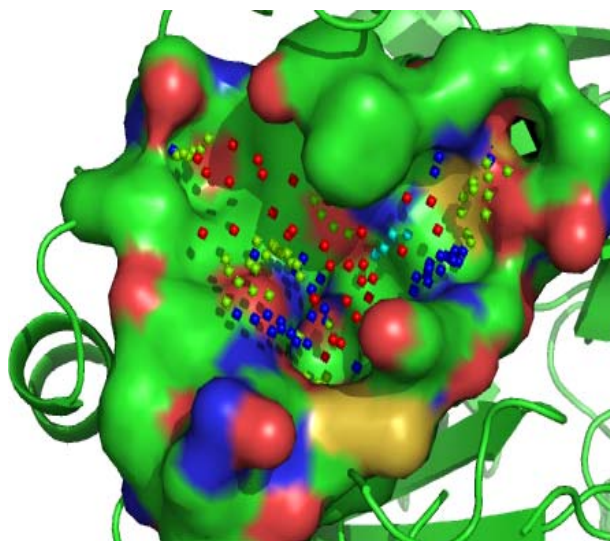
The binding site of thrombin filled with unbiased template points colored by chemical type.

and a `template.log` file, all in the `$SLIDE_DATA_DIR/1vr1/unbiased/in` directory. The `points.usr` file contains the initial points placed on the fine grid, which are subsequently clustered and filtered to create the hydrophobic template points.

3. Define key points in the template file corresponding to points deep in the thrombin binding pocket. By default, SLIDE marks all hydrogen-bonding points as key points (so that at least one hydrogen bond position is matched in the binding site), but the key points can instead be marked by the user so that SLIDE filters for molecules that can dock in a certain part of the binding site. Edit the `template` file created in the `$SLIDE_DATA_DIR/1vr1/unbiased/in` directory by marking only those 11 points with a * that are similarly marked in the file `$SLIDE_DIR/examples/slide_data/1vr1/unbiased/in/template`. (Note that this file is simply called `template`, rather than `template.pdb`, which is provided for display purposes.) These key points are template points within 7.0 Å of the CG carbon of Asp189 in the specificity pocket of thrombin. You could

Docked orientations of a known ligand (from x-ray structure 1tbz) in the binding site of thrombin. Figure created with PyMOL.

alternatively define a set of key points on your own by reading in the unbiased template along with the protein and identifying the residue numbers of the template points you would like to use as key points. Then, star only those points in the template file.

4. Run SLIDE using the database and template you have prepared:

   `run_slide 1vr1 unbiased mixed_sample 9.0 y result_examples_1vr1_unbiased`

5. Analyze your results:

   To sort the successfully docked ligands by their predicted binding affinity you can run the following command:

   `results_table.pl 1vr1 unbiased mixed_sample 25 > \`
   `1vr1_unbiased_example.results_table`

   which will output the molecules with the top 25 predicted binding affinity values.

   - Compare the `result_examples_1vr1_unbiased` from `$SLIDE_DATA_DIR/1vr1/unbiased/in/` and the `1vr1_unbiased_example.results_table` file just generated with the biased results in the `examples/slide_data/results` directory.

   - You can visually inspect the docked ligand orientations (mol2 files) generated with SLIDE, found in `$SLIDE_DATA_DIR/1vr1/unbiased/mixed_sample_ligands/`, with the protein side chains rotated by SLIDE, found in `$SLIDE_DATA_DIR/1vr1/unbiased/mixed_sample_targets/` using PyMOL. The original protein target can be read in from `1vr1.pdb`.

Tutorial 3. Docking using a merged template

Target protein : thrombin with ligand-free active site (PDB entry 1vr1)

Template : knowledge-based (unbiased) template merged with ligand-based (biased) template

Ligand database : same mixed sample of 7 known thrombin ligands from crystal structures, 8 individual organic molecules not known to be thrombin ligands, and 1 multi mol2 file, containing 10 organic molecules not known to be ligands

It is sometimes useful to combine the biased and unbiased templates, incorporating all information about favorable interaction sites from the protein perspective and from known ligands. This tutorial also shows how to reuse an existing ligand database for a new SLIDE run, using the script `slide_setup.pl`.

