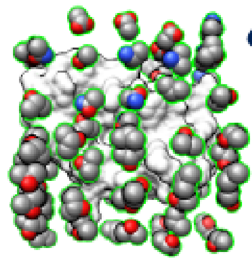


DruGUI

Druggability Suite for VMD

DruGUI Tutorial

Release 1.0

Ahmet Bakan

December 24, 2013

CONTENTS

1 Introduction **2**
- 1.1 Installation . . . 2
- 1.2 DruGUI Plugin . . . 2
- 1.3 Tutorial Files . . . 3
- 1.4 How to Cite . . . 4

2 Background **5**
- 2.1 Method & Theory . . . 5

3 Probe Molecules **6**
- 3.1 Topology & Parameter . . . 7

4 Simulation Setup **13**
- 4.1 Input Files . . . 13
- 4.2 Options & Parameters . . . 14
- 4.3 Output Files . . . 14
- 4.4 Simulation . . . 15

5 Probe Grid Calculation **16**
- 5.1 Input Files . . . 17
- 5.2 Options & Parameters . . . 17
- 5.3 Output Files . . . 17

6 Druggability Analysis **18**
- 6.1 Input Files . . . 18
- 6.2 Options & Parameters . . . 18
- 6.3 Output Files . . . 19

7 Analyze a Specific Site **20**
- 7.1 Input Files . . . 20
- 7.2 Options & Parameters . . . 20
- 7.3 Output Files . . . 21

8 Visualization & Analysis **22**
- 8.1 Input Files . . . 22
- 8.2 Options & Parameters . . . 22
- 8.3 Probe binding spots . . . 22
- 8.4 Druggable Sites . . . 26
- 8.5 Visualize Probe Grids . . . 26

9 Incorporate CGenFF Molecules 29

10 License 32

Bibliography 33

Learn how to use Druggability Suite for setting up and analyzing druggability simulations.

CHAPTER

ONE

INTRODUCTION

Druggability Suite is a **'VMD'_** plugin GUI and a Python module developed for setup and analysis of simulations described in [AB12].

1.1 Installation

1. **VMD_** 1.9.1 or later is required for using GUI. **NAMD_** is required for running druggability simulations. Following are required for performing druggability analysis calculations:

- **Python_** 2.7
- **NumPy_** 1.3 or later

2. Download one of the following archive files:

- drugui_plugin_files.tgz
- drugui_plugin_files.zip

3. Extract contents of the archive and copy `drugui` folder to VMD TCL plugins directory, i.e. `$VMDDIR/plugins/noarch/tcl/`.

   Then, insert following line into `$VMDDIR/scripts/vmd/loadplugins.tcl` at line 186:

```
vmd_install_extension drugui drugui_tk "Modeling/DruGUI"
```

   If you are not sure where VMD directory is located, run **vmd**, and type the following command line in the VMD console:

```
global env; puts $env(VMDDIR)
```

1.2 DruGUI Plugin

Druggability Suite GUI (DruGUI) plugin, shown below, has five panels to streamline setup, analysis, and visualization of druggability simulations:

- *Simulation Setup*
- *Probe Grid Calculation*
- *Druggability Analysis*
- *Analyze a Specific Site*

- *Visualization & Analysis*

The rest of the tutorial will show you how to use these panels, and described required inputs and outputs from different analysis steps.

1.3 Tutorial Files

Files in the following archives can be used to follow this tutorial:

- DruGUI Tutorial Files (TGZ)
- DruGUI Tutorial Files (ZIP)

Here is a list of these files:

```
112K  mdm2.pdb
335K  mdm2.psf
3.9K  sample_1t4e_inhibitor.pdb
2.2M  sample_ACAM.dx
2.2M  sample_ACET.dx
4.5K  sample_all_hotspots.pdb
 55K  sample_heavyatoms.pdb
2.2M  sample_IPAM.dx
2.4M  sample_IPRO.dx
5.7K  sample.log
1.3K  sample_site_1.pdb
```

```
553  sample_site_1_soln_1.pdb
553  sample_site_1_soln_2.pdb
553  sample_site_1_soln_3.pdb
```

1.4 How to Cite

If you benefited from Druggability Suite in your research, please cite the following paper:

CHAPTER

TWO

BACKGROUND

Druggability of a target protein is an important question in drug discovery. A reliable and physically relevant measure of druggability can help determining risks associated with pursuing a given target protein. Furthermore, a comprehensive analysis of druggability of a target can help identifying novel sites and alternate inhibition mechanisms.

To this extent, experimental NMR and X-ray and computational screening methods are utilized for druggable binding site identification. Protein pockets that bind a wide range of fragments or organic solvent molecules in these screening experiments usually coincide with known druggable sites.

2.1 Method & Theory

Based on these ideas, we developed a unbiased simulation based approach to assess druggability of target proteins with known structures. Our approach involves simulations of the target in presence of a set of small organic molecules (probes) with diverse physiochemical properties. Probes are selected to be small (four non-hydrogen atoms) so that they can diffuse fast and explore small and even transient pockets in simulations. This property also helps sampling large number of binding events and enables reaching equilibrium in relatively short simulation times.

Simulation trajectories are analyzed to calculate probe enrichment on protein surface and pockets using a grid based approach. Enrichment grids are converted to probe binding affinities using inverse Boltzmann relation. Binding sites are identified by locating clusters of high affinity probe binding spots. Druggability index for a binding site is calculated by considering the affinities of seven or eight probe molecules (28 to 32 non-hydrogen atoms), which is equivalent to a drug-like molecule in size.

Details of the approach can be found in [AB12]. We showed that probes mimic interactions of drug-like molecules, as well as substrates and inhibitors that are not necessarily drug like. Thus, the approach is suitable for assessing druggability or ligandability of a protein.

CHAPTER

THREE

PROBE MOLECULES

Probe molecules and their fractional composition, shown below, were selected based on analysis of FDA approved drugs [AB12]. The best composition for a specific system may, however, depend on the surface properties of the target. For a protein with a highly charged surface, increasing the proportion of acetate and isopropylamine might perform better in identifying ligandable sites.

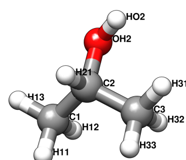


Figure 3.1: Isopropanol (70%)

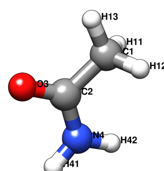


Figure 3.2: Acetamide (10%)

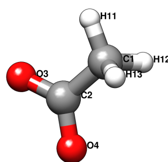


Figure 3.3: Acetate (10%)

Among these, isobutane is not included in the default configuration due to its high propensity to aggregate. If it is included, it should have a small fraction.

Figure 3.4: Isopropylamine (10%)

Figure 3.5: Isobutane (0%)

