VSS TUTORIAL

A VMD plugin for fast free energy scan based on a single trajectory

YingChih Chiang (Letitia), Yui Tik Pang (Andrew), Yi Wang

Department of Physics, CUHK

Introduction

VSS is a VMD plugin which is designed to post-process a single trajectory file for a fast free energy scan of different ligands. The typical example is to have a trajectory of a lead compound, and use this trajectory to calculate the free energy change of the alchemical transformation of this lead compound into its derivatives, e.g. free energy change when transforming a benzene into toluene.

In this tutorial, we will introduce how to use this plugin according to the following topics:

- Installation
- Set up input files for VSS
- Use the job script to run VSS
- Error Analysis

We stress that VSS fails if the protein conformation changes dramatically when binding to different ligands. For this reason, currently, it is better to use VSS for solvation free energy calculations.

Installation

The VMD plugin and its necessary dependencies are all included in the vssplugins.tgz file. To install the plugin, one simply needs to edit the first 3 directories provided in the install_vss.csh:

- VMDPLUGINDIR: the VMD plugin directory, where the original VMD plugin is installed.
- TCLINCDIR: the Tcl include files directory
- TCLLIBDIR: the Tcl library directory

Here is one example:

```
set VMDPLUGINDIR = ~/work/binary/vmd_plugins
set TCLINCDIR = /usr/include/tcl8.5
set TCLLIBDIR = /usr/lib/x86_64-linux-gnu
```

Then execute the installation script, ./install_vss.csh
VSS should be successfully installed if you see the following message:
“VSS plugin has been installed to your VMD. Enjoy!”

Set up input files

VSS requires the input files for free energy calculation. They are

- The dcd, xst trajectories and the corresponding psf file, e.g. a benzene in a water box.
- The pdb file of the ligand derivative and its corresponding psf file, e.g. a toluene.
- The parameter file.

1. Setting up the Parameter File

To set up the parameter file is easy: one simply pastes the parameter files that are used in collecting the dcd trajectory and the missing parameters (those reported with a penalty) reported by ParamChem, if there is any. For instance, for transforming a benzene into a toluene under the water solvation, the required parameters are par_all36_cgenff.prm and toppar_water_ions.str. Hence, the required parameter input file can be created via:

```
cat  par_all36_cgenff.prm > parfile.prm
cat  toppar_water_ions.str >> parfile.prm
```

The created parameter file (parfile.prm) can then be used in VSS calculation with the input argument -parfile parfile.prm

Set up input files

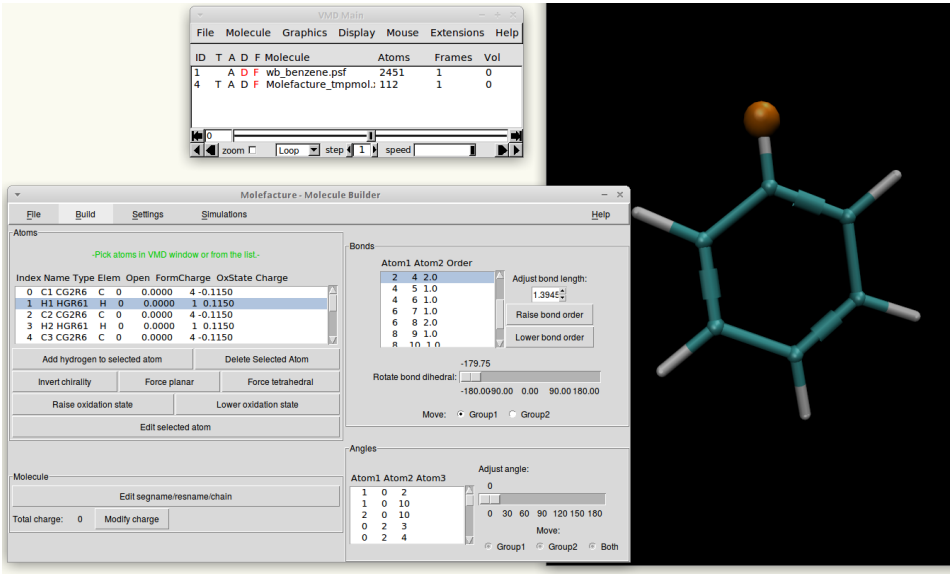
2. Preparing dcd, xst and psf files of the Lead Compound

VSS post-processes a given dcd trajectory for free energy calculations. Therefore, the user has to provide a dcd trajectory, together with the corresponding box size (xst file) and the molecule structure (psf file). Note that by default NAMD provides a rectangular box, so the VSS plugin processes only the rectangular box.

