# SUPPORTING INFORMATION

# Empirical Valence Bond

# Simulation Of *Cis* Stilbene-Oxide by Limonene Epoxide Hydrolase Using GROMACS And Q Software

## THE WORKFLOW

The workflow presented below is for the hydrolysis simulation of *cis*-stilbene oxide by the enzyme Limonene epoxide hydrolase. The force field (OPLS-AA), the topology files, tools, scripts and other intermediary files used during the simulation can be found at https://github.com/aderaissa/Thesis-Files.git.The tools and commands presented below were used during the simulation and they run and execute on Linux terminal. In this report we will consider the folders and file names as there as there are on the github page.

## PART A: SIMULATION WITH Q6

## LOCATION: EVB_Q6 (For wild type and the simulation in water)

### I. Preparation of the system (building the topology) using Qprep6 (Directory: 01_preparation)

#### a. PDB file preparation

The pdb file used here is a docked system of the substrate in the binding pocket of the Protein. Fix the pdb file obtained from Yasara to be suitable for Q6; remove all the water molecules except the catalytic water, rename the catalytic water as HO2, remove all the Hydrogens in the system, renumber the residues in the system, rename the substrate as STO instead of SUB, change the ASP101 to ASH101 (since it is a protonated Aspartate) and change the Histidine name in the system (HIS) to HIE so that it can be recognized by the Q-oplsaa library. The resulting pdb file is called clean.pdb

#### b. Substrate and Product Preparation

1. Calculate the RESP charges of the reactants and products with Antechamber from AmberTools. The command below can be used to do that from each of the output files of a Gaussian calculation.

   ```
   antechamber -fi gout -i Stilbene_oxide.log -fo ac -o Stilbene_oxide.ac -c resp -nc 0
   ```

2. Get the parameter and library files for the reactants (H2O and cis-stilbene oxide) and the various products (RR-Diol, SS-Diol) using the ffld_server tool in maestro. Start by generating the -mae file from your pdb structures in maestro and convert the file into ffld format.  Below is the command used for Stilbene oxide.

```
$PATH(ffld_server.exe)  -imae Stilbene_Oxide.mae -version 14  -print_parameters -out_file Stilbene_Oxide.ffld
```

3. Create the library(.lib) and Parameter file (.prm) using q_ffld2.py from q-tools.

```
q_ffld2q.py -o Stilbene_oxide Stilbene_oxide.ffld Stilbene_oxide.pdb
```

4. Add the parameter files of the product and reactants into that of the forcefield. Rename the parameter file as: mysystem.prm.

   **NOTE**: Q reads only one parameter file but can read numerous library files

5. Add the RESP charges of the reactants and products to the generated library files.

   **NOTE**: The sum of the RESP charges of every charge group must be an integer

### c. Creating the Topology file And New PDB file

Make and write the topology and pdb file using the commands below. These commands prepare a system encompassed by a spherical boundary made of water molecules at a radius of 20Å. The center of the sphere is built around the oxygen atom of the epoxide. These writes the topology file RRDiol.top and the pdb file RRD.pdb which will be subsequently used during the simulation.

```
Qprep6
readlib   qoplsaa.lib
readlib  ./Stilbene_oxide/Stilbene_oxide.lib
readlib  ./RRDiol/RRD_Product.lib
readlib  ./ff/H2O.lib
readprm   mysystem.prm
readpdb  clean.pdb
boundary sphere 300:O99 20.
solvate  300:O99 20. grid HOH
maketop  RRDiol.top
writetop RRDiol.top
writepdb RRDiol.pdb y
quit
```

## II. Create the FEP file (Directory: 01-Preparation, Sub-directory:Fep)

1. Start by generating the fep.qmap file. This file should contain the q-atoms (the atoms participating in the reaction) defined by their atom type, PDB ID and LIB ID. This file maps the transformation of the reactant species to products.
2. Run the script makeFEPcys.py to generate the FEP file; RRD_Diol.fep.
3. Add the missing softcore potentials, morse potentials and vander waals parameters into the generated file.