3.1 Topology & Parameter

Probe atom types and parameters are based on protein side-chains described in CHARMM force field. Below topology and parameter files distributed with the plugin are shown:

```
*  ---------------------------------------------------------------------------  *
*          Topology file for probes used in druggability calculations           *
*                              31 December, 2009                                *
*  ---------------------------------------------------------------------------  *
*
36  1

!  ---------------------------------------------------------------------------  !
!           ATOM NAME     SERIAL NO. SLOTS                                      !
!                         RESERVED FOR THE ATOMS                                !
!           Hydrogens        1 --  39                                           !
!           Carbons         40 --  99                                           !
!           Nitrogens      100 -- 149                                           !
!           Oxygens        150 -- 189                                           !
!           Sulphurs       190 -- 209                                           !
!           Halogens       210 -- 249                                           !
!           Miscellaneous  250 -- ---                                           !
!  ---------------------------------------------------------------------------  !

!hydrogens
MASS     1 H      1.00800 H ! polar H
MASS     2 HC     1.00800 H ! N-ter H
MASS     3 HA     1.00800 H ! nonpolar H
MASS           4 HT             1.00800 H ! TIPS3P WATER HYDROGEN
MASS    32 CC    12.01100 C ! carbonyl C, asn,asp,gln,glu,cter,ct2
MASS    55 NH2   14.00700 N ! amide nitrogen
MASS    56 NH3   14.00700 N ! ammonium nitrogen
MASS    70 O     15.99900 O ! carbonyl oxygen
MASS    72 OC    15.99900 O ! carboxylate oxygen
MASS          75 OT            15.99940 O ! TIPS3P WATER OXYGEN
MASS    22 CT1   12.01100 C ! aliphatic sp3 C for CH
MASS    24 CT3   12.01100 C ! aliphatic sp3 C for CH3
```

```
MASS    73 OH1   15.99900 O ! hydroxyl oxygen

DEFA FIRS NONE LAST NONE
AUTO ANGLES DIHE

RESI TIP3              0.000 ! tip3p water model, generate using noangle nodihedral
GROUP
ATOM OH2  OT         -0.834
ATOM H1          HT          0.417
ATOM H2          HT          0.417
BOND OH2 H1 OH2 H2 !H1 H2  ! the last bond is needed for shake
!ANGLE H1 OH2 H2           ! required
ACCEPTOR OH2
PATCHING FIRS NONE LAST NONE

RESI IPRO         0.000 ! ATOM TYPES FROM THR
GROUP
ATOM C2   CT1      0.181 !  H12  H13  H33 H32
ATOM H21  HA       0.049 !     \ /      \ /
GROUP                    ! H11--C1      C3--H31
ATOM C1   CT3    -0.147 !        \    /
ATOM H11  HA       0.049 !          C2
ATOM H12  HA       0.049 !        /    \
ATOM H13  HA       0.049 !      OH2    H21
GROUP                    !       |
ATOM C3   CT3    -0.147 !      HO2
ATOM H31  HA       0.049
ATOM H32  HA       0.049
ATOM H33  HA       0.049
GROUP
ATOM OH2  OH1    -0.660
ATOM HO2  H       0.430
BOND C2  C1   C2  C3   C2  OH2
BOND C1  H11  C1  H12  C1  H13
BOND C2  H21
BOND C3  H31  C3  H32  C3  H33
BOND OH2 HO2
DONOR HO2 OH2
ACCEPTOR OH2

RESI IBUT         0.000 ! ISOBUTANE
GROUP
ATOM C2   CT1    -0.049 !  H12  H13  H33 H32
ATOM H21  HA       0.049 !     \ /      \ /
GROUP                    ! H11--C1      C3--H31
ATOM C1   CT3    -0.147 !        \    /
ATOM H11  HA       0.049 !          C2--H21
ATOM H12  HA       0.049 !        /
ATOM H13  HA       0.049 ! H41--C4--H43
GROUP                    !       |
ATOM C3   CT3    -0.147 !      H42
ATOM H31  HA       0.049
ATOM H32  HA       0.049
ATOM H33  HA       0.049
GROUP
ATOM C4   CT3    -0.147
ATOM H41  HA       0.049
ATOM H42  HA       0.049
```

```
ATOM H43  HA      0.049
BOND C2  C1   C2  C3   C2  C4
BOND C1  H11  C1  H12  C1  H13
BOND C2  H21
BOND C3  H31  C3  H32  C3  H33
BOND C4  H41  C4  H42  C4  H43
IC C2  H13 *C1 H11  1.5472 117.4600  120.9800 107.1700  1.1145
IC C2  H13 *C1 H12  1.5472 117.4600 -124.6700 108.9800  1.1126
IC H13 C1   C2 C3   1.5543 117.4600  180.0000 110.4800  1.5361
IC C3  C1  *C2 C4   1.5361 110.4800  120.0000 112.5700  1.5360
IC C3  C4  *C2 H21  1.5361 110.2600  120.0000 108.0200  1.1168
IC C1  C2   C3 H31  1.5472 110.4800  177.3300 110.5400  1.1111
IC H31 C2  *C3 H32  1.1111 110.5400  119.9600 110.6200  1.1112
IC H31 C2  *C3 H33  1.1111 110.5400 -119.8500 110.6900  1.1108
IC C1  C2   C4 H41  1.5472 112.5700  178.9600 110.3200  1.1116
IC H41 C2  *C4 H42  1.1116 110.3200  119.7100 111.6900  1.1086
IC H41 C2  *C4 H43  1.1116 110.3200 -119.6100 110.4900  1.1115

RESI IPAM         1.000 ! ISOPROPYLAMINE, ATOM TYPES FROM LYS
GROUP
ATOM C2   CT1     0.252 !  H12  H13  H33 H32
ATOM H21  HA      0.049 !     \ /      \ /
GROUP                   ! H11--C1      C3--H31
ATOM C1   CT3    -0.147 !        \    /
ATOM H11  HA      0.049 !          C2--H21
ATOM H12  HA      0.049 !        /
ATOM H13  HA      0.049 ! H41--N4--H43
GROUP                   !       |
ATOM C3   CT3    -0.147 !      H42
ATOM H31  HA      0.049
ATOM H32  HA      0.049
ATOM H33  HA      0.049
GROUP
ATOM N4   NH3    -0.300
ATOM H41  HC      0.333
ATOM H42  HC      0.333
ATOM H43  HC      0.333
BOND C2  C1   C2  C3   C2  N4
BOND C1  H11  C1  H12  C1  H13
BOND C2  H21
BOND C3  H31  C3  H32  C3  H33
BOND N4  H41  N4  H42  N4  H43
IC C2  H13 *C1 H11  1.5472 117.4600  120.9800 107.1700  1.1145
IC C2  H13 *C1 H12  1.5472 117.4600 -124.6700 108.9800  1.1126
IC H13 C1   C2 C3   1.5543 117.4600  180.0000 110.4800  1.5361
IC C3  C1  *C2 N4   1.5361 110.4800  120.0000 112.5700  1.5360
IC C3  N4  *C2 H21  1.5361 110.2600  120.0000 108.0200  1.1168
IC C1  C2   C3 H31  1.5472 110.4800  177.3300 110.5400  1.1111
IC H31 C2  *C3 H32  1.1111 110.5400  119.9600 110.6200  1.1112
IC H31 C2  *C3 H33  1.1111 110.5400 -119.8500 110.6900  1.1108
IC C1  C2   N4 H41  1.5472 112.5700  178.9600 110.3200  1.1116
IC H41 C2  *N4 H42  1.1116 110.3200  119.7100 111.6900  1.1086
IC H41 C2  *N4 H43  1.1116 110.3200 -119.6100 110.4900  1.1115

RESI ACET        -1.000 ! ACETATE, ATOM TYPES FROM GLU
GROUP
ATOM C2   CC      0.520 !     H11    O4
ATOM O3   OC     -0.760 !      |    //
```

```
ATOM O4   OC     -0.760 ! H12--C1--C2
GROUP                   !      |    \
ATOM C1   CT3    -0.147 !     H13    O3 (-)
ATOM H11  HA      0.049
ATOM H12  HA      0.049
ATOM H13  HA      0.049
BOND C2  C1   C2  O3   C2  O4
BOND C1  H11  C1  H12  C1  H13
IMPR C2  C1  O3  O4
IC C2  H13 *CB  H11  1.5218 112.6000  119.2200 109.2300  1.1086
IC C2  H13 *CB  H12  1.5218 112.6000 -121.6100 110.6400  1.1080
IC H13 CB   C2  O4   1.5619 112.6000  180.0000 117.9900  1.2565
IC O4  CB  *C2  O3   1.2565 117.9900 -170.2300 117.7000  1.2541

RESI ACAM         0.000 ! ACETAMIDE, ATOM TYPES FROM GLN
GROUP
ATOM C2   CC      0.550 !     H11  O3    H41
ATOM O3   O      -0.550 !      |   ||   /
GROUP                   ! H12--C1--C2--N4
ATOM C1   CT3    -0.147 !      |        \
ATOM H11  HA      0.049 !     H13        H42
ATOM H12  HA      0.049
ATOM H13  HA      0.049
GROUP
ATOM N4   NH2    -0.620
ATOM H41  H       0.320
ATOM H42  H       0.300
BOND C2  C1   C2  O3   C2  N4
BOND C1  H11  C1  H12  C1  H13
BOND N4  H41  N4  H42
IMPR C2  N4  C1   O3    C2  C1  N4   O3
IMPR N4  C2  H41  H42   N4  C2  H42  H41
DONOR H41 N4
DONOR H42 N4
ACCEPTOR O3 C2
IC C2  H13 *CB H11  1.5319 114.3000  119.1700 107.8200  1.1120
IC C2  H13 *CB H12  1.5319 114.3000 -123.7400 110.3400  1.1091
IC H13 CB   C2 O3   1.5627 114.3000  180.0000 122.5600  1.2323
IC O3  CB  *C2 N4   1.2323 122.5600 -179.1900 116.1500  1.3521
IC CB  C2   N4 H41  1.5319 116.1500 -179.2600 117.3500  0.9963
IC H41 C2  *N4 H42  0.9963 117.3500  178.0200 120.0500  0.9951

END

*>>>>>>>>>>  Parameter file for druggability calculations  <<<<<<<<<<<<<
*>>>>>>>>>>>>>>>>>>>>>>>>> 31 December, 2009 <<<<<<<<<<<<<<<<<<<<<<<<<<<
*

! references
!
!PROTEINS
!
!MacKerell, A.D., Jr,. Feig, M., Brooks, C.L., III, Extending the
!treatment of backbone energetics in protein force fields: limitations
!of gas-phase quantum mechanics in reproducing protein conformational
!distributions in molecular dynamics simulations, Journal of
!Computational Chemistry, 25: 1400-1415, 2004.
!
```

```
!MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R.L.;
!Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.;
!Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos,
!C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, III,
!W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.;
!Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M.  All-atom
!empirical potential for molecular modeling and dynamics Studies of
!proteins.  Journal of Physical Chemistry B, 1998, 102, 3586-3616.
!
!

BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!

ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
!atom types     Ktheta    Theta0   Kub     S0
!

NH3  CT1  HA    45.000    107.50   35.00   2.10100 ! ALLOW   ALI POL
                ! new stretch and bend; methylammonium (KK 03/10/92)
                ! NH3  CT2  HA

DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types             Kchi    n   delta
!
!Neutral N terminus

IMPROPER
!
!V(improper) = Kpsi(psi - psi0)**2
!
!Kpsi: kcal/mole/rad**2
!psi0: degrees
!note that the second column of numbers (0) is ignored
```

