# 2. Free Energy Perturbation

Martin Almlöf, Martin Andér, Sinisa Bjelic, Jens Carlsson,
Hugo Gutiérrez de Terán, Martin Nervall, Stefan Trobro and Fredrik österberg

18th August 2005
updated:March 13, 2014

## Contents

# 1 FEP

## 1.1 Theoretical background

Free energy perturbation simulations, generally known as FEPs, can be used for theoretical prediction of binding energies. Direct calculation of free energies from the partition function generally does not work because of very slow convergence. Instead, the free energy is calculated as a difference between the two states, A and B, according to a formula derived by Zwanzig,

$$\Delta F = kT \ln < e^{-\Delta U_{BA}/kT} >_A \tag{1}$$

where $k$ is the Boltzmann constant, $T$ is the temperature in Kelvin and $U$ is the potential energy of the system. The notation $< \cdot >$ denotes an ensemble average. Eq. 1 is derived from the statistical mechanical expression for the free energy and the configurational integral, Z.

$$F = -kT \ln(Z) \tag{2}$$

$$Z = \int e^{-U/kT} d\Gamma \tag{3}$$

Depending on the ensemble (constant pressure or volume), either the Helmholtz free energy or the Gibbs energy are obtained, but for many applications the difference between5D these quantities is negligible, and we will denote the free energy by F. The free energy difference between the two different states A and B, that are represented with potential energies $U_A$ and $U_B$ respectively, can be rearranged into eq. 1 as follows

$$\begin{aligned}
\Delta F_{A\to B} &= -kT \ln(Z_B) - (-kT \ln(Z_A)) \\
&= -kT \ln(Z_B/Z_A) \\
&= -kT \ln(\int e^{-U_B/kT} d\Gamma / \int e^{-U_A/kT} d\Gamma) \\
&= -kT \ln(\int e^{-U_B/kT} e^{-U_A/kT} e^{U_A/kT} d\Gamma / \int e^{-U_A/kT} d\Gamma) \\
&= -kT \ln < e^{-\Delta U_{A\to B}/kT} >_A
\end{aligned} \tag{4}$$

FEP is very accurate concerning small perturbations, when the difference between the potential $U_A$ and $U_B$ is smaller than a few $kT$'s. The free energy of association of ligands to a protein is calculated with FEP as relative differences between the ligands. The absolute binding energy is defined as a free energy associated with moving the ligand from water into a solvated protein, legs II and IV in Figure 1. The relative binding energies are defined as a difference between two simulations where the ligand A is stepwise transformed to ligand B in water and in the solvated protein, legs I and III in Figure 1.

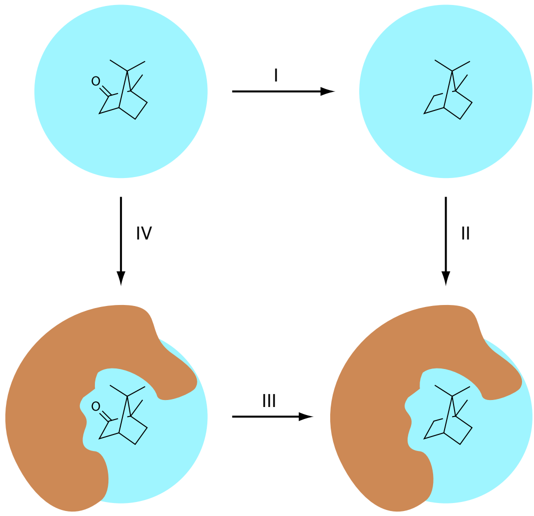


Figure 1: Thermodynamic cycle for ligand binding.

### 1.2 FEP simulations in `qdyn5`

Before starting the practical change the directory to `Protein` or `Water` in the `FEP` folder.

In this practical we are going to analyze the perturbation of camphane (CMA) to camphor (CAM) in water and in P450cam. The P450 enzymes catalyze the hydroxylation of the unactivated alkanes, and P450cam catalyzes specifically the hydroxylation of CAM. Although the system investigated here is P450cam the rules in this practical are general and can be applied to whatever system of interest.

A FEP simulation is carried out in `Q` with a `name.fep` file where the transformation is defined in detail. Every `name.fep` is divided in a number of sections where we define atoms, charges, bonds, angles etc that are changing during the FEP simulation (Figure 2). Note that a section in `Q` is defined by [...].

