Relative Binding Free Energy Calculation of a Pair of Congeneric Small Molecules Towards the Same Protein Target: A Free Energy Perturbation (FEP) Tutorial Using NAMD

Quantao Sun

quantaosun@gmail.com

https://quantaosun.github.io/post6.html for a side-to-side reading.

Free energy perturbation is one of the state-of-the-art technologies in drug design, it has a very strong theoretical foundation and has been verified in various hit-to-lead research.

However, the tedious setting up and careful preparation procedure FEP requires is a major factor that contributes to its limited usage. Commercial sources like FEP plus do provide a user-friendly interface but the heavy cost is not suitable for everyone.

In this tutorial, a step-by-step procedure of how to set up a basic FEP, how to run the simulation with an open-source molecular dynamic package NAMD and how to analyze the result, was introduced.

The time to set up the input files for this FEP calculation is around 1.5 hours.

The time to running the simulation need around 10 hours, with 32 CPUs on NAMD/2.13-mpi.

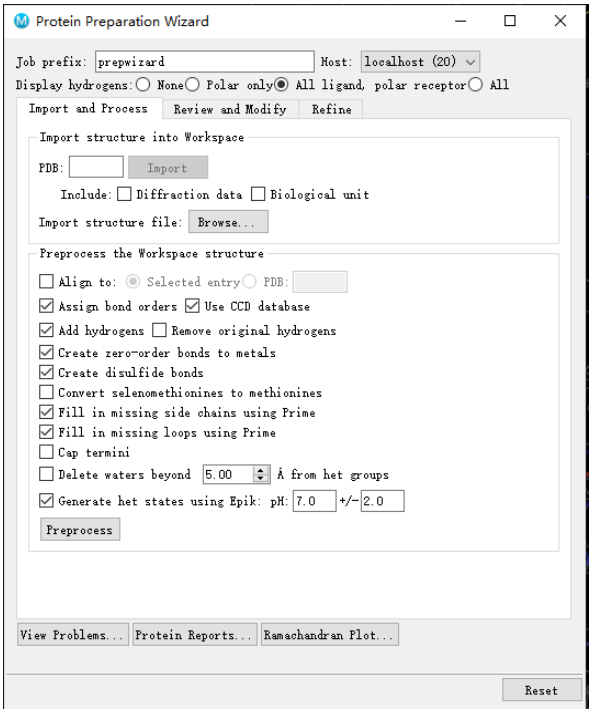
----------Resources Required ----------

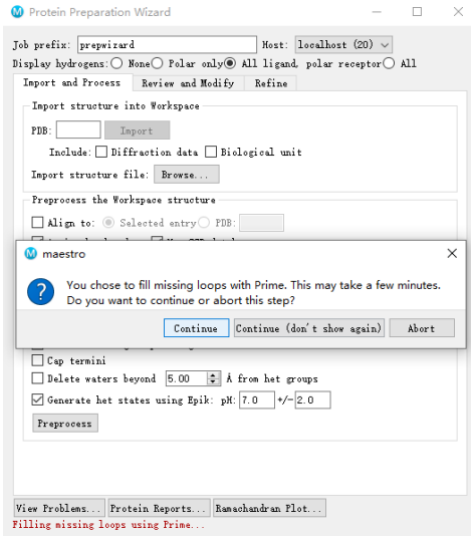
1. https://github.com/quantaosun/fep_prepare
2. LigPargen webserver
3. Feprepare webserver
4. Schrodinger Maestro 2018-4 version and above
5. NAMD/2.13-mpi or another version
6. VMD 2019 and above version
7. Sublime Text editor
8. FileZilla client.

----------*Step 1 Target protein preparation*----------

The crystal structures are 1MQ5 and 1MQ6, the ligand from 1MQ5 (XLC) will be taken as reference, ligand from 1MQ6 (XLD) as mutation and the fundamental question is, is XLD a better ligand compared to XLC, in terms of binding to the same protein target?