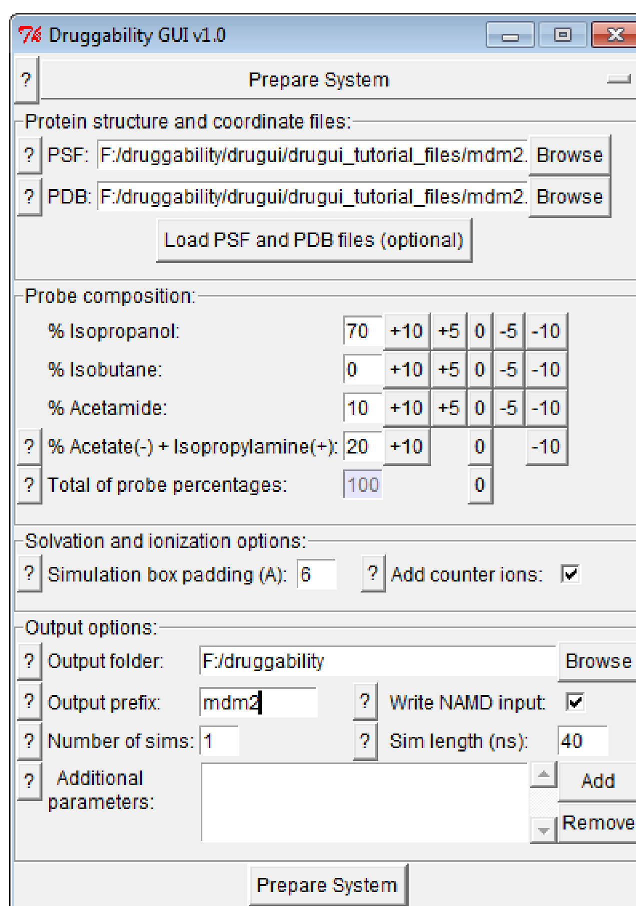
```
!
!atom types           Kpsi                    psi0
!

END
```

CHAPTER

FOUR

SIMULATION SETUP

System that contains target, probes, water, and counter ions for druggability simulations can be prepared using the following interface:



4.1 Input Files

`.psf` and `.pdb` files for the target protein are required from the user. You can learn how to prepare these files from NAMD tutorials.

4.2 Options & Parameters

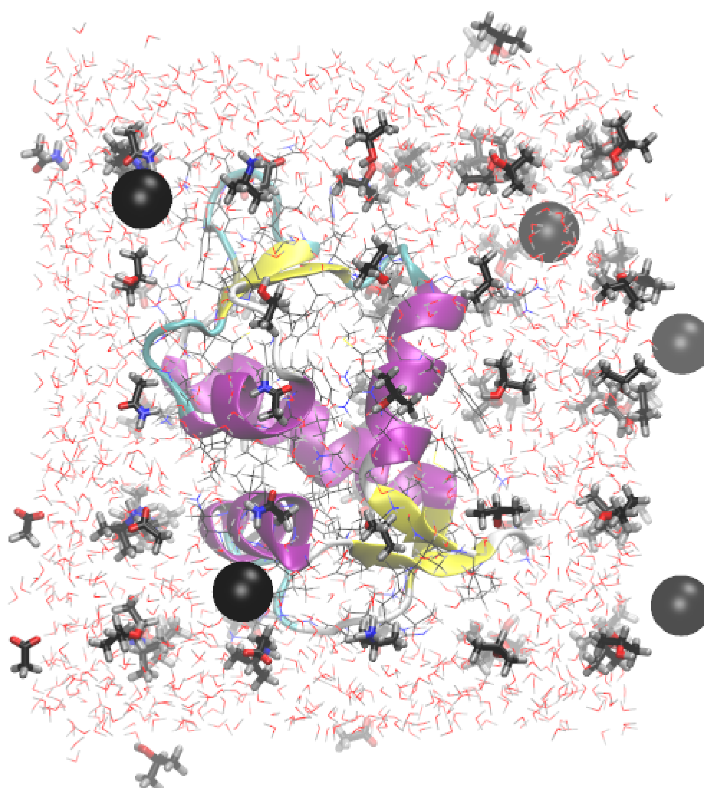
1. First, select `.psf` (structure) and `.pdb` (coordinate) files. You can use the MDM2 files provided for this tutorial. Alternatively, structure and coordinate files for a protein of interest can also be used. These files should contain all atoms required for protein stability and function, these may include cofactors and metal atoms. Crystallographic water molecules may also be retained in the
2. Select a probe composition that will complement the surface properties of the target protein. Note that, probe percentages must sum up to 100. Acetate and isopropylamine percentages will be equal to each other so that opposite probe charges are balanced.

   Note: It is possible to set all probe percentages to zero. The final system will not contain any probes. This may be used as a control simulation, to see how the target protein behaves in the absence of probes.

3. Enter simulation box padding (distance from the protein to box surfaces) and select whether you would like to add counter ions. Druggability GUI uses Solvate and Autoionize plugins to add water, probes, and ions.
4. Select output folder, file prefix, and number of simulations that you want to perform. Performing multiple simulations to see whether results are reproducible is always a good idea.