1. Set up the SLIDE directory tree structure for the target: `1vr1`, template: `merged`, database: `mixed_sample` as follows:

   `slide_setup.pl 1vr1 merged mixed_sample`

2. Concatenate the unbiased and biased templates from Examples 1 & 2 into a new template file in the merged directory:

   `cat $SLIDE_DATA_DIR/1vr1/unbiased/in/template \`
   `$SLIDE_DATA_DIR/1vr1/biased/in/template > $SLIDE_DATA_DIR/1vr1/merged/in/template`

3. Edit the template file created in the `$SLIDE_DATA_DIR/1vr1/unbiased/in` directory by marking only those 7 points with a * that are similarly marked in the file `$SLIDE_DIR/examples/slide_data/1vr1/merged/in/template`. These key points are template points from either the biased or unbiased templates within 7.0 Å of the CG carbon of Asp189 in the S1 specificity pocket of thrombin. For each ligand, at least one of the key points must be matched.

4. Run SLIDE:

   `run_slide 1vr1 merged mixed_sample 9.0 y result_examples_1vr1_merged`

5. Run results_table.pl:

   `results_table.pl 1vr1 merged mixed_sample 25 > 1vr1_merged_example.results_table`

   The summary of your run, `result_examples_1vr1_merged`, will be in the directory:
   `$SLIDE_DATA_DIR/1vr1/merged/in/`

   The `results_table` file will be in the directory in which you executed the `results_table.pl` command.

   The docked ligand orientations will be in:
   `$SLIDE_DATA_DIR/1vr1/merged/mixed_sample_ligands/`

   The rotated protein side chains will be in:
   `$SLIDE_DATA_DIR/1vr1/merged/mixed_sample_targets/`

- Compare the `result_examples_1vr1_merged` from `$SLIDE_DATA_DIR/1vr1/merged/in/` and the `1vr1_mergeded_example.results_table` files with the merged results in the `$SLIDE_DIR/examples/slide_data/results` directory.

- Visually inspect the docked ligand orientations (mol2 files) you generated with SLIDE found in `$SLIDE_DATA_DIR/1vr1/merged/mixed_sample_ligands/` and the corresponding rotated protein side chains found in `$SLIDE_DATA_DIR/1vr1/merged/mixed_sample_targets/` using PyMOL. Read in the protein target (`1vr1.pdb`) and display the surface. Compare the dockings to those found in the runs you did with ligand-based and unbiased templates.

Tutorial 4. Docking using ligand conformers

Target protein   : thrombin with ligand-free active site (PDB entry 1vr1)
Template   : knowledge-based (unbiased) template
Ligand database : same mixed sample as previous tutorials, all with conformers generated using Omega 2.1

In a typical SLIDE run, you will want to screen all the low-energy 3-dimensional conformers for each molecule to represent their full flexibility. We recommend using Omega for conformer generation (OpenEye Scientific Software, Santa Fe, NM), which is freely available to academic groups. SLIDE will then make small adjustments in ligand single bonds and protein side chains during the docking process.

1. For the database of ligand conformers, make a directory in your home directory called `ligands_to_be_docked_confs`, and copy into this directory all the mol2 files from the `$SLIDE_DIR/examples/slide_data/ligands_to_be_docked_confs` as follows:

   mkdir $HOME/ligands_to_be_docked_confs

   **cp $SLIDE_DIR/examples/slide_data/ligands_to_be_docked_confs/*.mol2 \
   $HOME/ligands_to_be_docked_confs**

2. Set up the `mixed_sample_confs` ligand database and the SLIDE data directory structure for the target `1vr1`; template, `merged`; and database, `mixed_sample_confs`, as shown. The entire path for the database directory containing the mol2 file should be given.

   **setup_dbase 1vr1 unbiased mixed_sample_confs $HOME/ligands_to_be_docked_confs \
   $SLIDE_DATA_DIR/1vr1.pdb**

   2458 molecular conformers should be found.

   [Note: these conformers were generated using Omega 2.1 using the following command:

   **omega2 –in <ligand>.mol2 –out confs_<ligand>.mol2 –warts true –commentEnergy true
   –includeInput true –buildff mmff94s –fraglib /openeye/data/omega2/fraglib.oeb.gz**

   Then the MOLECULE name of the first entry (the included input after the `–in` flag) had the conformer ID "_0" added using the script `mol2_add_conf_0.pl` found in the `$SLIDE_DIR/bin` directory. (If the `$PATH` was set properly, all SLIDE commands should run by just typing the name of the command).]