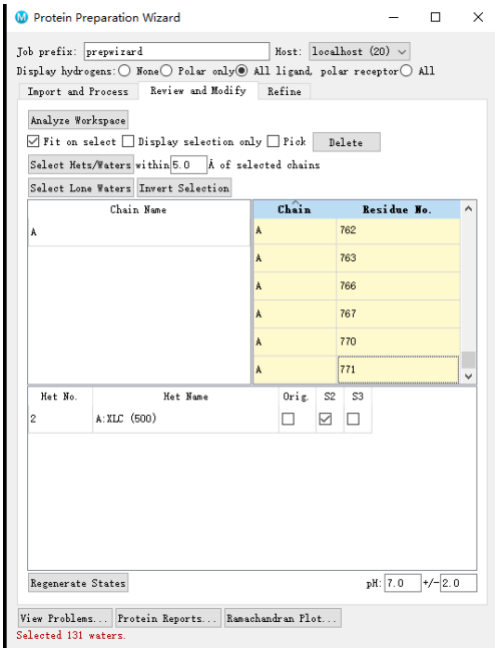
The first thing to do, is to CAREFULLY prepare your protein in Maestro, make sure you tick the "fill missing loops" option.



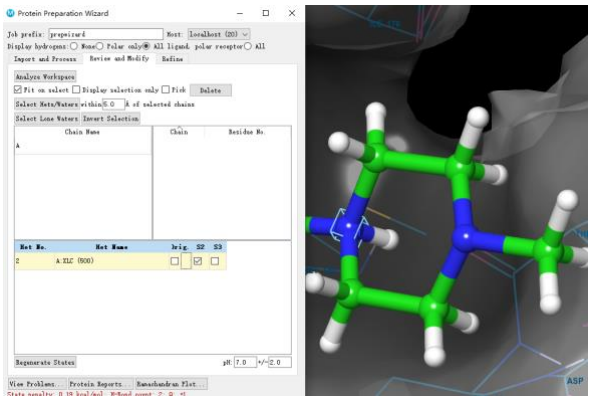


Use Get PDB by ticking the "biology unit" option, otherwise filling missing loops may fail

There are two chains, A and L, the ligand binds to the green chain A, for the sake of this tutorial, chain L was deleted, but ideally, you should keep it in an FEP calculation, especially when it plays a role in helping the small molecule binding.



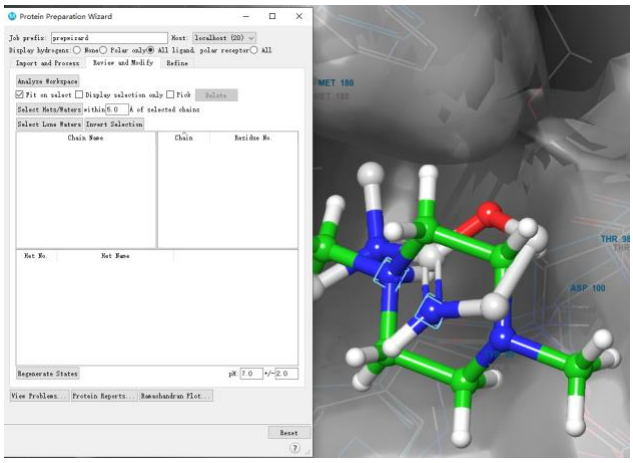
All waters and other co-factors were deleted, only XLC ligand was kept as the Reference ligand.



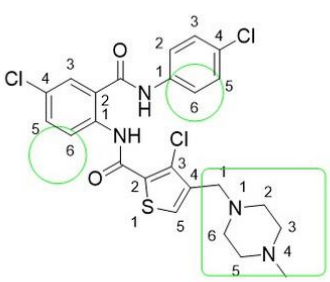
You may have noticed that the generated state for XLC is protonated at the nitrogen position, i.e., the molecule brings one positive charge, keep that in mind.

Continue to finish the protein preparation procedure.