4.3 Output Files

Output will be `prefix.psf` and `prefix.pdb` files of the system that contain target protein, water, ions, and probe molecules (if selected). You system should look like the following:



In addition, you will see `prefix_min`, `prefix_sim`, `parameters` folders that contain input for molecular dynamics simulations.

For a summary of contents of the final system, see `prefix.log` file.

4.4 Simulation

Now you need to run druggability simulations. See `prefix.sh` file for NAMD commands that you need to execute. When simulations are complete, you can continue with following analysis steps.

CHAPTER

FIVE

PROBE GRID CALCULATION

When simulations are complete, you need to perform grid calculations using the following interface:

5.1 Input Files

Input files are `prefix.psf` and `prefix.pdb` files, and simulation trajectory files, e.g. `prefix_sim/sim.dcd`. If you performed multiple simulations for the same system, you can include all productive simulation trajectories in grid calculations. Ideally, trajectories from equilibration simulations should not be included in grid calculations. *Selection* specifies atoms used to align the target conformations in trajectory frames.

Note: Selection is an important input for grid calculations. If the binding site move internally when all atoms of the protein are used for alignment, you may want to restrict the alignment to binding site residues excluding those that are mobile. This will help capturing probe enrichment at a binding site properly. If there are multiple binding sites that move internally in a protein, it would be better to analyze those sites one by one.

5.2 Options & Parameters

1. For probe enrichment (or occupancy) grid calculations, probe and solvent molecules needs to be wrapped. If you like and have enough disk space, you may write trajectory frames after they are wrapped. This may be good for later visualization and movie making purposes.
2. By default, grids will be calculated for different probe types using their central carbon atoms. These grids will be merged in druggability analysis. Default grid resolution 0.5 Å has been found to capture probe locations ideally.

   For visualization purposes, you may select to output occupancy grids for hydrophobic, polar, charged, and water atoms.

   Grids can be visualized using Chimera Volume Viewer or VMD.
3. You may also select to evaluate grids immediately after their calculation. Details of this step is discussed in the next part.

5.3 Output Files

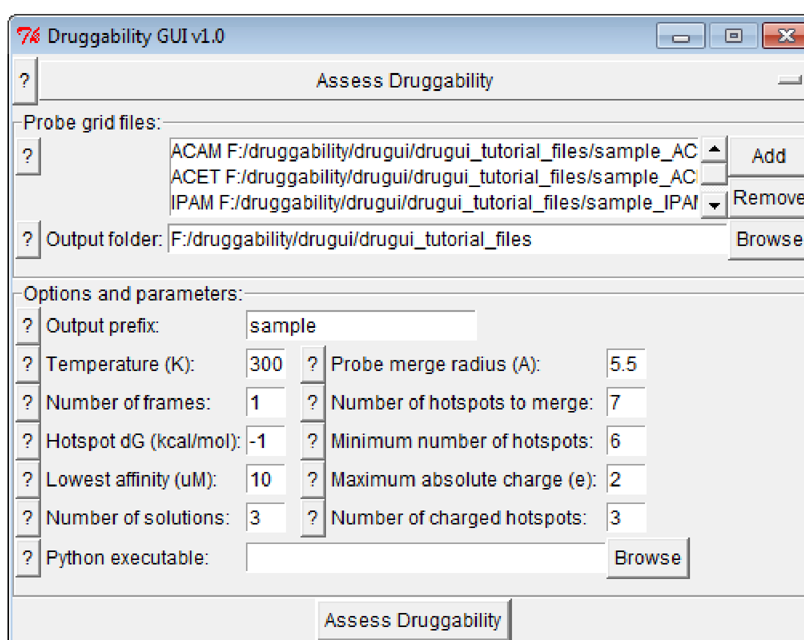
Output files are occupancy grids for each probe type, e.g. `prefix_IPRO.dx`, and selected atom types.

CHAPTER

SIX

DRUGGABILITY ANALYSIS

Druggability index based on probe occupancy grids can be calculated using the following interface:



6.1 Input Files

Input files are probe occupancy grids calculated in the previous step, e.g. `prefix_IPRO.dx`, `prefix_ICAM.dx`, etc.

6.2 Options & Parameters

1. This step involves selecting high affinity probe binding spots, clustering them, and then merging to assess druggability. Depending on parameter choice you may see different number of binding sites and some variation in their druggability. This will usually effect weakly druggable site. Ideal parameters that worked best for a set of diverse proteins are set as defaults and their detailed discussions can be found in [AB12].

2. GUI will try to locate Python executable path, but if you do not see an entry, you will need to specify it manually. Results will be visualized immediately, in high resolution if selected so.

6.3 Output Files

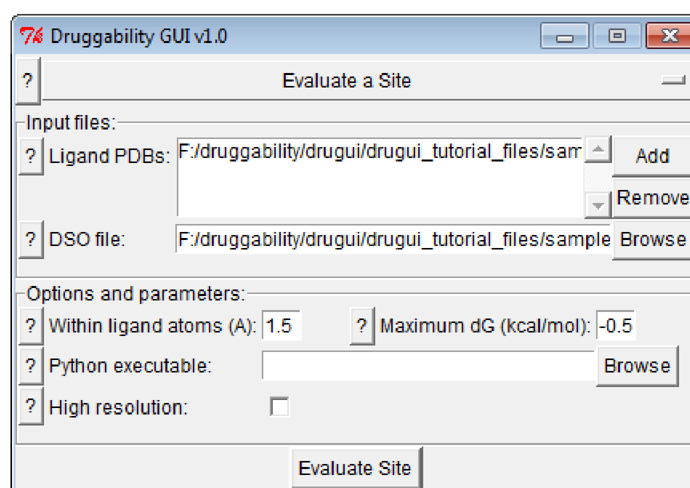
Output from this step is a set of PDB files written into `prefix` folder and a `.dso` file that contains Python objects containing probe grids.

CHAPTER

SEVEN

ANALYZE A SPECIFIC SITE

This interface can be used to select probe binding hotspots that overlap with a given ligand.



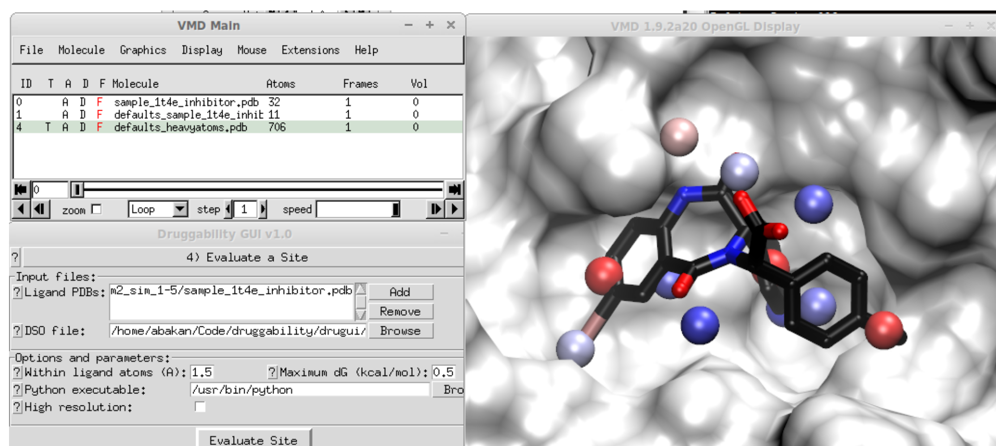
7.1 Input Files

Input files area list of ligand `.pdb` files and `.dso` file from previous step. Ligand-bound form of the protein should be aligned to the simulated structure. You can use `prefix_heavyatoms.pdb` file for alignment.

7.2 Options & Parameters

1. Parameters determined how many probe binding spots are incorporated in assessment of the specified site. Selected probe should be close to the ligand, so 1.5 Å should be sufficient. If the ligand is bound to a weak site, maximum free energy may be lower that that is used in the previous step, such as -0.5 kcal/mol.
2. GUI will try to locate Python executable path, but if you do not see an entry, you will need to specify it manually.