3. Run SLIDE:

   **run_slide 1vr1 unbiased mixed_sample_confs 9.0 y \
   result_examples_1vr1_unbiased_confs**

4. Analyze your results:

To sort the docked ligands by their predicted binding affinity values, use the following command:

```
results_table.pl 1vr1 unbiased mixed_sample_confs 25 > \
1vr1_unbiased_example.results_table
```

which will output the molecules with the top 100 predicted binding affinity values

- Compare `the result_examples_1vr1_unbiased_confs` from the `$SLIDE_DATA_DIR/1vr1/unbiased/in/` and `1vr1_unbiased_confs_example.results_table` files with the `unbiased_confs` results in the `examples/slide_data/results` directory.

Tutorial 5. Interaction matchprint table generation

Target protein   : thrombin with ligand-free active site (PDB entry 1vr1)
Template   : ligand-based (biased) template
Ligand database : same mixed sample as previous tutorials

A useful tool now included in SLIDE is the generation of interaction tables for Matchprint analysis. Using the slide_score module and the –interactions_table flag, SLIDE will output a table of all the protein side chains that make hydrophobic interactions with the ligand, or the specific protein atom that makes a hydrogen bond or salt-bridge with the ligand. By comparing these interactions across all the output molecules of a screen it is possible to get an idea of what interactions are important for the strong binding (and correctly docked) molecules. These interactions can even be used as a way to filter the results, especially if interaction tables have been generated for known ligands of the target protein.

For more information using matchprints together with scoring, see: M. I. Zavodszky, A. Rohatgi, J. Van Voorst, H. Yan, and L. A. Kuhn (2009) "Scoring Ligand Similarity in Structure-Based Virtual Screening", J. Molecular Recognition 22(4), 280-292.

1. First a file will need to be generated containing the names and locations of the docked molecules, e.g.:

```
~/slide_data/1vr1/biased/mixed_sample_ligands/1tbz_ligand_0000.mol2
~/slide_data/1vr1/biased/mixed_sample_ligands/ZINC00000185_0000.pdb
~/slide_data/1vr1/biased/mixed_sample_ligands/ZINC00092235_0000.pdb
```

One way of doing this follows:

`touch ligand.list`

`foreach fn ($SLIDE_DATA_DIR/1vr1/biased/mixed_sample_ligands/*mol2)`
{system prompts something like "tell me more" and you reply} **`echo $fn >> ligand.list`**
{system prompts "tell me more" and you reply} **`end`**

2. Next a file containing the names and locations of all the rotated side chains will need to be generated:

`touch side_chain.list`

`foreach fn ($SLIDE_DATA_DIR/1vr1/biased/mixed_sample_targets/*pdb)`
{system prompts something like "tell me more" and you reply} **`echo $fn >> side_chain.list`**
{system prompts "tell me more" and you reply} **`end`**

It is crucial that your `ligand.list` and `side_chain.list` files show one-to-one correspondence in the molecule names (order as well as number) of docking entries. Check the files before proceeding.