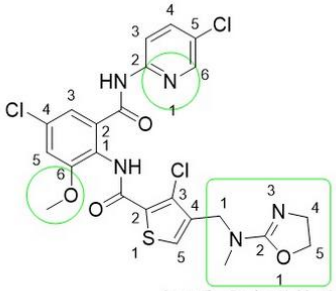
DO THE SAME PEROCEDURE FOR 1MQ6.



Having finished all preparation for XLD and XLC, both have a charge near the mutation area (image above). You could have a better vision by the selected rectangle atoms shown below.



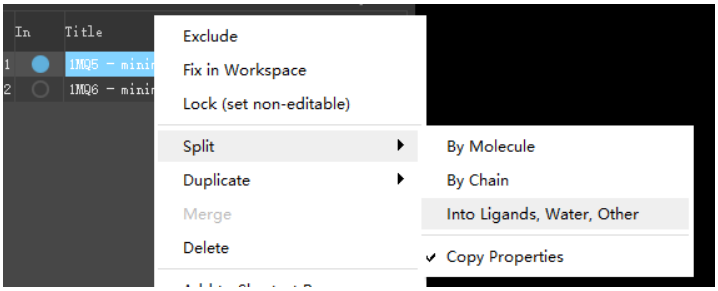
XLC



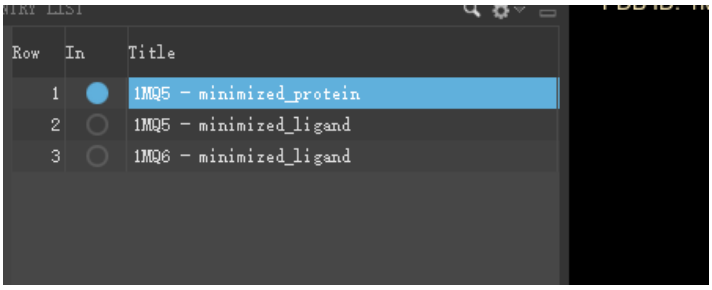
XLD

It has turned out with a positive charge, the Feprepare web serves returned an error, making it not possible to generate the input files for further simulation.

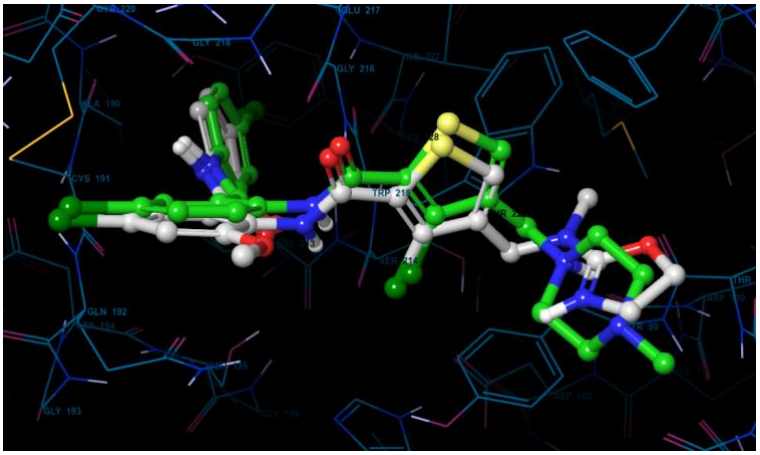
Both small molecules were then decided to be put back to their original state, i.e., the neutral states, with a proper record of the corresponding protonation state penalty, S1 and S2, at the end of the FEP, the ΔΔG = ΔΔG(FEP) +(S2-S1).



Now, split the entry into ligands, water, others, then delete anything else but the next three entries, please double-check the protonation state of small molecules are neutral.