- Open the `cma_cam.fep`, in the `Protein` or `Water` folder, that is used for the CAM to CMA perturbation. Try to locate the different sections that are used in the FEP file. Can you understand how CMA is transformed to CAM?

- What effect do they have on the simulation? (HINT: For the detailed explanation of the different sections consult Appendix A!)

Having a FEP file the perturbation is defined in `Q` by adding a specifier

```
fep             name.fep
```

in the input files for `qdyn5`. Additionally it is necessary to add the section `[lambdas]` where the lambda values are defined, $\lambda_1$ and $\lambda_2$. The lambda values are used to transform the potential $U_A$ to $U_B$ in small steps to improve the convergence. The energies are sampled

```
[atoms]
...

[FEP]
...

[change_charges]
...

[atom_types]
...

.
.
.
```

Figure 2: FEP file format.

on the potential $U$,

$$U = \lambda_1 U_A + \lambda_2 U_B = \lambda_1 U_A + (1 - \lambda_1) U_B \tag{5}$$

where $\lambda_1$ varies between 0 and 1. How many lambda steps are used in the simulation depends on difference between $U_A$ and $U_B$ and the type of perturbation.

- Open an input file `name.inp`. In what section is the fep file read?

- Find the `[lambdas]` section. How many lambda steps are used in the perturbation and how big is each step? (HINT: Look in several `cma_camN.inp` files, where `N`=0..30.)

- How many steps are used in the simulation in each file?

- Also note the specifiers

```
energy          name.en
energy          25
```

that are used for saving the energies every 25th step to the file `name.en`.

## 1.3 Analysis of FEP simulations by **`qfep5`**

The results from the FEP simulations, which take several hours to run, are analyzed in `qfep5`. `qfep5` uses energies saved in the `name.en` files to calculate the free energy difference between the states. Do the following for both water and protein simulations:

- Type `qfep5 < qfep.inp > qfep.out`. The `qfep.inp` file contains all necessary parameters that are loaded in `qfep5` and the `qfep.out` file is the output summary.

- Open `qfep.inp` and `qfep.out`. Try to understand the different commands. What are they specifying? (HINT: Check out Appendix B!)

- The free energy difference for the perturbation from CMA to CAM can be found at the end of table summarized in `Part 1`. What is the difference in water and in the protein?

Water simulation: $\Delta G^{w}_{A\to B} =$ (kcal/mol)

Protein simulation: $\Delta G^{p}_{A\to B} =$ (kcal/mol)

The free energy is calculated as an average of forward and backward calculations at every lambda point. The forward and backward free energy values are summarized also in `Part 1` of the `qfep.out` file in columns `sum(dGf)` and `sum(dGr)`. The theoretical error of the FEP simulation is half the value of the difference between forward and backward calculations.

-What is the theoretical error of FEP for water and protein simulations?

Water simulation: $\Delta G^{w,error}_{A\to B} =$ (kcal/mol)

Protein simulation: $\Delta G^{p,error}_{A\to B} =$ (kcal/mol)

The relative free energy of binding is determined as the difference between the free energies for the protein and water simulations.

Relative binding free energy:

Calculated: $\Delta\Delta G^{calc}_{bind,rel} = \Delta G^{p}_{A\to B} - \Delta G^{w}_{A\to B} =$ (kcal/mol)

Experimental: $\Delta\Delta G^{exp}_{bind,rel} = \Delta G^{p}_{A\to B} - \Delta G^{w}_{A\to B} =$ -2.0 (kcal/mol)

Alternatively the free energy profiles can be plotted and investigated graphically. The program `gnuplot` is used for plotting different graphs.

- Open `gnuplot` by typing `gnuplot` in the shell window and then write `load 'fep.part1.pgp'`. This script plots the free energy as a function of lambda, as summarized in the table in `Part 1` of `qfep.out`.

- What is the shape of the curve? What is the spacing between the points? Does it look ok?

Molecular dynamics simulations also gives us an opportunity to investigate the structures as they are propagated through time.

- Type `pymol` and open `cma_cam.pse` which loads `cma_cam0.pdb`, `cma_cam10.pdb` and `cma_cam30.pdb` files in the viewing program `PYMOL`. The `name.pdb` files represent structures at the lambda state (1,0), (0.63,0.37) and (0,1) respectively.

- What is the difference between the different ligand structures?