## III. Relax the system (Directory: 02_relax)

1. Carry out eight sequential relaxation steps using input files generated from the "genrelax.proc" file. Each relaxation file should contain the number of steps, the time step, the temperature and other relevant MD parameters. The files: genrelax.proc, the topology file, the generated pdb file and the fep file should be passed through the qtools script; q_genrelax.py. The Relaxation input files can be gotten by running the file "commands". This generates a folder relax_001 which contains the eight relax input files.
2. Run these files using the Qdyn6 module in Q by the script; run_relax_q.sh or individually.

```
Qdyn6 relax_001.inp > relax_001.log
```

This results in an output file (.log), a restart file(.re) and a trajectory file(.dcd).

## III. Running the Free Energy Perturbation (FEP) calculation (Directory:03-FEP)

1. Free Energy Perturbation input files should be created using the file "genfeps.proc" for ten replicas (depending on your choice) with each replica having 51 frames characterized by different coupling parameters. Prior to FEP, equilibration should be done on the relaxed system at 300K whereby the last equilibrated system is passed to the FEP calculation. The various replica files containing the equilibration and FEP input files can be created by running the input file "commands". This generates replicas contained in different directories labelled accordingly.
2. Successive equilibration and FEP calculation can be done by simply submitting a batch job using the batch script; run_feps_q.sh. This script runs Qdyn6 on each equilibrated state and FEP state resulting in an output file (.log), a restart file (.re), trajectory file(.dcd) and in the case of FEP an energy file (.en) in addition. The movie for each FEP can be seen on VMD by concatenating the various dcd files and loading the file or

by placing the script "load_evb.py" and the pdb file into the FEP directory and using the command below.

```
Python load_evb.py RRDiol.pdb $path_dcd_files
```

## IV. EVB Runs

The Energy files gotten from Qdyn6 are then passed to Qfep6 using the input file qfep.inp. The input file specifies the number of energy files, the off-diagonal energy value, the gas shift constant and other relevant parameters.

```
Qfep6 < qfep.inp > qfep.out
```

The output file contains the Gibbs free energy needed to generate plots for the various replicas.

## V. Post Processing Analysis Using Q-Tools (Directory: 04-Calibartion)

The various scripts and packages contained in qtools should be used to plot the results obtained from Qfep6. In the folder 04-calibration, run the file "commands" to plot the data. Interpret the results and compare with experimentally reported data. If the free energy is not same as the experimentally determined values, keep adjusting the off-diagonal value and the gas shift constant. When you get the most suitable values for these variables, use them for all the mutants you have.

**NB:** Relevant files and scripts pertaining to the quantum cluster calculation are found in the directory QM Files.

**Table S1.** EVB parameters for the reaction of *cis*-stilbene oxide substrate catalyzed by Limonene Epoxide Hydrolase. The parameters include the RESP charges for the EVB atoms in reactant (RS) and product (PS) states, soft exponential repulsion for the atoms that form or break chemical bonds and Morse potential parameters. The EVB atoms include the side chain of Aspartate 101 (TYR), Aspartate132 and the substrate (STO). The atom names are the same as in topology and are shown in Figure S1. The residue names are the same as in topology as well. These parameters ca be found in the fep file in the FEP sub-directory

