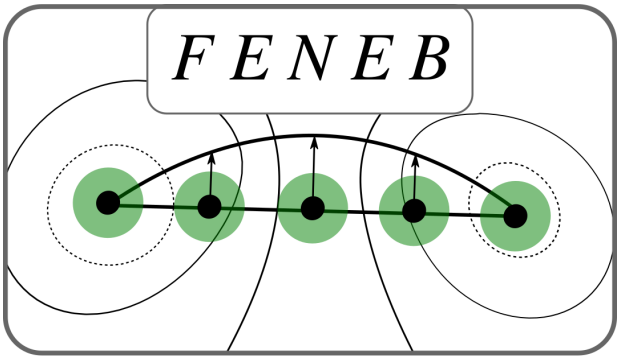


Minimum free energy pathways with nudged elastic bands

FENEB TUTORIAL

Jonathan A. Semelak
jsemelak@qi.fcen.uba.ar

---

1. What is FENEB

A tool for performing nudged elastic bands simulations on the free energy surface using Amber .nc files.

2. Requirements

This code requires NETCDF Libraries.

3. Compilation

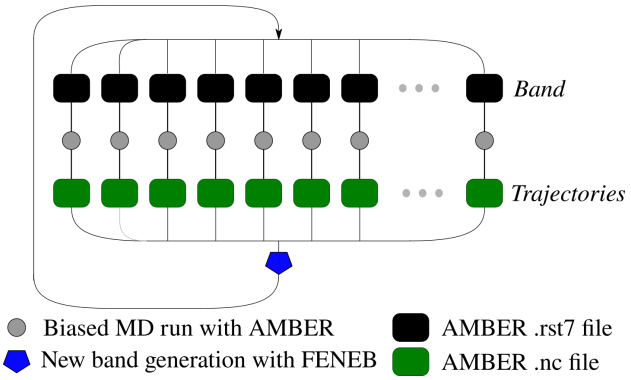
Execute the following script:

>>$ ./compile-feneb.sh

The location of the nf-config file is assumed to be \textit{/usr/bin/nf-config}. If that is not the case, the \textit{compile-feneb.sh} file should be edited.

4. Usage

Run molecular dynamics (MD) with AMBER, feed this code with .rst7 and .nc files, and you will get new .rst7 files to continue the optimization. Additionally, every time this code is executed, an output file will be generated with thermodynamic information.

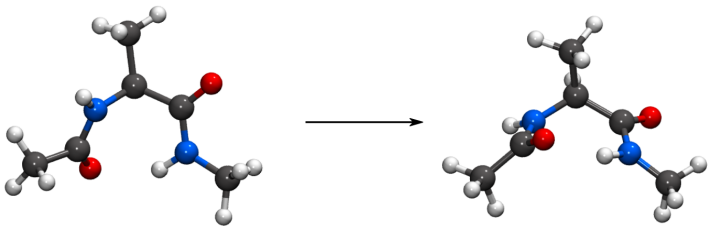


Knowledge of the AMBER package is assumed, and we recommend checking the following two references for theory and notation:

- Semelak, et. al. (2022). Minimum free energy pathways with the nudged elastic band method in combination with a QM-MM Hamiltonian. ChemRxiv. Cambridge: Cambridge Open Engage. This content is a preprint and has not been peer-reviewed.
- Bohner, et. al. (2014). Nudged-elastic band used to find reaction coordinates based on the free energy. The Journal of Chemical Physics, 140(7), 074109.

5. Tutorial

Here we will show how to optimize a minimum free energy pathway (MFEP) for the alanine dipeptide conformational change from the local minima C^{7eq} to C^{7ax}, using the FENEB method:



C^{7eq} C^{7ax}

We will use minimal inputs, following the simplest protocol to optimize a MFEP. Other capabilities of the code will be soon added to this tutorial (soon becoming a manual too).

5.1. Defining the free energy surface

The first step is to define free energy surface (FES) by selecting which atoms will integrate the *reaction coordinate space* (or RCS). The thumb rule for this is to choose all atoms involved in large geometrical modifications during the process. In the case of chemical reactions, atoms participating on the forming and breaking of bonds are recommended to be included on the RCS. In the case we address here, we will choose those atoms forming the Ramachandran angles of the alanine dipeptide.

According to the files from the **tutorial** folder, they correspond to indexes 5,7,9,15 and 17.

5.2 Optimizing reactants and products

This is not a chemical reaction, but we will refer to the band extrems as reactants and products for simplicity. Inside the **1-REACTANTS** folder, you will finde the following files:

ALAD.prmtop	This is the topology file
prod.mdin	This is an AMBER input for NVT biased MD
feneb.in	This is the input for this code
feopt_script.sh	This is a bash script to automate the FENEB run

Additionally, the folder **0** contains the initial structure for the reactants optimization, and the folder **1** contains the outputs (we will get to that point later).