- What are the interactions between the ligands and the surrounding. (HINT: Check out the Tyr87 residue!)

Here ends the FEP part of the lab course!

# A FEP file format

Table 1: FEP file format

[**atoms**]: Define Q-atoms.

| **column** | **description** |
|---|---|
| 1 | Q-atom number (counting from 1 up). |
| 2 | Topology atom number. |

[**PBC**]: For periodic boundary conditions.

| **keyword** | **value** | **comment** |
|---|---|---|
| switching-_atom | Topology atom number. | Required with periodic boundary conditions. |

[**FEP**]: General perturbation information.

| **keyword** | **value** | **comment** |
|---|---|---|
| states | Number of FEP/EVB states. | Optional, default 1. |
| offset | Topology atom number. | Optional, default 0. This number is added to all topology atom numbers given in the FEP file. |
| offset_residue | Residue/fragment number. | Optional. Set offset to the topology number of the first atom in the given residue minus one. |
| offset_name | Residue/fragment name. | Optional. Set offset to the topology number of the first atom in the first residue with the given name minus one. |
| qq_use-_library-_charges | This is a special feature for studying $e.g.$ electrostatic linear response. Set to 'on' to use the library charges from the topology for intra-Q-atom interactions, i. e. change only Q-atom-surrounding electrostatic interactions. | Optional, default off. |
| softcore-_use_max-_potential | Set to 'on' if the values entered in the [**softcore**] section are the desired maximum potentials (kcal/mol) at $r = 0$. Qdyn will then calculate pairwise $\alpha_{ij}$ to be used in equation 6. 'off' means the values are to be used directly in equation 6. | Optional, default off. |

[**change_charges**]: Redefine charges of Q-atoms.

| **column** | **description** |
|---|---|
| 1 | Q-atom number (referring to numbering in atoms section). |
| 2... | Charge (e) in state 1, state 2, ... |

Table 1: FEP file format

[**atom_types**]: Define new atom types for Q-atoms: Standard LJ parameters and parameters for the exponential repulsion potential $U_{soft} = C_i \cdot C_j \epsilon^{-a_i \cdot a_j \cdot r_{i,j})}$.

| | |
|---|---|
| 1 | Name (max 8 characters). |
| 2 | Lennard-Jones A parameter (kcal$^{\frac{1}{2}}$·mol$^{-\frac{1}{2}}$·Å$^{-6}$) for geometric combination or R* (kcal·mol$^{-1}$·Å$^{-12}$)for arithmetic combination rule. |
| 3 | LJ B parameter (kcal$^{\frac{1}{2}}$·mol$^{-\frac{1}{2}}$·Å$^{-3}$) or $\epsilon$ (kcal·mol$^{-1}$·Å$^{-6}$). |
| 4 | Soft repulsion force constant $C_i$ (kcal$^{\frac{1}{2}}$·mol$^{-\frac{1}{2}}$) in $U_{soft}$. |
| 5 | Soft repulsion distance dependence parameter $a_i$ (Å$^{-\frac{1}{2}}$) in $U_{soft}$. |
| 6 | Lennard-Jones A parameter (kcal$^{\frac{1}{2}}$·mol$^{-\frac{1}{2}}$·Å$^{-6}$) or R* (kcal·mol$^{-1}$·Å$^{-12}$) for 1-4 interactions. |
| 7 | LJ B parameter (kcal$^{\frac{1}{2}}$·mol$^{-\frac{1}{2}}$·Å$^{-3}$) or e (kcal·mol$^{-1}$·Å$^{-6}$) for 1-4 interactions. |
| 8 | Atomic mass (u). |

[**change_atoms**]: Assign Q-atom types to Q-atoms.

| | |
|---|---|
| 1 | Q-atom number. |
| 2... | Q-atom type name in state 1, state 2, ... |

[**soft_pairs**]: Define pairs which use soft repulsion.

| | |
|---|---|
| 1 | Q-atom number of first atom in pair. |
| 2 | Q-atom number of second atom in pair. |

[**excluded_pairs**]: Define pairs to exclude from non-bonded interactions. Note: also non-Q-atoms can be excluded.

| | |
|---|---|
| 1 | Topology atom number of first atom in pair. |
| 2 | Topology atom number of second atom in pair. |
| 3... | Exclusion effective (1) or not (0) in state 1, state 2, ... |