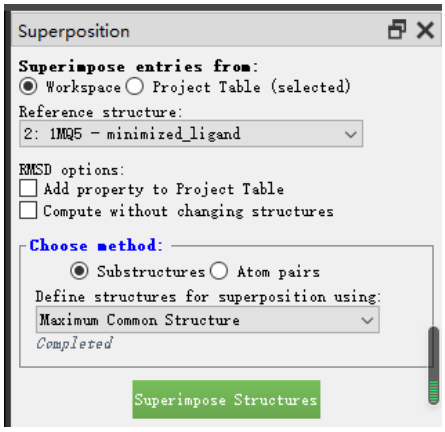
Now let's align the mutation XLD to the reference XLC



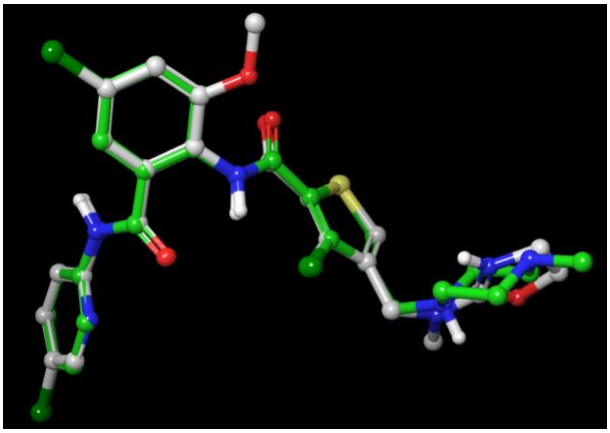
the two ligands are already in a good position with each other, but for the sake of an easier fep calculation, we want an even better alignment.

Note, this alignment tool may not exist in another version of Maestro or is slightly different.

In 2018-4 Schrodinger (maestro 11.8) the tool we need to use is "flexible ligand alignment"

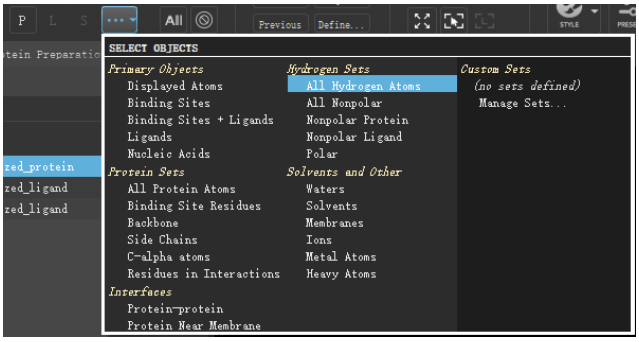


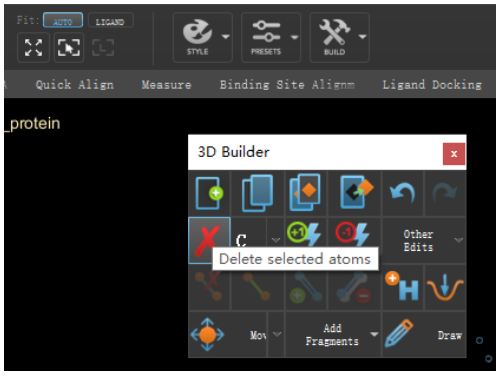
While the module you should use on the 2020 version of Schrodinger (maestro 12.6) is called "superimpose structures", after alignment, now the two looks like



Note: In some other cases, you derive your mutation from your reference with the build tool in Maestro, in that situation, the two ligands are already naturally aligned so you don't have to do anything else to align them again.

After the alignment is done. Let's save the XLC.pdb, XLD.pdb.





Select all the hydrogens for the protein and use the build tool to delete all the selected H atoms, then save the protein.PDB

