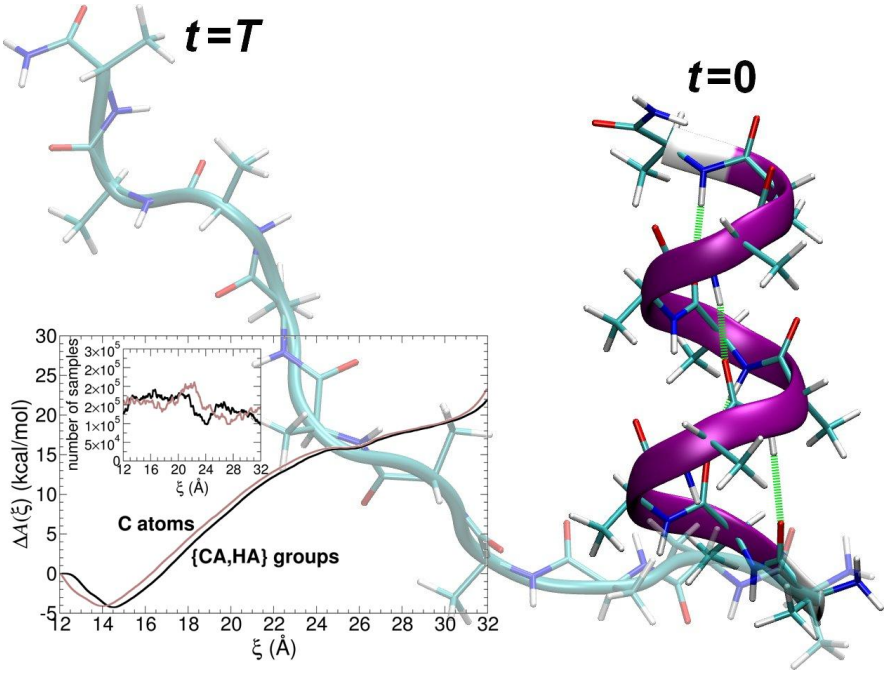
Université de Lorraine
Centre National de la Recherche Scientifique
Laboratoire International Associé CNRS-University of Illinois

Centre National de la Recherche Scientifique
Institut de Biologie Physico-Chimique

University of Illinois at Urbana-Champaign
Beckman Institute for Advanced Science and Technology
Theoretical and Computational Biophysics Group

Free energy calculations along a reaction coordinate: A tutorial for adaptive biasing force simulations



Jérôme Hénin
James Gumbart
Christophe Chipot

September 21, 2017

Current editor: Lela Vuković (Lvukov1@ks.uiuc.edu)

Please visit www.ks.uiuc.edu/Training/Tutorials/ to get the latest version of this tutorial, to obtain more tutorials like this one, or to join the `tutorial-l@ks.uiuc.edu` mailing list for additional help.

Abstract

The purpose of this tutorial is to get the NAMD user familiarized with the calculation of free energy along a reaction coordinate through a variety of applications of the adaptive biasing force (ABF) method. The reversible unfolding of deca-alanine serves as an introductory example, wherein the reaction coordinate is defined as a simple end-to-end distance in the peptide chain. The ABF scheme is then applied to the transfer of a methane molecule across the water liquid-vapor interface to estimate its hydration free energy. The result is then compared to free energy perturbation calculations. In the next example, the Ramachandran, (φ, ψ) free-energy map of N–methyl–N'–acetylalanylamide is constructed using both ABF and metadynamics with two coupled torsional variables. Last, the reversible association of a chloride and a sodium ion in water is examined. Prior knowledge of NAMD and standard molecular dynamics simulations is assumed.

Contents

Introduction

The goal of this tutorial is to familiarize the NAMD user with free energy calculations along a chosen reaction coordinate and provide guidance when setting up these simulations within NAMD. [1,2] Determination of free energy differences along the reaction coordinate, ξ, is achieved employing the adaptive biasing force (ABF) method [3], in its NAMD formulation and implementation, [4]. ABF is based on the computation of the mean force along ξ, which is then canceled out by an equal and opposite biasing force, allowing the system overcome barriers and escape from minima of the free energy landscape. Ultimately, the dynamics of ξ corresponds to a random walk with zero mean force, and only the fluctuating part of the instantaneous force exerted along ξ remains. Virtual erasure of the roughness of the free energy landscape yields a uniform sampling along ξ.