3. Now run the slide_score command to generate the matchprint output (in comma separated value format):
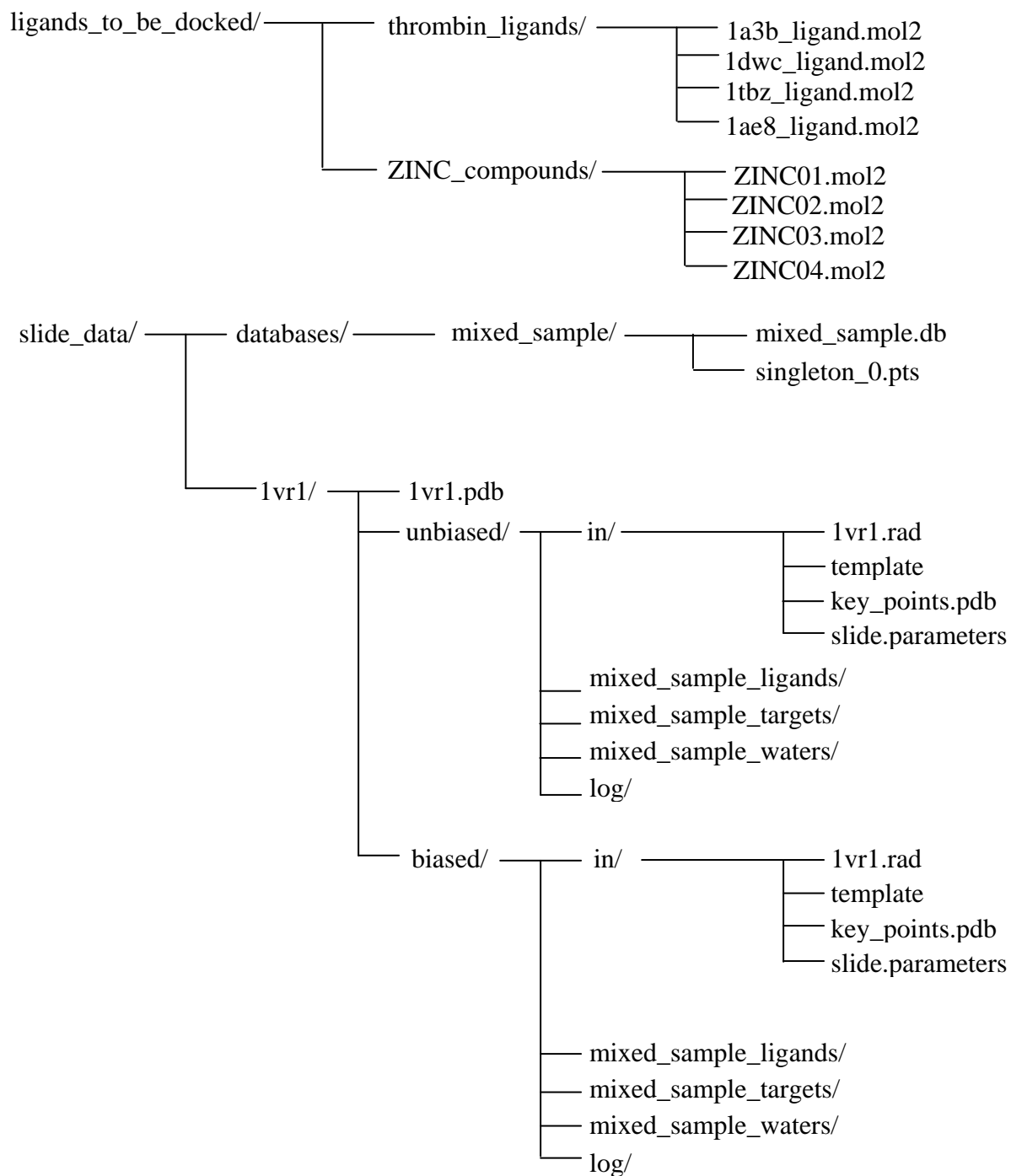
```
slide_score -p $SLIDE_DATA_DIR/1vr1/biased/in/1vr1.rad \
--lig_list ligand.list --sc_list side_chain.list --interactions_table \
>> 1vr1_biased_interactions_table.csv
```

Open this file using a spreadsheet program (i.e. Excel), and for the interaction table values on the bottom of the file, change the background shade color of all the cells with non-zero values to grey. (The method to do this varies with the spreadsheet program. In Excel it is a function called "Conditional Formatting" where you can select cells greater than 0 to be shaded). The result should look something like the following table:

	ligand	LEU 41(H)	CYS 42(H)	HIS 57(H)	HIS 57(H)	HIS 57(H)	TYR 60A(H)	TRP 60D(H)	LYS 60F(H)	LEU 99(H)	GLU 146(H)	GLU 146(H)	GLU 146(H)	ILE 174(H)	ASP
47	1vr1 interactions table:														
49	1a2c_ligar	0	0	0	0	0	0	0	0	0	1	1	1	2	
50	1a46_ligar	0	0	0	1	1	2	3	0	4	0	0	0	1	
51	1a4w_liga	0	0	0	0	1	1	3	1	4	0	0	0	1	
52	1a61_ligar	0	0	0	1	0	2	3	0	4	0	0	0	1	
53	1ad8_ligar	0	0	0	0	0	3	3	0	4	0	0	0	1	
54	1dwc_liga	0	0	0	0	1	6	4	0	1	0	0	0	3	
55	1tbz_ligar	0	0	0	0	1	3	3	0	4	0	0	0	1	
56	ZINC00000	0	0	0	0	0	4	5	0	1	0	0	0	0	
57	ZINC00000	0	0	0	0	0	2	4	0	4	0	0	0	2	
58	ZINC00000	1	1	0	2	0	0	1	2	0	0	0	0	0	
59	ZINC00000	0	1	0	0	0	1	4	1	0	0	0	0	0	
60	ZINC00000	0	0	0	0	0	0	4	0	0	0	0	0	0	
61	ZINC00000	0	1	0	2	0	0	7	2	0	0	0	0	0	
62	ZINC00000	0	0	0	0	0	0	0	0	0	0	0	0	0	
63	ZINC00000	0	1	1	1	0	0	3	0	0	0	0	0	0	
64	ZINC00000	0	0	0	0	1	6	3	0	1	0	0	0	0	
65	ZINC00000	0	0	0	1	0	0	3	0	0	0	0	0	0	
66	ZINC00000	0	0	0	0	0	0	3	0	2	0	0	0	0	
67	ZINC00092	0	0	0	0	0	2	3	0	1	0	0	0	0	
68	ZINC00171	0	0	0	0	0	3	3	0	1	0	0	0	0	
69	ZINC00538	0	0	0	1	0	4	5	0	2	0	0	0	0	