## PARAMETERS UTILIZED FOR THE EVB SIMULATION

| #Atom_type | LJ_A | LJ_B | SP_Ci | SP_ai | LJ_A_14 | LJ_B_14 | mass |
|---|---|---|---|---|---|---|---|
| sto.O99 | 445.125 | 18.2511 | 170.0 | 2.5 | 314.7509 | 12.9055 | 15.999 |
| rrd.O99 | 760.645 | 25.0448 | 180.0 | 2.5 | 537.8572 | 17.7093 | 15.999 |
| sto.C98 | 944.518 | 22.0296 | 190.0 | 2.5 | 667.8751 | 15.5773 | 12.011 |
| rrd.C98 | 944.518 | 22.0296 | 110.0 | 2.5 | 667.8751 | 15.5773 | 12.011 |
| sto.C96 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C96 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C1 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C1 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C2 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C2 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C3 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C3 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C4 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C4 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C5 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C5 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C99 | 944.518 | 22.0296 | 1 | 2.5 | 667.8751 | 15.5773 | 12.011 |
| rrd.C99 | 944.518 | 22.0296 | 1 | 2.5 | 667.8751 | 15.5773 | 12.011 |
| sto.C97 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C97 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C10 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C10 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C11 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C11 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C12 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C12 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.C13 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C13 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| sto.C14 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| rrd.C14 | 1059.1297 | 23.6736 | 1 | 2.5 | 748.9178 | 16.7398 | 12.011 |
| sto.H98 | 84.5728 | 5.4127 | 1 | 2.5 | 59.802 | 3.8274 | 1.0079 |
| rrd.H98 | 84.5728 | 5.4127 | 1 | 2.5 | 59.802 | 3.8274 | 1.0079 |
| sto.H1 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H1 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H2 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H2 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H3 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H3 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H4 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H4 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H5 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H5 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H99 | 84.5728 | 5.4127 | 1 | 2.5 | 59.802 | 3.8274 | 1.0079 |
| rrd.H99 | 84.5728 | 5.4127 | 1 | 2.5 | 59.802 | 3.8274 | 1.0079 |
| sto.H10 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H10 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H11 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H11 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H12 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H12 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H13 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H13 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| sto.H14 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| rrd.H14 | 69.5797 | 4.9095 | 1 | 2.5 | 49.2003 | 3.4715 | 1.0079 |
| HW | 0.0 | 0.0 | 1 | 2.5 | 0.0 | 0.0 | 1.008 |
| HO | 0.0 | 0.0 | 1 | 2.5 | 0.0 | 0.0 | 1.008 |
| OW | 762.89 | 24.39 | 180.0 | 2.5 | 539.4447 | 17.2463 | 16.0 |
| rrd.O88 | 760.645 | 25.0448 | 160.0 | 2.5 | 537.8572 | 17.7093 | 15.999 |
| rrd.H88 | 0.0054 | 0.0433 | 1 | 2.5 | 0.0038 | 0.0306 | 1.0079 |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| C2 | 1802.24 | 34.18 | 1 | 2.5 | 1274.3761 | 24.1689 | 12.01 |
| ODE | 601.15 | 22.27 | 1 | 2.5 | 425.0772 | 15.7473 | 16.0 |
| O2 | 616.44 | 23.77 | 1 | 2.5 | 435.8889 | 16.8079 | 16.0 |
| O1 | 616.44 | 23.77 | 1 | 2.5 | 435.8889 | 16.8079 | 16.0 |
| rrd.H89 | 0.0054 | 0.0433 | 165.0 | 2.5 | 0.0038 | 0.0306 | 1.0079 |
| CT | 944.52 | 22.03 | 1 | 2.5 | 667.8765 | 15.5776 | 12.01 |
| HC | 84.57 | 5.41 | 1 | 2.5 | 59.8 | 3.8254 | 1.008 |

## PART B: EVB SIMULATION WITH GROMACS

### I. Preparing the system and building the topologies (Topology)

1. Calculate the RESP charges using the same method above. This time, also do the calculation for an Aspartate molecule (Ash101) and a deprotonated Aspartate (Asp132) capped with a methyl group.
2. Extract the pdb from the gaussian output file for all the reactants and products. Below is an example illustrated for cis- Stilbene_oxide (STO).