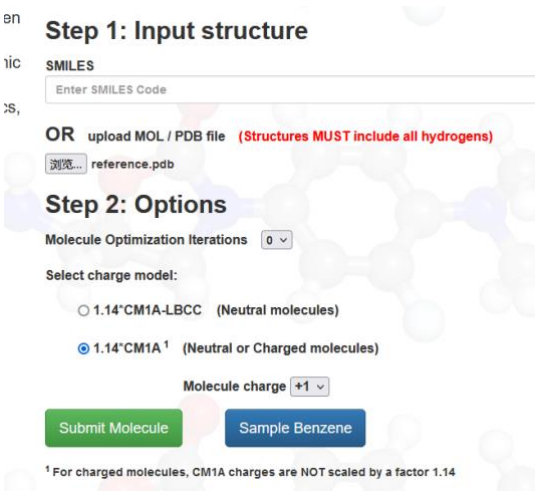
```
1862 ATOM   1842  CG  LYS A 243      35.665  16.987  10.130  0.50 39.43           C
1863 ATOM   1843  CD  LYS A 243      35.230  16.410  11.484  0.50 38.66           C
1864 ATOM   1844  CE  LYS A 243      36.072  16.880  12.671  0.50 36.06           C
1865 ATOM   1845  NZ  LYS A 243      35.513  16.377  13.938  0.50 30.26           N1+
1866 CONECT   47   83
1867 CONECT   83   47
1868 CONECT  206  324
1869 CONECT  324  206
1870 CONECT 1240 1351
1871 CONECT 1351 1240
1872 CONECT 1433 1644
1873 CONECT 1644 1433
1874 END
1875
```

now you are suggested to open the protein. PDB with Sublime text editor and scroll down to the bottom to delete all the "**CONNECT**" lines, and save the file.

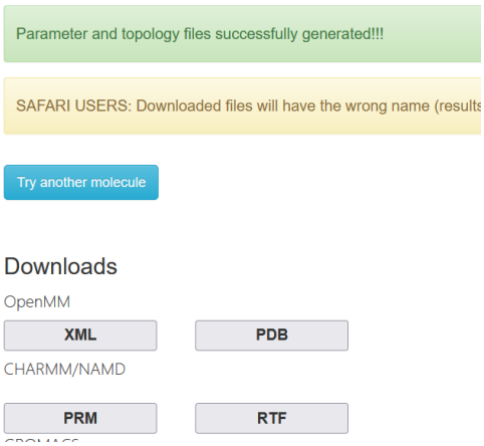
Do the same manipulation for your two small-molecule files as well.

----------*Step 2 Files uploaded to LigParGen webserver*----------

http://zarbi.chem.yale.edu/ligpargen/



Upload reference.PDB and mutation. PDB to the server one by one, download RTF and PRM files, separately. Make sure you select the right charge, 0, in our case (ignore the image above).

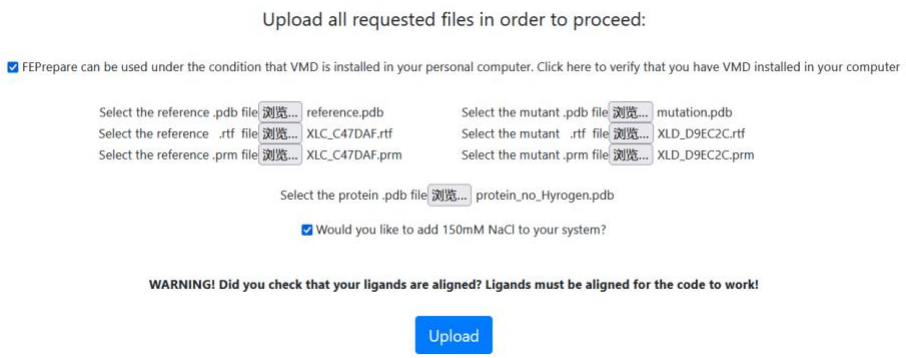


Be careful about the names, the LigPargen may have renamed your file, just be sure which is

which.