The ABF scheme assumes that the reaction coordinate, ξ, is *fully* unconstrained. This implies that in the course of the simulation, the complete reaction pathway discretized in small bins of width $\delta\xi$ will be explored in a continuous fashion. Samples of the instantaneous force acting along ξ are accrued in the different bins until a user-defined threshold is attained, beyond which the adaptive biasing force will be applied. Although ABF eventually guarantees uniform sampling along ξ, convergence properties of the free energy calculation can be strongly affected by the coupling of ξ to other, slowly relaxing degrees of freedom. Under such circumstances, a larger number of force samples ought to be accumulated prior to applying the adaptive biasing force onto ξ.

The concept of *unconstrained* reaction coordinate imposes that ξ be fully decoupled from any other constrained degrees of freedom [5] — *viz.* by means of the SHAKE or RATTLE algorithm [6,7].

The reader of this tutorial is assumed to be familiar with the use of NAMD to perform "standard" calculations, including energy minimization and MD simulations. General documentation, tutorials and templates of NAMD configuration files are available from the Documentation section of the NAMD web page.

ABF is implemented as part of the "collective variable calculations" (colvars) module [8] of NAMD. The colvars module is extensively documented in the NAMD user's guide. Other information about the ABF method can be found on a dedicated website, `http:// www.edam.uhp-nancy.fr/ABF`. A basic working knowledge of VMD is highly recommended.

Completion of this tutorial requires:

- various files contained in the archive `ABF_tutorial.zip`, provided with this document;
- NAMD 2.9 or later (`http://ks.uiuc.edu/Research/namd`);
- VMD 1.8.7 or later (`http://ks.uiuc.edu/Research/vmd`; using the latest version is recommended).

Many of the simulations in this tutorial can, out of necessity, take a significant amount of time to run on a single processor. If this time becomes prohibitive to completing the tutorial, example output is provided throughout to enable the reader to still carry out the desired analysis.

1. Reversible unfolding of deca-alanine

In this prototypical example, the equilibrium unfolding of a short, ten-residue peptide chain will be investigated in the gas phase by means of the ABF method. Initially organized in a compact α-helical structure, deca-alanine will be pulled reversibly into an ensemble of extended conformers. The main objective of this section is to demonstrate the efficiency of the free energy method, while emphasizing the complete reversibility of the transformation, which indicates that the simulation proceeds under near-equilibrium conditions. This problem has been used as a test-case for the early NAMD implementation of ABF [5, 9], and studied in greater detail in later work [4].

1.1. System setup

A structure file (PSF) and a coordinate file (PDB) describing the α-helical conformation of deca-alanine are provided, so that the reader may get started quickly with ABF simulations.

The peptide sequence consists of ten L-alanine amino acids, with an unblocked (yet unprotonated) N-terminus — *i.e.* ending with an $-NH_2$ moiety, and blocked at the C-terminus by an $-NH_2$ moiety, forming an amide group. Note that charged terminal groups (more typical of hydrated peptides) would interact strongly due to the lack of electrostatic screening in such a gas-phase simulation, and alter the results significantly.

To appreciate the role played by the choice of the reaction coordinate on the resulting free energy profile, use will be made of two distinct definitions of ξ, *i.e.* two functional forms for the collective variable `colvar`:

- `AtomDistance`: ξ corresponds to a simple interatomic distance between the carbon atoms pertaining to the first and the last carbonyl moieties of the peptide chain. This definition is compatible with the constraint of all hydrogens' degrees of freedom in deca-alanine — *viz.* `rigidbonds all`, ξ being fully decoupled from frozen chemical bonds.

- `COMDistance`: ξ corresponds to the separation between two groups of atoms formed by the α-carbon atom and its aliphatic hydrogen in the first and the last residues of the peptide chain. This definition is also compatible with the constraint of all hydrogens' degrees of freedom (`rigidbonds all`), because it involves the center of mass of the constrained `CA`–`HA` bond, rather than individual atoms. While bond constraints exert forces on individual atoms, they do not affect the center of mass of the bonded pair.

1.2. Running the adaptive biasing force calculation

Now that the structure (PSF) file is ready to use and that a reaction coordinate has been defined, let us detail how the free energy calculations will be carried out with NAMD, in particular how the ABF aspect of the simulations will be set up. The files given specify two 2.5-ns simulations, one initiating the calculations and one restarting from the end of the first, for 5 ns total. Since the system is very small, consisting only of 104 atoms, the ABF simulations run quickly on a single processor (about 30-60 minutes each). If this is too long, however, see the example output for what one would obtain from running them to completion.