Lets inspect the content of the prod.mdin and feneb.in files:

>>$ cat prod.mdin

```
 &cntrl
  imin=0,
  ntx=1,
  ntwr=100,
  ntpr=10,
  ntwx=10,
  ioutfm=1,
  ntxo=1,
  nstlim=10000,
  dt=0.001,
  ntt=3,
  tempi=300.0,
  temp0=300.0,
  ig=2022, !This is the seed. It should be set to random (-1), but here is set  to 2022 for reproducibility
  gamma_ln=5.0,
  ntp=0,
  ntb=0,
  cut=999.,
  ntr=1, !Turn on restraints on cartesian coordinates
  restraint_wt=50.0,  !Specify restrain force constant
  restraintmask='@5,7,9,15,17',  !Specify which atoms will be affected by the restraint
 &end
```

The prod.mdin file corresponds to a biased MD run in the NVT ensemble, with a restraint in the cartesian coordinates of the atoms specified in restraintmask. The restraintmask selection must be the same as the desired RCS.

>>$ cat feneb.in

```
infile ALAD_f
reffile ALAD_r
outfile ALAD_o
topfile ALAD.prmtop
per F
velin F
velout F
nrep 1
nrestr 5
mask 5 7 9 15 17
kref 100
skip 500
ftol 2.25
smartstep T
```

Where **infile**, **reffile**, and **outfile** correspond to the prefix for the initial coordinates, the reference coordinates, and the output (optimized) coordinates. Files will be named by the FENEB code as “ALAD_f_numberofimage.rst7”, etc. The **topfile** corresponds to the name of the topology file, **per, velin** and **velout** indicate whether the system is periodic, and whether the input and output files contain or must contain velocity info.
The **nrep** variable corresponds to the number of images the FENEB band contains.

IMPORTANT: When **nrep** is set to 1, the code automatically understands that we are only performing a free energy optimization, and nod a FENEB optimization.

The **nrest** and **mask** variables indicate how many and which atoms form the RCS. The kref variable is the force constant used in the biased MD (and must be twice the one from the prod.mdin file, because of how AMBER defines the restraint potential). The units are kcal/(mol $Å^2$)
The **skip** variable indicates how many steps from the MD trajectory will be omitted for the analysis (the first skip **steps** are omitted). The **ftol** variable indicates the convergence criterion, The optimization is assumed to be converged when the maximum (root mean square) atomic free energy gradient is lower than **ftol**. Finally, the **smartstep** variable indicates that the optimization algorithm will be a steepest descent one, in which the length of the step will be adjusted depending on the maximum gradient.

Now we are ready to execute the script, that basically will run MD, execute the FENEB code, run MD, execute the FENEB code, until convergence. As note of caution, in case the user wanted to modify the script, it is mandatory that the feneb.in file is present in the folder where FENEB is executed,as well as the topology file, and any coordinates and trajectories used in the AMBER run.

The script requires the user to first must modify the SANDER and FENEB, which indicate were AMBER and FENEB are installed.

Then, we execute it:

>>$ ./feopt_script.sh

In this case the code we will only execute a single step (notice the variable MAXSTEPS is set to 1 in the feopt_script.sh file), because that is enough to optimize the structure.

Several files will be generated, being the most important ones the **ALAD_o_1.rst7**, **ALAD_o_av_1.rst7** and **feneb.out** files.

ALAD_o_av_1.rst7	Mean coordinates obtained from the MD (only for RCS, the rest correspond to the initial coordinates). If the system is converged, here is where you want to look at.
ALAD_o_av_1.rst7	Optimized coordinates obtained by a single FENEB step. This coordinates are used as reference coordinates for the following optimization step.
feneb.out	This the output of FENEB.

The most important thing to look at the feneb.out file is the “Max force:” line, where the maximum atomic free energy gradient is printed. In this case, it is equal to 0.67 kcal/(mol Å), wich satisfies the 2.25 kcal/(mol Å) convergence criterion.

If more optimization steps were performed, the script would generate a folder for each of them, and every folder would contain the “same” files as folder **1** from this tutorial.

Since the initial structure actually possessed coordinates from the RCS that corresponded to a local minimum in our FES, we can take the **ALAD_o_av_1.rst7** file, change its name to **ALAD_r.rst7**, and continue our life.

>>$ mv ALAD_o_av_1.rst7 ALAD_r.rst7

Now we repeat the same with the products, working in the **2-PRODUCTS** folder. We will change the name of the new **ALAD_o_av_1.rst7** file, but this time to **ALAD_p.rst7**:

>>$ mv ALAD_o_av_1.rst7 ALAD_p.rst7

NOTE: FENEB always numbers as “1” the image of a free energy optimization. It does not know whether it corresponds to reactants or products.

5.3 Generating the initial band

The FENEB code includes a tool called BANDBUILDER which takes reactant and products coordinates, and generates the initial band. Reactants and products coordinates should be aligned (with VMD or CPPTRAJ, for instance), previous to executing BANDBUILDER. The folder **3-INITIALBAND** contains:

ALAD.prmtop	This is the topology file
ALAD_r.rst7	Reactants optimized coordinates
ALAD_p.rst7	Products optimized coordinates
bandbuilder.in	Input for BANDBUILDER
bandbuilder_script.sh	This is a bash script to run BANDBUILDER and generate a trajectory file for easy visualization
ALAD_BAND_0.nc	Trajectory file with the images generated.