3. Preparing pdb file of the Derivative by Molefacture

Once the dcd is provided, one should construct the desired mutation based on the ligand inside the dcd trajectory. This can be done easily via the Molefacture plugin, provided by VMD:

- Load one frame from the dcd trajectory and the psf file into VMD.
- Open the plugin: Extensions → Modeling → Molefacture.
- Modify the ligand after selecting an H: Build → Replace hydrogen with fragment.
- Save the structure as a pdb file: File →Save.



Set up input files

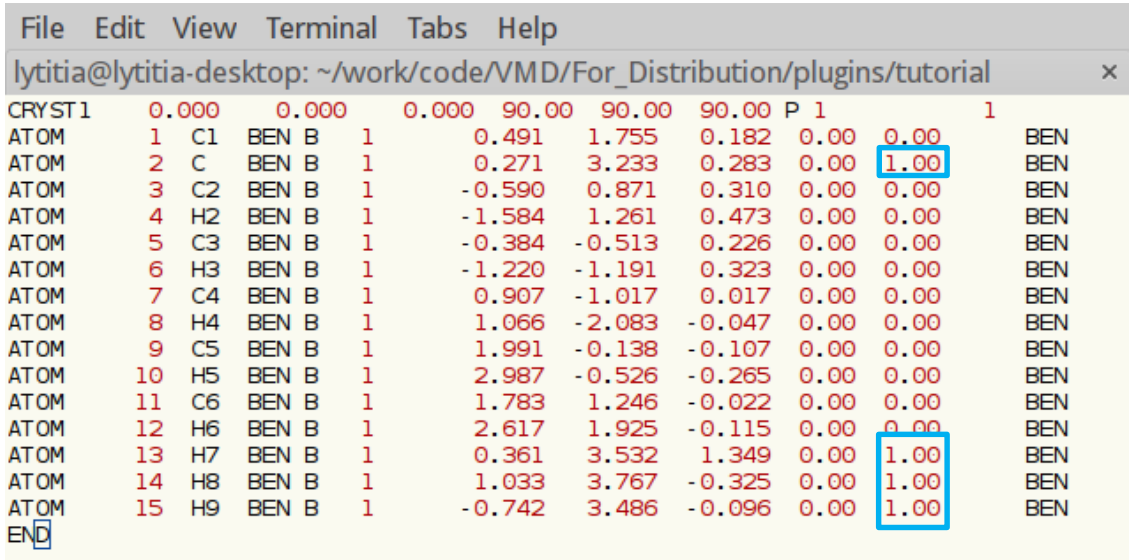
4. Performing Energy Minimization for the Derivative

The pdb file created by the Molefacture needs to be energy minimized before using it in VSS calculations. To do so, the psf file should be generated through ParamChem and the psfgen plugin. Sample configuration files for performing minimization by NAMD is provided in the directory “pdbprep”.

Please pay attention that one should use a **cutoff 0.1 and switchdist 0.05**, such that the van der Waals’ interaction is practically turned off. It is necessary to generate a ghost molecule that is minimized without the van der Waals interaction.

After energy minimization, label the beta column of the newly changed substitution, e.g. CH3 group, as 1. (Blue rectangular)

Finally, the pdb file can be used in VSS calculation.