The traditional MD section of the provided NAMD configuration files requests an MD run at a constant temperature of 300 K. A smooth spherical cut-off of 12 Å (shifted above 10 Å) is used to handle non-bonded forces. A time step of 0.5 fs is appropriate for gas phase simulations. The NAMD configuration file also contains statements that trigger the initialization of the colvars module:

```
colvars on
colvarsConfig Distance.in
```

If ξ corresponds to the distance separating the carbon atoms of the first and the last carbonyl moieties, the colvars configuration file (`Distance.in`) should read:

```
colvarsTrajFrequency      2000
colvarsRestartFrequency  20000

colvar {
   name AtomDistance

   width 0.1

   lowerboundary 12.0
   upperboundary 32.0

   lowerwallconstant 10.0
   upperwallconstant 10.0

   distance {
      group1 {
         atomnumbers { 10 }
      }
      group2 {
         atomnumbers { 92 }
      }
   }
}

abf {
   colvars AtomDistance
   fullSamples   500
   hideJacobian
}
```

Most of the file defines a colvar named `AtomDistance`, while the two-line block at the end sets up an ABF calculation on this variable. Because we anticipate that the free energy can vary rather rapidly, we purposely set the bin width $\delta\xi$ for `AtomDistance` to 0.1 Å. The colvar has a range of allowed values from 12 to 32 Å, and boundary potentials with a force constant of 10 kcal/mol/Å^2. In the gas phase and given the nature of the molecular system, we further expect relaxation to be rather fast, so that a threshold of 500 force samples should be enough to obtain a reasonable estimate of the force distribution, $\varrho(F_\xi)$, and, hence, of the average force, $\langle F_\xi \rangle_\xi$, which will be subsequently applied along the reaction coordinate.

In many cases, it is useful to define coordinates involving groups of atoms - for example, grouping a

heavy atom with its bound hydrogen atoms prevents constraint forces applied to the bond from affecting the ABF calculation. In the case of deca-alanine, ξ can be chosen as the distance separating the first and the last CA–HA groups of the peptide chain, by replacing the `Distance.in` file above with the following:

```
Colvarstrajfrequency     2000
Colvarsrestartfrequency 20000

colvar {
   name COMDistance

   width 0.1

   lowerboundary 12.0
   upperboundary 32.0

   lowerwallconstant 10.0
   upperwallconstant 10.0

   distance {
      group1 {
         atomnumbers {  4   5 }
      }
      group2  {
         atomnumbers { 99 100 }
      }
   }
}

abf {
  colvars COMDistance
  fullSamples      500
   hideJacobian
}
```

1.3. Analyzing the results

ABF produces the following output files:

- `<outputName>.grad`: current estimate of the free energy gradient (grid), in multicolumn;
- `<outputName>.count`: histogram of samples collected, on the same grid;

- `<outputName>.pmf`: only for one-dimensional calculations, integrated free energy profile or PMF.

The data is in text format. It can be read with any plain text editor, and plotted with a variety of software.

Convergence of the free energy calculations usually occurs within 5 ns, although it is recommended to extend sampling further to ascertain that the potential of mean force does not vary any more. You can compare the obtained free energy profile with those of Figure 1, and note that ABF results agree well with the free energy differences measured by Park *et al.* [10] in their pulling experiments with steered MD. [11]

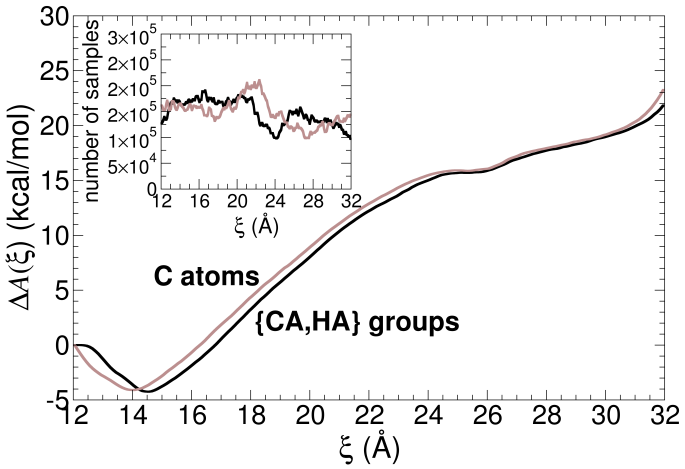


Figure 1: Free energy profile delineating the reversible unfolding of deca-alanine in the gas phase, using two distinct reaction coordinates.