[**el_scale**]: Define q-atom pairs for scaling of the electrostatic interaction. Can be useful e.g. when highly charged intermediates appear in FEP/EVB. The scale factor applies to all states. Note: only Q-atom pairs can be scaled.

| | |
|---|---|
| 1 | q-atom number of first atom in pair |
| 2 | q-atom number of second atom in pair |
| 3 | electrostatic scale factor (0..1) |

[**softcore**]: Define q-atom softcore potentials. The meaning of these entries depends on the value of softcore_use_max_potential.

| | |
|---|---|
| 1 | q-atom number |
| 2... | Desired potential at $r = 0$ for all of this q-atom's vdW interactions in state 1, state 2, ... or the actual $\alpha$ value used in equation 6. An $\alpha$ of 200 yields vdW potentials at $r = 0$ of 10-50 kcal/mol for heavy atom - heavy atom interactions. Set to 0 if softcore is not desired for this q-atom. |

Table 1: FEP file format

[**monitor_groups**]: Define atom groups whose non-bonded interactions are to be monitored (printed in the log file).

| | |
|---|---|
| 1... | Topology atom number of first and following atoms in group. |

[**monitor_group_pairs**]: Define pairs of monitor_groups whose total non-bonded interactions should be calculated.

| | |
|---|---|
| 1 | First monitor_group number. |
| 2 | Second monitor_group number. |

[**bond_types**]: Define Q-bond types using Morse or harmonic potentials,
$E_{Morse} = D_e \left(1 - e^{-\alpha(r-r_0)}\right)^2$ $E_{Harmonic} = \frac{1}{2} k_b \left(r - r_0\right)^2$.
Morse and harmonic potentials can be mixed (but each bond type is either kind). Entries with four values are Morse potentials and entries with three values are harmonic.

| | **Morse potential** | **Harmonic potential** |
|---|---|---|
| 1 | Q-bond type number (starting with 1). | |
| 2 | Morse potential dissociation energy, $D_e$ (kcal·mol$^{-1}$). | Harmonic force constant $k_b$ (kcal·mol$^{-1}$·Å$^{-2}$). |
| 3 | Exponential co-efficient $\alpha$ in Morse potential (Å$^{-2}$). | Equilibrium bond length $r_0$ in harmonic potential (Å). |
| 4 | Equilibrium bond length $r_0$ in Morse potential (Å). | |

[**change_bonds**]: Assign Q-bond types. Note: shake constraints for the redefined bonds are removed. The order in which atoms are given is not important.

| | |
|---|---|
| 1 | Topology atom number of first atom in bond. |
| 2 | Topology atom number of second atom in bond. |
| 3... | Q-bond type number (referring to numbering in bond_types section) or 0 to disable bond in state 1, state 2, ... |

[**angle_types**]: Define Q-angle types.

| | |
|---|---|
| 1 | Q-angle type number (starting with 1). |
| 2 | Harmonic force constant (kcal·mol$^{-1}$·rad$^{-2}$). |
| 3 | Equilibrium angle (°). |

[**change_angles**]: Assign Q-angle types.

| | |
|---|---|
| 1 | Topology atom number of first atom in angle. |
| 2 | Topology atom number of middle atom in angle. |
| 3 | Topology atom number of third atom in angle. |
| 4... | Q-angle type number (referring to numbering in angle_types section) or 0 to disable angle in state 1, state 2, ... |

[**torsion_types**]: Define Q-torsion types.

| | |
|---|---|
| 1 | Q-torsion type number (starting with 1). |
| 2 | Force constant = $\frac{1}{2}$·barrier height (kcal·mol$^{-1}$). |
| 3 | Periodicity (number of maxima per turn). |

Table 1: FEP file format

| | |
|---|---|
| 4 | Phase shift (°). |

[**change_torsions**]: Assign Q-torsion types. Note: The order of atoms (1, 2, 3, 4 or 4, 3, 2, 1) is not important.

| | |
|---|---|
| 1 | Topology atom number of first atom in torsion. |
| 2 | Topology atom number of second atom in torsion. |
| 3 | Topology atom number of third atom in torsion. |
| 4 | Topology atom number of fourth atom in torsion. |
| 5... | Q-torsion type number (referring to numbering in torsion_types section) or 0 to disable torsion in state 1, state 2, ... |

[**improper_types**]: Define Q-improper types.