When you use the tutorial files for MDM2 inhibitor, you should get a representation similar to the following:



In the figure, 11 probe binding spots that overlap with MDM2 inhibitor is selected. Sum of their binding free energies looks reasonable (this will be displayed in the logfile viewer). For a large ligand, however, you may end up with a large selection of probe binding spots and sum of their binding free energies may result in a very high affinity. If this is the case, you should disregard the total. The approach (merging probe binding spots and adding their binding free energies) works well for drug-size molecules.

7.3 Output Files

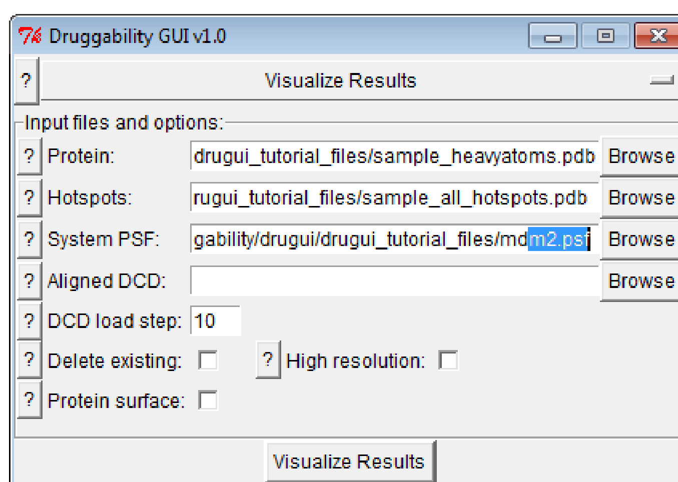
Output from this step is a set of PDB files written into `prefix` folder. Binding free energies of selected probe binding spots will be appended to the log file in this folder.

CHAPTER

EIGHT

VISUALIZATION & ANALYSIS

Following interface can be used to generate a quick visualization of results from druggability analysis.



8.1 Input Files

Input files are `prefix_heavyatoms.pdb` and other PDB files in `prefix` folder.

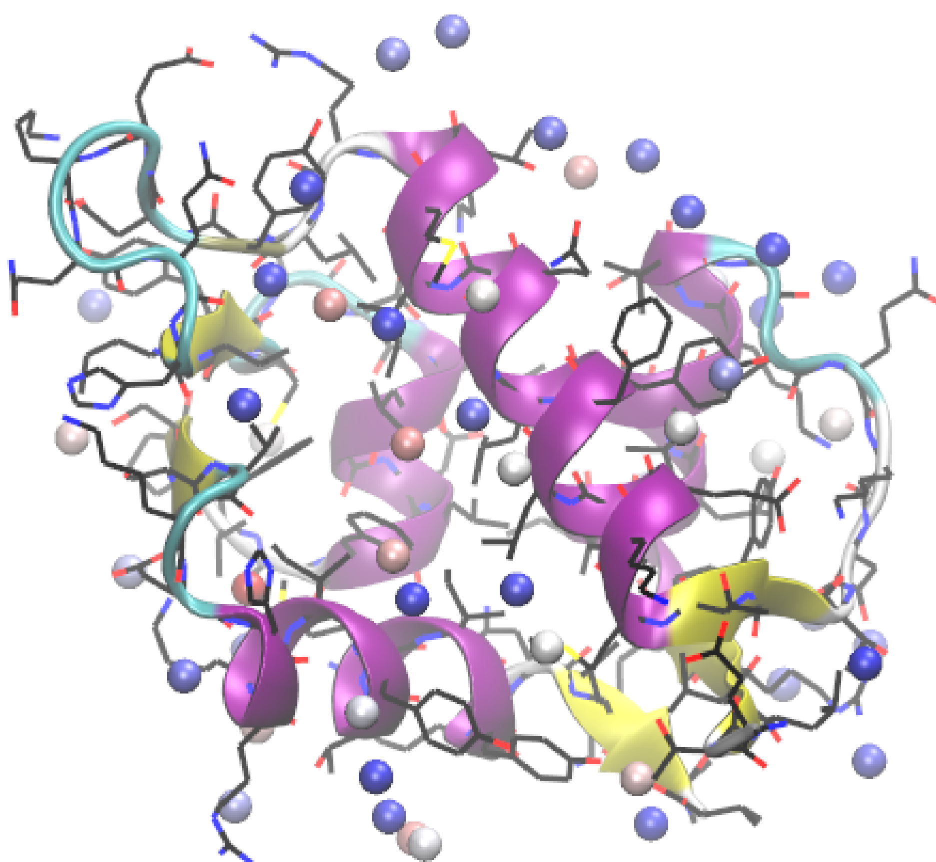
8.2 Options & Parameters

1. If you have outputted aligned trajectory in grid calculation step, you can select to load it too.
2. Optionally, molecules present in VMD can be deleted, high resolution representations and a protein surface representation can be generated.

8.3 Probe binding spots

When results are loaded, you will see a representation similar to the following:

Each sphere corresponds to a probe binding spot. Spheres are colored according to their binding free energies. Red most sphere has the lowers binding free energy.