Matchprint interactions table for 1vr1 biased template screen of the mixed_sample database. Notice the pattern of highly conserved hydrophobic interactions with the TYR 60A, TRP 60D, and LEU 99 side chains.

Structure of the SLIDE data directory

```
ligands_to_be_docked/ ─────── thrombin_ligands/ ─────── 1a3b_ligand.mol2
                                                         1dwc_ligand.mol2
                                                         1tbz_ligand.mol2
                                                         1ae8_ligand.mol2

                              ZINC_compounds/ ─────── ZINC01.mol2
                                                      ZINC02.mol2
                                                      ZINC03.mol2
                                                      ZINC04.mol2

slide_data/ ─────── databases/ ─────── mixed_sample/ ─────── mixed_sample.db
                                                             singleton_0.pts

                    1vr1/ ─────── 1vr1.pdb
                                  unbiased/ ─────── in/ ─────── 1vr1.rad
                                                               template
                                                               key_points.pdb
                                                               slide.parameters

                                  mixed_sample_ligands/
                                  mixed_sample_targets/
                                  mixed_sample_waters/
                                  log/

                                  biased/ ─────── in/ ─────── 1vr1.rad
                                                             template
                                                             key_points.pdb
                                                             slide.parameters

                                  mixed_sample_ligands/
                                  mixed_sample_targets/
                                  mixed_sample_waters/
                                  log/
```

Preparing the protein target for docking with SLIDE

1. Obtain target PDB file. This is the file that contains the protein which will be used in the docking/screening run.

2. Remove any ANISOU records from the file

   `pdb_remove_anisou.pl <target>.pdb > <target>_noanisou.pdb`

3. Keep only one "Occupancy" occurrence of any multiple occupancy atoms in the PDB file by scanning through the file for any case where there are multiple lines per atom. There will generally be letters in column 17 indicating there are alternate locations for this atom. Delete all but the line with the highest occupancy (found in columns 55-60). For the line with the highest occupancy, replace the alternate location indicator (column 17) with a space [Warning: PDB format is column specific, so one must not simply delete the indicator, as this will shift columns].

4. Remove hydrogens

   `pdbdehydrogen <target>_noanisou.pdb > <target>_no_anisou_noH.pdb`

5. Remove ligands from binding site. Scan through the file and remove any ligands that occur in the binding site. (If these ligands will be used in the screen, then save the coordinates to a new file.) If co-factors exist that make important interactions, then it may be advisable to follow the instructions in the Quick Guide section 2.

Ligand database preparation

1. Gather ligands from PDB files, and convert to mol2 format, or download a database of small molecules (in mol2 format, or converted to mol2 format).

2. Make sure the name following the line @<TRIPOS>MOLECULE only contains alphanumeric characters and underscores '_', and no spaces. If other characters exist, use the script `mol2_format_name.pl` in `$SLIDE_DIR/bin` to remove everything after the first space character.

3. Add partial charges if your mol2 files do not already have them. We recommend using the program molcharge (OpenEye Scientific Software, Santa Fe, NM), with the am1bcc option. If the molecules contain atoms other than C, N, O, F, S, P, Cl, Br, I, Si, or H, then the am1 option will need to be used instead. SLIDE uses partial charges to assess salt bridges.

4. Generate low-energy conformers for each ligand candidate. We recommend using the program Omega (OpenEye Scientific Software, Santa Fe, NM) with the following parameters.

   a. Warts true (to set the conformer ID for each conformer). This is necessary for use with SLIDE.
   b. commentEnergy true   This gives Omega calculated energies for each conformer
   c. includeInput true (if you wish to include the input molecule, in the given conformation, in the run. If this is set to true, the input molecule will not be given a conformer ID, and one needs to be added afterwards using the script `mol2_add_conf_0.pl`.)
   [NOTE: Omega 2.1 can only handle atoms of the type C, N, O, F, S, P, Cl, Br, I, Si, or H]