| | |
|---|---|
| 1 | Q-improper type number (starting with 1). |
| 2 | Harmonic force constant (kcal·mol$^{-1}$·rad$^{-2}$). |
| 3 | Equilibrium angle (°). |

[**change_impropers**]: Assign Q-improper types. Note: The order of atoms (1, 2, 3, 4 or 4, 3, 2, 1) is not important.

| | |
|---|---|
| 1 | Topology atom number of first atom in improper. |
| 2 | Topology atom number of second atom in improper. |
| 3 | Topology atom number of third atom in improper. |
| 4 | Topology atom number of fourth atom in improper. |
| 5... | Q-improper type number (referring to numbering in improper_types section) or 0 to disable improper in state 1, state 2, ... |

[**angle_couplings**]: Couple Q-angles to Q-bonds, $i.e.$ scale angle energy by the ratio of the actual value of the Morse bond energy to the dissociation energy.

| | |
|---|---|
| 1 | Q-angle number (line number within change_angles section). |
| 2 | Q-bond number (line number within change_bonds section). |

[**torsion_couplings**]: Couple Q-torsions to Q-bonds.

| | |
|---|---|
| 1 | Q-torsion number (line number within change_torsions section). |
| 2 | Q-bond number (line number within change_bonds section). |

[**improper_couplings**]: Couple Q-impropers to Q-bonds.

| | |
|---|---|
| 1 | Q-improper number (line number within change_impropers section). |
| 2 | Q-bond number (line number within change_bonds section). |

[**shake_constraints**]: Define extra shake constraints. The effective constraint distance will be the sum of the distances given for each state, weighted by their l values. Note: constraints defined here do not override constraints imposed by setting the shake flag to *on* in the Qdyn input file. To remove a constraint the bond must be redefined as a Q-bond. The order in which atoms are given is not important.

| | |
|---|---|
| 1 | Topology atom number of first atom. |
| 2 | Topology atom number of second atom. |

Table 1: FEP file format

| | |
|---|---|
| 3... | Constraint distance (Å) in state 1, state 2, ... |

[**off-diagonals**]: Define off-diagonal elements of the Hamiltonian, represented by $H_{i,j} = A_{i,j} \cdot \epsilon^{-\mu_{i,j} \cdot r_{k,l}}$ where i and j are states and k and l are Q-atoms.

| | |
|---|---|
| 1 | State i. |
| 2 | State j. |
| 3 | Q-atom k. |
| 4 | Q-atom l. |
| 5 | $A_{i,j}$ (kcal·mol$^{-1}$). |
| 6 | $\mu_{i,j}$ (Å$^{-1}$). |

Softcore equation:

$$U_{vdW}(r) = \frac{A_{ij}}{(r^6 + \alpha_{ij})^2} - \frac{B_{ij}}{r^6 + \alpha_{ij}} \qquad \text{or} \qquad U_{vdW}(r) = \epsilon \cdot (\frac{{R_{ij}^*}^{12}}{(r^6 + \alpha_{ij})^2} - 2 \cdot \frac{{R_{ij}^*}^{6}}{r^6 + \alpha_{ij}}) \tag{6}$$

# B qfep5 input summary

Some of the commands are affecting only analysis of the empirical valance bond (EVB) simulations, while others are common for both FEP and EVB. The ones affecting FEP are marked in bold text and are of concern in FEP simulations.
(HINT: –> specifies the input commands and # denotes the written output.)

```
->Number of energy files:
# Number of files = 31
-> No.  of states, no.  of predefined off-diag elements:
# Number of states = 2
# Number of off-diagonal elements = 0
-> Give kT & no, of pts to skip:
# kT = 0.596
# Number of data points to skip = 80
-> Give number of gap-bins:
# Number of gap-bins = 40
-> Give minimum # pts/bin:
# Minimum number of points per bin= 10
-> Give alpha for state 2:
# Alpha for state 2 = 0.00
-> Number of off-diagonal elements:
# Number of off-diagonal elements = 0
-> linear combination of states defining reaction coord:
# Linear combination co-efficients= 1.00 0.00
name1.en
```

```
name2.en
.
.
nameN.en
```

The `# Number of files` is the total number of files used for the FEP simulation, `# Number of data points to skip` is the number of points that are discarded as the equilibration at each lambda step and `# kT` specifies the temperature of the simulation. The energy files `name.en` are read last in `qfep5`.