Binding free energies of probes can be found in the logfile:

```
defaults is initialized.
defaults working directory is set to "defaults".
Druggability Analysis defaults is initialized.
Parameter: temperature 300.00 K
Parameter: delta_g -1.000 kcal/mol
Parameter: n_probes 7
Parameter: min_n_probes 6
Parameter: merge_radius 5.5 A
Parameter: low_affinity 10.00 uM
Parameter: n_solutions 3
Parameter: max_charge 2.0 e
Parameter: n_charged 3
Parameter: n_frames 1
Parsing OpenDX file defaults_IPRO.dx.
defaults_IPRO was parsed in 0.45s
Parsing OpenDX file defaults_IPAM.dx.
defaults_IPAM was parsed in 0.45s
Parsing OpenDX file defaults_ACAM.dx.
defaults_ACAM was parsed in 0.45s
Parsing OpenDX file defaults_ACET.dx.
defaults_ACET was parsed in 0.45s
Searching probe binding hotspots with deltaG less than -1.00 kcal/mol (~5 folds enrichment).
51 all-probes binding spots were identified in 1.07s.
Minimum binding free energy is -2.73 kcal/mol.
Hotspot   1 -2.73 kcal/mol 100.0% IPRO
Hotspot   2 -2.36 kcal/mol 100.0% IPRO
Hotspot   3 -2.34 kcal/mol  95.3% ACET   2.5% IPRO   2.2% ACAM
Hotspot   4 -2.33 kcal/mol 100.0% IPRO
Hotspot   5 -2.28 kcal/mol  93.4% IPRO   3.8% ACAM   2.8% IPAM
Hotspot   6 -2.23 kcal/mol  98.8% IPRO   1.2% ACAM
Hotspot   7 -2.19 kcal/mol  94.6% IPRO   2.9% ACAM   2.5% ACET
Hotspot   8 -2.13 kcal/mol  99.5% IPRO   0.3% ACET   0.2% ACAM
Hotspot   9 -2.13 kcal/mol  96.9% IPRO   3.1% ACAM
Hotspot  10 -2.02 kcal/mol  84.5% IPRO  15.5% ACAM
Hotspot  11 -2.00 kcal/mol  83.9% IPAM  12.7% IPRO   3.4% ACAM
Hotspot  12 -1.98 kcal/mol  93.2% IPRO   4.7% ACAM   2.1% IPAM
Hotspot  13 -1.91 kcal/mol  97.2% IPRO   2.8% ACAM
Hotspot  14 -1.89 kcal/mol  80.7% IPRO  18.0% ACET   1.3% ACAM
Hotspot  15 -1.85 kcal/mol  94.5% IPRO   4.7% IPAM   0.7% ACET   0.2% ACAM
Hotspot  16 -1.79 kcal/mol  82.5% IPRO  10.2% ACAM   5.6% ACET   1.6% IPAM
Hotspot  17 -1.75 kcal/mol  97.6% IPRO   2.4% ACAM
Hotspot  18 -1.72 kcal/mol  98.9% IPRO   0.6% ACAM   0.5% IPAM
Hotspot  19 -1.70 kcal/mol  46.3% ACET  45.4% IPRO   5.4% ACAM   2.9% IPAM
Hotspot  20 -1.56 kcal/mol  63.6% IPRO  22.7% ACET  11.7% ACAM   2.0% IPAM
Hotspot  21 -1.53 kcal/mol  79.7% ACET  20.3% IPRO
Hotspot  22 -1.44 kcal/mol  95.7% IPRO   4.3% ACAM
Hotspot  23 -1.40 kcal/mol  87.4% IPRO  11.0% IPAM   1.6% ACAM
Hotspot  24 -1.36 kcal/mol  50.2% ACET  49.8% IPRO
Hotspot  25 -1.30 kcal/mol  80.4% ACET  18.9% IPRO   0.7% ACAM
Hotspot  26 -1.30 kcal/mol  72.1% IPRO  16.0% ACET   6.6% IPAM   5.4% ACAM
Hotspot  27 -1.27 kcal/mol  54.5% IPRO  24.3% ACET  21.3% ACAM
Hotspot  28 -1.26 kcal/mol  88.4% ACET   9.0% IPRO   2.5% ACAM
Hotspot  29 -1.26 kcal/mol  78.4% ACET  18.6% IPRO   3.0% ACAM
Hotspot  30 -1.25 kcal/mol  89.8% IPRO   8.7% ACAM   1.5% IPAM
Hotspot  31 -1.24 kcal/mol  97.7% IPRO   2.3% ACAM
Hotspot  32 -1.23 kcal/mol  83.9% IPRO  16.1% ACAM
Hotspot  33 -1.23 kcal/mol  85.7% IPAM  13.5% IPRO   0.8% ACAM
```

```
Hotspot  34 -1.21 kcal/mol  94.0% IPRO   5.8% ACAM   0.3% IPAM
Hotspot  35 -1.20 kcal/mol  90.2% IPRO   6.4% IPAM   3.4% ACAM
Hotspot  36 -1.15 kcal/mol  90.3% IPRO   8.8% ACAM   0.9% ACET
Hotspot  37 -1.14 kcal/mol  88.9% IPRO  10.2% ACET   0.9% ACAM
Hotspot  38 -1.13 kcal/mol  94.7% IPRO   5.3% ACAM
Hotspot  39 -1.11 kcal/mol  99.0% ACET   0.6% IPRO   0.3% ACAM
Hotspot  40 -1.10 kcal/mol  97.4% IPRO   2.3% ACAM   0.3% IPAM
Hotspot  41 -1.09 kcal/mol  85.9% IPRO   8.7% ACAM   5.0% IPAM   0.3% ACET
Hotspot  42 -1.08 kcal/mol 100.0% IPRO
Hotspot  43 -1.08 kcal/mol 100.0% IPRO
Hotspot  44 -1.08 kcal/mol  96.6% IPRO   3.4% ACAM
Hotspot  45 -1.08 kcal/mol  55.8% IPAM  33.6% IPRO  10.6% ACAM
Hotspot  46 -1.06 kcal/mol  97.9% IPRO   2.1% ACAM
Hotspot  47 -1.06 kcal/mol  86.0% ACET  11.6% IPRO   2.1% ACAM   0.4% IPAM
Hotspot  48 -1.05 kcal/mol  96.8% IPRO   3.2% ACAM
Hotspot  49 -1.03 kcal/mol 100.0% IPRO
Hotspot  50 -1.01 kcal/mol  86.7% IPRO  11.8% IPAM   1.5% ACAM
Hotspot  51 -1.01 kcal/mol  85.6% IPRO   8.4% ACET   6.1% ACAM
IPRO: 39 isopropanol binding hotspots were identified.
IPRO: lowest binding free energy is -2.73 kcal/mol.
IPAM: 3 isopropylamine binding hotspots were identified.
IPAM: lowest binding free energy is -2.00 kcal/mol.
ACAM: 0 acetamide binding hotspots were identified.
ACET: 9 acetate binding hotspots were identified.
ACET: lowest binding free energy is -2.34 kcal/mol.
Clustering probe binding hotspots.
Clustering completed in 2.64ms.
1 potential sites are identified.
Calculating achievable affinity ranges.
Site 1: 16 probe binding hotspots
Site 1: Lowest probe binding free energy -2.36 kcal/mol
Site 1: Average probe binding free energy-1.56 kcal/mol
Site 1: Total of 70 solutions.

Achievable affinities for site 1
-log10(affinity)
     #-------------#
9.53 |o            |
9.28 |-o           |
9.03 |--------o    |
8.79 |----------o  |
8.54 |-----------o |
8.29 |---------o   |
8.05 |--------o    |
7.80 |-----o       |
7.56 |----o        |
7.31 |----o        |
     #-------------#
     0    5   10
Site 1: Lowest drug-like binding free energy -13.07 kcal/mol
Site 1: Highest drug-like affinity 0.298 nM
Site 1: Solution 1 binding free energy -13.07 kcal/mol
Site 1: Solution 1 affinity 0.298 nM
Site 1: Solution 1 total charge 0.02 e
Site 1: Solution 1 number of hotspots 7
Site 1: Solution 1 approximate volume 450.58 A^3
Site 1: Solution 2 binding free energy -12.66 kcal/mol
Site 1: Solution 2 affinity 0.593 nM
```

```
Site 1: Solution 2 total charge -0.03 e
Site 1: Solution 2 number of hotspots 7
Site 1: Solution 2 approximate volume 449.28 A^3
Site 1: Solution 3 binding free energy -12.49 kcal/mol
Site 1: Solution 3 affinity 0.780 nM
Site 1: Solution 3 total charge 0.03 e
Site 1: Solution 3 number of hotspots 7
Site 1: Solution 3 approximate volume 451.70 A^3
Hotspots are written into file defaults/defaults_all_hotspots.pdb.
defaults is cPickled into file defaults/defaults.dso.gz.
```

Logfile lists all probe binding spots, their binding free energies, and fractional contribution of different probe types to the hotspot.

8.4 Druggable Sites

Druggable sites are identified by clustering probe binding spots and merging them to identify subsets of binding spots that have a size similar to that of a drug-like molecule. After results load, you will see a list of molecules in *VMD Main* for each druggable site and solutions therein. You can toggle displayed molecules to see locations of different sites and solutions.

Figure shows the best solution for protein MDM2. The maximal achievable affinity (druggability index) for this solution is 0.3 nM or, in terms of free energy, it is -13 kcal/mol. You can find such information in the log file shown above.