----------Step 3 Upload files to Feprepare webserver----------

https://feprepare.vi-seem.eu/



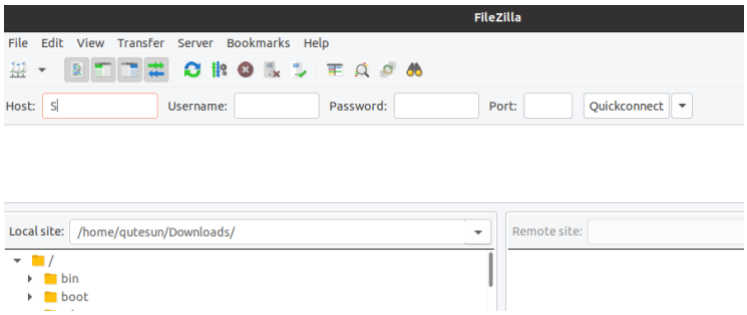
Upload all necessary files as above, then click upload. After finish,

Click the download files button, and save the zip file

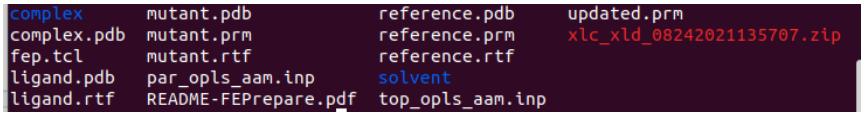
----------Step 4 Transfer files to where NAMD installed----------

Upload the downloaded zip file from Feprepare to your remote cluster or your working states if you have one. In my case, I upload this file through the FileZilla client.

You could also use the SSH command if you can copy the zip file to the remote cluster.

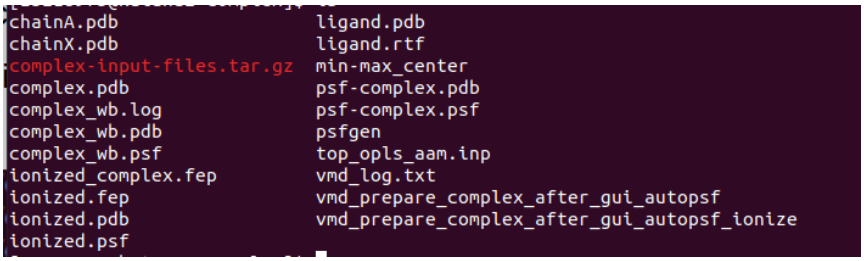


Next, we need to unzip and modify these files a bit before they could be used in the simulation.



The structure of the files is there are two folders namely “complex” and “solvent”, the plan is we do both forward and backward simulation for complex and solvent, respectively, then ΔΔG(FEP) = dG(complex) – dG(solvent). Don’t worry, we will learn how to calculate dG(complex) and dG(solvent) by VMD later. Inside the folder like below

Now let’s start with entering the complex folder.



The red tar.gz file contains all the configuration files we need, just extract it with tar command.

----------Step 5 Modify NAMD configuration files----------

https://github.com/quantaosun/fep_prepare

Download the toppar_modified.zip to the relative path or your simulation working directory as defined in the configuration file, as per the path defined by the NAMD configuration file. Please also download the three namd configuration files end with a .namd end.

Indeed, there are already a lot of configuration files that come back from Feprepare, but they are not complete based on my test, and the three purple ones are the modified version and should be ready to simulate,

Just remember to make sure all defined parameters inside the .namd file could be found, e.g., the location of the toppar folder should be placed correctly to the current folder displayed below. Change the simulation steps based on your need, at the end of the *.namd file, change the periodical condition values as well if you use a crystal structure different from this tutorial.

It should be emphasized that I did not apply the way Feprepare default simulation logics, i.e., run the md_forward_namd one by one from 1 to 16, then backward from 1 to 16, you could do that if you are confident enough to handle everything, but in this tutorial, I just decide to only handle three configuration files, namely the nvt_equil_modi.namd, npt_equil_modi.namd and the md_forward_modi.namd and md_backward_modi.namd, since inside md_forward and md_backward file, there are also a 16 lambda windows defined, so they are still a fair equivalent to the default way to do the job. The only reason I go this way is it is unlikely to mess things up since there are fewer files to deal with.

----------Step 6 Simulate with NAMD----------

There are at least two scenarios, one is you use a remote cluster; the other one is you use a local working station or local computer. The second one is pretty straightforward; we will not discuss it here.