Run VSS

1. Edit "runvss.tcl"

To run VSS, open the runvss.tcl. Possible usages are explained in the file. One only needs to set up the parameters listed in the script and decides which arguments to put in under "Run your VSS job".

```
# VSS usage:
# 4 possible cases:
# Both Dihedral and Angle sampling: vss -sel1 $sel1 -file1 $file1 -file2 $file
# Only Dihedral sampling:           vss -sel1 $sel1 -file1 $file1 -file2 $file
# Only Angle sampling:              vss -sel1 $sel1 -file1 $file1 -file2 $file
# No Dihedral nor Angle sampling:   vss -sel1 $sel1 -file1 $file1 -file2 $file
#
# Only sel1, file1, sel2, file2, parfile, mutvec, are mandatory input argument
# Additionally assign initial findex (fint), step of findex (fstep), end of fi
##############################################################################

# Set the parameters
set file1 "dcd/wb_benzene";     # lead compound's dcd, xst, psf (no extension)
set file2 "min_toluene";        # derivatives' pdb, psf (no extension)
set parfile "parfile.prm";      # used parameters (Charmm, CGenFF only)
set sel1 "resname BEN";         # ligand in dcd trajectory
set mutvec "0 1 0 1";           # site of mutation
set samdihedlist "2 0 1 12 ;";  # index of the dihedral angle to be sampled
set samdihedval "0 30 60 90 ;"; # value of the dihedral angle to be sampled
set samanglelist "0 1 12 ;";    # index of the angle to be sampled
set samangleval  "108.57 ;";    # value of the angle to be sampled

# Load the VSS plugin
source ./vss.tcl

# Run your VSS job
vss -fstep 1 -temp 300.0 -sel1 $sel1 -file1 $file1 -file2 $file2 -parfile $par
```

Parameters are self-explanatory, except for mutvec, which denotes the mutation sites on the lead compound and its derivative. It is composed of two set of indices: the indices of mutation site C-H in the dcd file of the lead compound, where the H is going to be replaced by a functional group A; and indices of corresponding sites (C-A) in the pdb file of the derivative.

2. Execute the script by VMD

After setting up the script file, run VSS by the command (in shell): **vmd -dispdev none -e runvss.tcl**

Here we provide a 1ns trajectory of benzene, and toluene.pdb for practice. You may want to try it now.

Advanced usage

1. Multiple dihedral angles or multiple angles

If the dihedral angle links to a methyl group, its sampling can be ignored. The current VSS code does not support the sampling of multiple dihedral angles or multiple angles in the correct product form. So, one has to generate different conformations for the second dihedral angle and submit all of them for VSS calculation. VSS will provide energy differences and Boltzmann weights for each conformation at each frame. One can combine them to get the final result. The next VSS version will allow this operation.

2. Sampling one dihedral angle and one bond angle

The current VSS code supports such sampling in product form. Simply input the angle and dihedral angle parameters in runvss.tcl and follow the comments in the script.

3. Using different Hamiltonians

The current VSS code supports different Hamiltonians (HamI or HamII) for sampling. We recommend to use HamII by including –D Ham in the Makefile under the vss directory (default). In this way one only needs to sample the dihedral angles to achieve an accurate sFEP calculation. Other internal degrees of freedom of the ligand (bond and angle) are irrelevant. However, if one wishes to compare the VSS result with NAMD, please remove –D Ham from the Makefile.

Error Analysis

The error analysis can be performed for VSS calculation, using the subsampling method, developed by Ytreberg and Zuckerman (J Comput Chem 25: 1749-1759, 2004).

Here we provide two small C routines in the directory “erroranalysis”, one for evaluating the correlation time (autocorr.c), another for evaluating the error (subsampling.c) based on the subsampling method. Note that one has to **first evaluate the correlation time, and use only the non-correlated data for error analysis!**

To run the two routines, one need to prepare a file that contains the time (or frame index) and the DeltaE, which are printed out by VSS plugin in file vss.log. Execute the following command to produce a file (dE.log) that can be used in the error analysis:

```
cat vss.log | sed '/^#/d' | awk '{print $1, $4}'  > dE.log
```

The two routines are written in an interactive way to ask for input. After compiling the routines, simply execute them. They will ask for necessary inputs. Note that the sample calculation provided is too short, so a very large error bar is expected.

Remark

- Different from other VMD plugins, VSS calculations are quite heavy. Therefore, one should avoid any non-necessary tasks, e.g. X-window. For this reason, no GUI is implemented for VSS. Additionally, OpenMP is employed for parallelization for further speedup. With this implementation, one can also run the VSS calculation on a multicore cluster.

- For the VSS calculation, the dcd trajectory and xst file need to be stored according to the expected correlation time. For instance, if the lead compound is placed in water, its correlation time is quite short, typically within 40 fs. In this case, the output frequency of 40 fs will be sufficient. However, if the lead compound binds to a protein, the correlation time can be very long, typically of hundreds of fs, and hence a less frequent dcd output (0.1ps to 1ps) will be sufficient.

- If you use this plugin, please cite our paper “Virtual substitution scan via single-step free energy perturbation” Biopolymer **105**, 324-336 (2016).

Thank you for your interest in VSS!