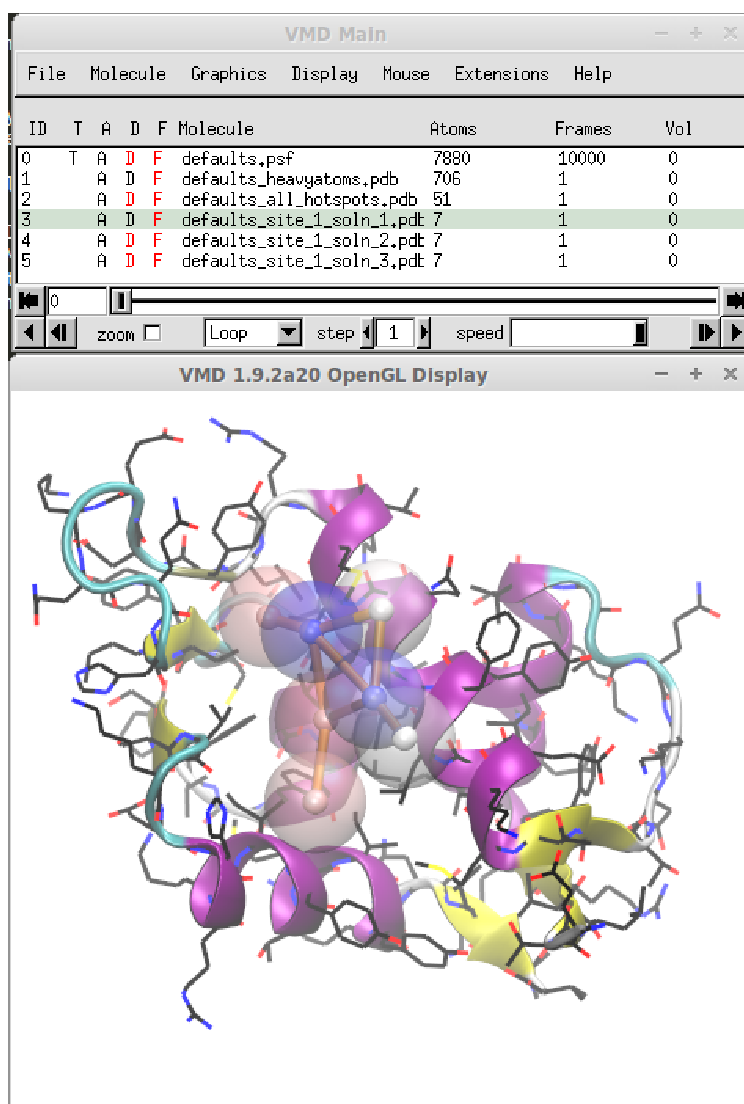
Probe binding hotspots and protein structure shown above can be found among tutorial files. These results of course deserve a more detailed analysis, and some things that can be done include:

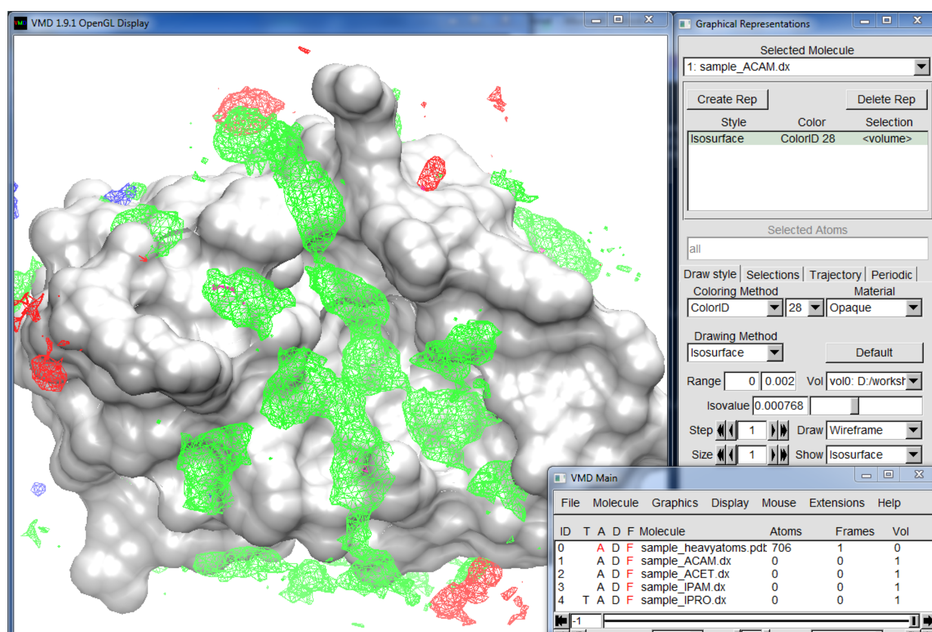
- looking into types of probes that contribute to a given binding spot and types of amino acid residue interacting with the binding spot
- visualizing trajectories (in which probes are wrapped) to see specific interactions and residence time of probes at a given binding spot
- comparing results from simulations in presence and absence of probes to see how binding site shape is affected by ligand binding
- looking into other structures of the target protein (ligand binding sites, crystal contacts, protein interfaces) to see whether observations in simulations are supported by interactions determined experimentally

8.5 Visualize Probe Grids

Finally, you can visualize probe occupancy grids using VMD. Simply load `.dx` files and create *isovolume* representations. An example is shown below for `sample_IPRO.dx`, and other grid files. Mesh surfaces correspond to locations that were highly enriched with probes. Coloring is as follows:

- isopropanol: green (high enrichment at the binding site)
- acetate: red (enrichment at the surface, not that proetein has +5 net charge)
- isopropylamine: blue (few interaction spots)
- acetamide: magenta (not observed to interact with this protein much)


VMD Main
File Molecule Graphics Display Mouse Extensions Help
ID T A D F Molecule Atoms Frames Vol
0 T A D F defaults.psf 7880 10000 0
1 A D F defaults_heavyatoms.pdb 706 1 0
2 A D F defaults_all_hotspots.pdb 51 1 0
3 A D F defaults_site_1_soln_1.pdb 7 1 0
4 A D F defaults_site_1_soln_2.pdb 7 1 0
5 A D F defaults_site_1_soln_3.pdb 7 1 0
zoom
Loop
step
1
speed
VMD 1.9.2a20 OpenGL Display



Note that values in occupancy grids is the count of central carbon atoms of probe molecules. Since the grid elements (voxels) are small (0.5A dimension), the occupancy numbers are small. You will need to adjust *Isovalue* value in VMD Representations window to make grid elements visible.

Similar representations can be generated for water or other atom type specific grids too.

CHAPTER NINE

INCORPORATE CGENFF MOLECULES

DruGUI also has a command utility, **drugui** for setting up a system containing molecules from CHARMM General Force Field.

In VMD *Tk Console*, typing the following command, for example, will prepare a system with probe composition of 30% imidazole, 30% isopropanol, and 10% of each of acetate, acetamide, isopropylamine, and isobutane:

```
% drugui mdm2.psf mdm2.pdb -IPRO 30 -IMID 30 -ACTT 10 -ACAM 10 -IPAM 10 -IBTN 10
```