For the first scenario, we need to be a bit careful to correctly submit the simulation job, in my case, I just run an interactive job with fewer steps first, if everything is right, then I change the number of steps to a bigger number then submit the job in a batch mode. The queue system varies from cluster to cluster, mine is PBS, an example for an interactive job may look like this:

```
$ qsub -I -l select=1: ncpus=32:mem=40gb,walltime=10:00:00
$ cd $PBSS_O_WOKDIR
$ module load intel/18.x.x  openmpi/3.1.2-intel namd/2.13-mpi
```

(you need double check the preconditions for namd/2.13-mpi, it varies)

Then simulation was started by issuing commands as next two lines,

```
$mpirun  -oversubscribe -np 4 /apps/namd/2.13-mpi/arch/Linux-x86_64-icc/namd2 +ppn 7 \ nvt_equil_modifiled.namd > nvt_equil.out.
```

NOTE: If it is your first-time run, it is pretty normal you get ERRORS, in most cases, all these errors could be debugged by modifying your configuration file, or add (cross off) parameter (toppar) lines in the conf file.

If everything looks good, what you then need to do is waiting for the NVT to finish, then similarly do the NPT, then run PRODUCTION run with md_forward_modi.namd, at last run the md_backward_modi.namd.

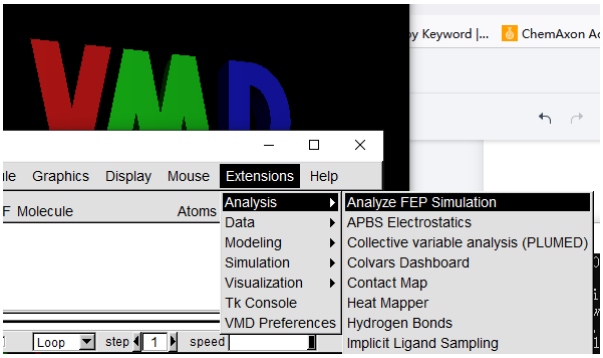
AFTER THE COMPLEX FOLDER IS DONE, DO THE SAME THING TO THE SOLVENT FOLDER.

When you do the solvent folder, there may be an error related to "original water box too small", you could just try to increase the margin in the configuration file from 1 to a bigger number, e.g., 10, and try again, usually, it will solve this.

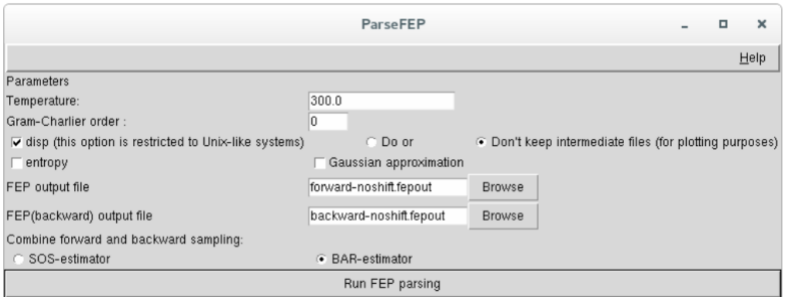
----------*Step 7 Result analysis*----------

Well, time for the LAST section, analysis and get the ΔΔG result!

Ideally, you should use a Linux version of VMD, if you can't access one, then just use the Windows version.



Find analyze FEP simulation from Extensions > analysis panel.



Change the temperature to your simulation value, select both forward and backward report files, click Run FEP parsing, then you will get the ΔG complex, similarly, ΔG solvent was obtained.

ΔΔG FEP = ΔG complex- ΔG solvent

Remember the protonation state penalty we have discussed earlier, S1, S2.

ΔΔG = ΔΔG FEP + (S2-S1)

Things to consider using this FEP protocol

How good do you understand your protein, like are you sure there is no broken loops or modified residues?

Are these two molecules share a common mother core or at least share a 60% of similarity?

How do you align your two molecules, which one is the reference, and which one is mutation?

Is your protein a dimer or multiple polymers when executing its biological functions?

Finally, are you in a situation, that one of the pairs of molecules is charged while the other one is not?