Additionally, the folder **0** contains the coordinates of the images generated: **ALAD_r_1.rst7**, **ALAD_r_2.rst7**, **…**, **ALAD_r_25.rst7**.

The **bandbuilder.in** file syntax is similar to that of the feneb.in method:

>>$ cat bandbuilder.in

```
prefix ALAD_r
rcfile ALAD_r.rst7
pcfile ALAD_p.rst7
nrep 15
nrestr 5
```

```
mask 5 7 9 15 17
velin F
velout F
per F
```

Heere the only new variables are **rcfile** and **pcfile**, that correspond to the reactants and coordinates files. With this input file, BANDBUILDER will generate coordinate files for the band interpolating reactants and products RCS coordinates, and the rest of the coordinates will be those of the reactants files. The band will contain 15 images, according to the **nrep** variable.

5.4 FENEB Optimization

Now we are ready to optimize a MFEP for the process $C^{7eq} \rightarrow C^{7ax}$. We have an initial band consisting of 2 previously optimized extremes, and 13 structures ready to be optimized until they lay in a MFEP.

In the **4-REACTION** folder, we have both the inputs and the outputs of the FENEB optimization. Inputs are the **0** folder obtained in the section 5.3, which contains the initial band, plus the following files:

ALAD.prmtop	This is the topology file
prod.mdin	This is an AMBER input for NVT biased MD
feneb.in	This is the input for this code
feneb_script.sh	This is a bash script to automate the FENEB run

The prod.mdin file is identical to that of section 2, but with a force constant ten times higher. The feneb_script.sh coordinates the calls to AMBER and FENEB, and creates folders to save the outputs of each optimization step (give it a look!).

The feneb.in file is similar to that of the free energy optimizations:

>>$ cat feneb.in

```
infile ALAD_f
reffile ALAD_r
outfile ALAD_o
topfile ALAD.prmtop
per F
velin F
velout F
nrep 15
nrestr 5
mask 5 7 9 15 17
kref 1000
```

```
kspring 500
skip 500
ftol 2.25
smartstep T
```

Where the only new variable is **kspring**, which specifies the value for this constant. Additionally, we note again that the force constant is ten times higher than in the free energy optimizations of reactants and products. By default, the “uncoupled” version of FENEB is performed.

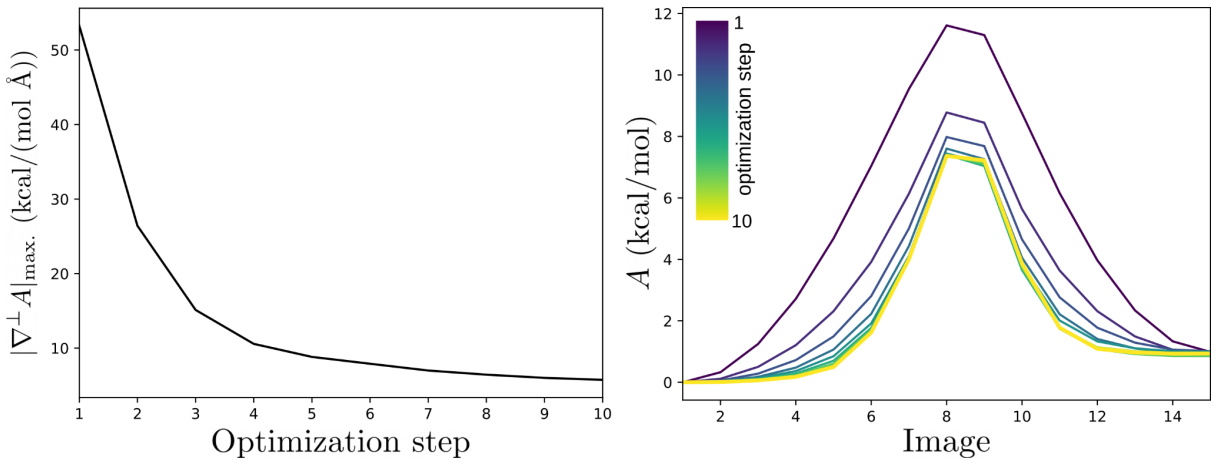
The **feneb_script.sh** requires the user to specify the location of SANDER and FENEB, as well as the maximum optimization steps to be performed. In this case, we perform 10 steps.

We are ready to execute the feneb_script.sh, which assumes that at least 15 cpu threads are available.

>>*$ ./f*eneb_script.sh

After a few seconds, 10 folders will be generated, each containing mean coordinates (“**_o_av_**” files), optimized coordinates (“**_o_**” files) and others. The feneb.out file contains information of the FENEB optimization step. The profile.dat file contains the free energy evolution along the corresponding band.

The feneb_script.sh will also grep useful information to the files **bandevolution.dat** and **maxgradevolution.dat.** The content of these files allows us to monitor the FENEB optimization process, and can be easily accessed with xmgrace, or a python script:



The obtained barrier is of 7.3 kcal/mol, which shows an excellent agreement with the 7.9 kcal/mol value that is obtained with longer simulation times and more optimization steps.