There is a large selection of potential probes you can incorporate in a simulation. You can get a list of them by running **drugui** command, which will print:

```
% drugui
Info) Usage: drugui <psffile> <pdbfile> [options]
Info)
Info) Options:
Info)     -probes <Yes/no>
Info)       (use default probe mixture; see below)
Info)     -PROBE <percentage>
Info)       (probe type and percentage, e.g. -IPRO 80 -ACET 10 -IPAM 10)
Info)     -prefix <outname>
Info)       (data will be written to outname.psf/outname.pdb/etc.)
Info)     -outdir <directory>
Info)       (data will be written into specified directory, default is .)
Info)     -rotate <yes/No>
Info)       (rotate molecule to minimize water volume)
Info)     -padding <distance>
Info)       (minimum solvent box padding in all directions, default is 6 A)
Info)     -boundary <distance>
Info)       (minimum distance between water/probe and solute, default is 2.4 A)
Info)     -neutral <Yes/no>
Info)       (add counter ions, chloride or sodium, to make the system neutral)
Info)     -lipid <yes/No>
Info)       (system is solvated by considering lipid bilayer in xy plane)
Info)     -nsim <number>
Info)       (number of independent simulations, default 0)
Info)     -simlen <ns>
Info)       (length of individual simulations in ns)
Info)     -constrain <Heavy/calpha>
Info)       (which protein atoms to contraint in equilibration step)
Info)     -parameter <filename>
Info)       (additional parameter files; multiple occurrence is handled)
Info)
Info) Available probes -RESI [default_percentage] (name, charge, source)
```

```
Info)
Info) Core probes
Info)     -IPRO 70% (isopropanol, 0.0e, PBDA)
Info)     -ACAM 10% (acetamide, 0.0e, PBDA)
Info)     -ACTT 10% (acetate, -1.0e, PBDA)
Info)     -IPAM 10% (isopropylamine, 1.0e, PBDA)
Info)     -IBTN (isobutane, 0.0e, PBDA)
Info)     -IMID (imidazole, 0.0e, PBDA)
Info)
Info) Polar probes
Info)     -PRO2 (2-propanol, 0.0e, CGenFF)
Info)     -GLYN (glycine, 0.0e, CGenFF)
Info)     -ACEH (acetic acid, 0.0e, CGenFF)
Info)     -ACEM (acetamide, 0.0e, CGenFF)
Info)     -PRAM (propionamide, 0.0e, CGenFF)
Info)     -NMA (N-methylacetamide, 0.0e, CGenFF)
Info)     -ACO (acetone, 0.0e, CGenFF)
Info)     -TFE (trifluoroethanol, 0.0e, CGenFF)
Info)     -MP_0 (neutral methylphosphate, 0.0e, CGenFF)
Info)     -UREA (urea, 0.0e, CGenFF)
Info)     -MSAM (methanesulfonamide, 0.0e, CGenFF)
Info)     -MMSM (N-methylmethanesulfonamide, 0.0e, CGenFF)
Info)     -MMST (methyl methanesulfonate, 0.0e, CGenFF)
Info)     -DMSN (dimethyl sulfone, 0.0e, CGenFF)
Info)     -MESN (methyl ethyl sulfone, 0.0e, CGenFF)
Info)     -DMSO (dimethylsulfoxide, 0.0e, CGenFF)
Info)     -2PDO (2-pyrrolidinone, 0.0e, CGenFF)
Info)     -TMAO (trimethylamine N-oxide, 0.0e, CGenFF)
Info)
Info) Hydrophobes
Info)     -IBUT (isobutane, 0.0e, CGenFF)
Info)     -BUTA (butane, 0.0e, CGenFF)
Info)     -EMS (ethylmethylsulfide, 0.0e, CGenFF)
Info)     -DMDS (dimethyldisulfide, 0.0e, CGenFF)
Info)     -DFET (difluoroethane, 0.0e, CGenFF)
Info)     -TFET (trifluoroethane, 0.0e, CGenFF)
Info)     -DCLE (1,1-dichloroethane, 0.0e, CGenFF)
Info)     -TCLE (1,1,1-trichloroethane, 0.0e, CGenFF)
Info)
Info) Negatively charged
Info)     -ACET (acetate, -1.0e, CGenFF)
Info)     -PROA (propionic acid, -1.0e, CGenFF)
Info)     -CO3 (ionized carbonate, -2.0e, CGenFF)
Info)     -MP_1 (anionic methylphosphate, -1.0e, CGenFF)
Info)     -MP_2 (dianionic methylphosphate, -2.0e, CGenFF)
Info)     -MSNA (methyl sulfonate, -1.0e, CGenFF)
Info)     -ESNA (ethyl sulfonate, -1.0e, CGenFF)
Info)
Info) Positively charged
Info)     -GUAN (guanidinium, 1.0e, CGenFF)
Info)     -MGUA (methyl-guanidinium, 1.0e, CGenFF)
Info)     -AMDN (amidinium cation, 1.0e, CGenFF)
Info)
Info) 5-membered rings
Info)     -IMIA (imidazole, 0.0e, CGenFF)
Info)     -IMIM (imidazolium, 1.0e, CGenFF)
Info)     -MIMI (4-methylimidazole, 0.0e, CGenFF)
Info)     -THAZ (thiazole, 0.0e, CGenFF)
```

```
Info)     -TRZ4 (triazole124, 0.0e, CGenFF)
Info)     -PYRL (pyrrole, 0.0e, CGenFF)
Info)     -FURA (furan, 0.0e, CGenFF)
Info)     -THIP (thiophene, 0.0e, CGenFF)
Info)     -OXAZ (oxazole, 0.0e, CGenFF)
Info)     -ISOX (isoxazole, 0.0e, CGenFF)
Info)     -ISOT (isothiazole, 0.0e, CGenFF)
Info)     -PYRZ (pyrazole, 0.0e, CGenFF)
Info)     -OXAD (oxadiazole123, 0.0e, CGenFF)
Info)     -2HPR (2H-pyrrole, 0.0e, CGenFF)
Info)     -2PRL (2-pyrroline, 0.0e, CGenFF)
Info)     -2PRZ (2-pyrazoline, 0.0e, CGenFF)
Info)     -2IMI (2-imidazoline, 0.0e, CGenFF)
Info)     -PRLD (pyrrolidine, 0.0e, CGenFF)
Info)     -3PRL (3-pyrroline, 0.0e, CGenFF)
Info)     -PRLP (pyrrolidine protonated, 1.0e, CGenFF)
Info)     -3PRP (3-pyrroline protonated, 1.0e, CGenFF)
Info)     -2PRP (2-pyrroline protonated, 1.0e, CGenFF)
Info)     -2IMP (2-imidazoline protonated, 1.0e, CGenFF)
Info)     -2HPP (2H-pyrrole protonated, 1.0e, CGenFF)
Info)     -3HPR (3H-pyrrole, 0.0e, CGenFF)
Info)     -CPDE (cyclopentadiene, 0.0e, CGenFF)
Info)     -DIOL (1,3-Dioxolane, 0.0e, CGenFF)
Info)     -IMDP (Imidazolidine protonated, 1.0e, CGenFF)
Info)     -PRZP (Pyrazolidine protonated, 1.0e, CGenFF)
Info)     -2DHF (2,3-dihydrofuran, 0.0e, CGenFF)
Info)     -MCPE (methylcyclopentane, 0.0e, CGenFF)
Info)     -OXD4 (oxadiazole124, 0.0e, CGenFF)
Info)     -THF (tetrahydrofuran, 0.0e, CGenFF)
Info)     -THFM (Methyl-tetrahydrofuran, 0.0e, CGenFF)
Info)     -THFO (3'-hydroxyl-tetrahydrofuran, 0.0e, CGenFF)
Info)     -CPEN (cyclopentane north types, 0.0e, CGenFF)
Info)     -CPES (cyclopentane south types, 0.0e, CGenFF)
Info)
Info) 6-membered rings
Info)     -BENZ (benzene, 0.0e, CGenFF)
Info)     -PY01 (4H-Pyran, 0.0e, CGenFF)
Info)
Info)
Info) Notes:
Info)     - Passing "y" or "n" (case-insensitive) is sufficient for applicable options.
Info)     - When probe types are specified, probe percentages must add up to 100.
Info)     - When probe is "no", only water (and ions) will be added.
Info)     - Water segment name prefix is "WT".
Info)     - Ion segment name is "ION".
Info)     - Input molecule dimensions are used to determine size of the solvation box.
Info)     - When specified, all atoms of the system is rotated by 10 degree increments.
Info)     - Sodium and chloride ions are used to neutralize the system.
Info)     - Minimum distances from solute and between ions are set to 5 A.
```

CHAPTER TEN

LICENSE


```
Copyright (c) 2013, University of Pittsburgh
All rights reserved.

Redistribution and use in source and binary forms, with or without
modification, are permitted provided that the following conditions are met:
    * Redistributions of source code must retain the above copyright
      notice, this list of conditions and the following disclaimer.
    * Redistributions in binary form must reproduce the above copyright
      notice, this list of conditions and the following disclaimer in the
      documentation and/or other materials provided with the distribution.
    * Neither the name of the <organization> nor the
      names of its contributors may be used to endorse or promote products
      derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND
ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED
WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE
DISCLAIMED. IN NO EVENT SHALL <COPYRIGHT HOLDER> BE LIABLE FOR ANY
DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES
(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES;
LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND
ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT
(INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS
SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
```

BIBLIOGRAPHY

[AB12] Bakan A, Nevins N, Lakdawala AS, Bahar I Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules *J Chem Theory Comput* **2012** 8(7):2435-2447