   ```
   antechamber -fi gout -i STO_gesp.log -fo pdb -o STO.pdb -rn STO
   ```
3. Use the extracted pdb files to calculate the ffld_server parameters using maestro as done in the case of Q6.
4. Convert the ffld_server parameters into GROMACS' OPLS-AA force field format using the python code (ffld2gmx.py). For cis-Stilbene oxide substrate we will obtain the following files: sto_atomtypes.opls, sto_vdw.opls, sto_bonds.opls, sto_angles.opls, sto_torsions.opls, and sto_impropers.opls.

   ```
   ffld2gmx.py -n STO -f STO.ffld -a STO.ac
   ```
5. Add the content from sto_atomtypes.opls to atomtypes.atp, add 3ov_vdw.opls to ffnonbonded.itp, and 3ov_bonds.opls, sto_angles.opls, sto_torsions.opls, and sto_impropers.opls to ffbonded.itp file of GROMACS' force field. (For convenience, you can write only the parameters for RS in ffbonded.itp file and substitute the bonding types for the PS atom types inside ffnonbonded.itp with the corresponding bonding types for RS). In these files, the representation of H2O is changed to WHO.

6. Add dummy atom types for all EVB atoms in ffnonbonded.itp and atomtypes.atp files (see the files at the address indicated above).
7. Build the corresponding residue inside aminoacids.rtp file (see STO residue inside aminoacids.rtp); we must build only the residues corresponding to Reactant state.
8. Build the corresponding residue inside aminoacids.rtp file for the Reactant state. Edit the aminoacids.rtp file and the aminoacid.hbd file so that the residues of the protonated Aspartate and the protonated Histidine are represented in a similar manner.
9. Build GROMACS topology.

```
gmx pdb2gmx -f STBO.pdb -o STBO-start.pdb -water spc -merge all
```

10. Build the periodic box.

```
gmx editconf -f STBO-start.pdb -o STBO-box.pdb -c -d 2 -bt
dodecahedron
```

11. Solvate the system

```
gmx solvate -cp STBO-box.pdb -cs spc216.gro -o STBO-solv.pdb -p
topol.top
```

12. Add ions. First check the topol.top file and if some of the improper angles of STO are missing, copy them from the ffbonded.itp into the topol.top file.Then proceed with the following commands.

```
echo > dummy.mdp
gmx grompp -f dummy.mdp -o dummy.tpr -p topol.top -c STBO-solv.pdb
-maxwarn 1
gmx genion -s dummy.tpr -o STBO_ion.pdb -p topol.top -neutral
choose Group 16 (SOL)
```

13. The qmatoms.dat file should be created. It should contain the various qatoms in the system, the RESP charges and the morse parameters. Ensure the charges in the reactant state and the product state for each species in the Qmatoms.dat file should be an integer (If you are dealing with amino acids, you can use charges of the amino acids species from the forcefield database. Do not forget to label the charge groups appropriately.
14. Topology files (51) for the various frames are created using the code gmx4evb.py and the command below:

```
python gmx4evb.py -f 51 -r STO AAH WHO -p RRD AAW
```

15. The table for the softcore potentials should also be generated based on your beta value. This is done using the command:

```
gfortran gen_table.f90 -o gen_table.out
./gen_table.out > table_r1_hr.xvg
```

## II. Minimization And Equilibration Of System (Directory: Equilibration)

Successive Equilibration of the system is carried out using the topology file topo.000.top. This is done by running the script equil_protocol.sh for the various input files. This directory should also contain the tabulated potential files and the modified force field directory.

## III. Free Energy Perturbation (Directory FEP)

Copy the last equilibrated .gro file inside the equilibration folder, the tabulated potential files, all topology files, and the force field directory into the new directory. First run **restart.mdp** file which further equilibrates the system starting from randomized velocities - this way we can provide a different starting point to each FEP replica. Then submit the FEP frames sequentially, passing to each frame its corresponding topology (topol_000.top to fep_000.mdp, topol_001.top to fep_001.mdp and so on).

**NB: This protocol ended at this step. So, the github will not contain the results obtained for step III.**