The most natural criterion for probing the convergence of the simulation is the evolution of the free energy profile. Reversibility of the transformation is another key aspect of ABF simulations. Employing VMD, it can be verified that not only is sampling of the reaction coordinate uniform, but that the different conformational states of the peptide chain are visited frequently. This can be done, for instance, by plotting ξ as a function of time. The latter provides a rough idea of how reversible the conformational transition is. The analysis can be further refined by measuring the time-evolution of the intramolecular $i \rightarrow i+4$ hydrogen bonds involved in the formation of the α-helix. [5]

2. Hydration free energy of methane

In the second example of this tutorial, the hydration free energy of methane will be determined following two different routes — *viz.* (i) based on the free energy derivative with respect to an order parameter, *i.e.* ABF, and (ii) using free energy perturbation (FEP). It will be shown that these two routes are equivalent and yield the same answer at the quantitative level.

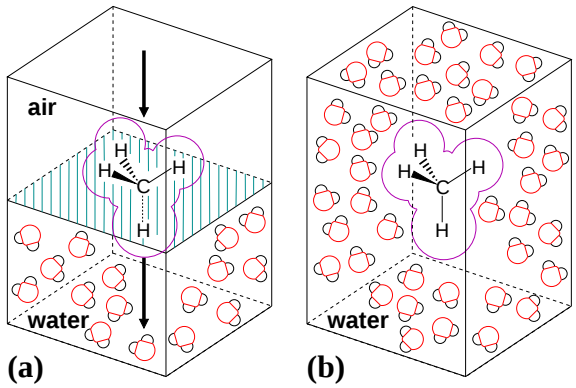


Figure 2: Two pathways for calculating the hydration free energy of methane.

Hydration of methane can be described by either of the processes depicted in Figure 2. The horizontal transformation corresponds to the *direct* hydration process, wherein the solute is brought from the gas phase to the aqueous medium. This process can be modeled using a water liquid-vapor interfacial system, consisting of a water lamella in equilibrium with its low pressure gaseous phase (on a microscopic scale, essentially vacuum). Methane is placed in the aqueous environment, and is translocated in a reversible fashion across the Gibbs dividing surface, into the gas phase. ABF can be employed to serve this purpose, choosing as a reaction coordinate the projection along the z-direction of Cartesian space of the distance separating the center of mass of water from that of methane.

If a standard state is chosen where methane has the same concentration in the gas and aqueous phases, then the standard hydration free energy of methane is the difference between the two plateau values of the transfer free energy profile, corresponding to either homogeneous phase far enough from the interface.

Standard state corrections for comparison with experiment

A common convention in physical chemistry would set the standard state to be a methane partial pressure of 1 bar in the gas phase, and a 1 mol/L concentration in the water solution. Computing standard free energies under such conventions requires a "standard state correction" due to the fact that standard volume concentrations are not the same in both phases.

Alternately, computation on the hydration free energy can be based on decoupling of methane from bulk water (Figure 2b), This transformation can be performed using the alchemical FEP methodology, [12] whereby interactions between methane and water are "turned off", and the resulting free energy change is calculated. A stratification strategy is used, which consists in connecting the reference and target states in a series of stages. [13] If use is made of $N-2$ intermediate states, it follows that the net free energy change is:

$$\Delta G = \sum_{k=1}^{N-1} \Delta G_{k\to k+1} = -\frac{1}{\beta} \sum_{k=1}^{N-1} \ln\langle \exp(-\beta \Delta U_{k\to k+1})\rangle_k \tag{1}$$

$\beta = (k_B T)^{-1}$, where k_B is the Boltzmann constant and T is the temperature. $\Delta U_{k\to k+1}$ denotes the change in potential energy between state k and state $k+1$. $\langle \cdots \rangle_k$ represents an ensemble average over configurations of the reference, k state.

Here, the free energy of hydration is simply the opposite of the decoupling free energy as computed by FEP. Again, we assume standard states based on equal volume concentrations in both phases, otherwise correcting for the standard concentration would be necessary.

2.1. System setup

The goal of this section is to build two systems:

- **FEP calculation**: A cubic box of water, with the methane molecule immersed in it.
- **ABF calculation**: An interfacial system consisting of a water lamella containing a methane molecule, separated from its periodic image in the z direction by a water vapor-filled space (in practice, essentially vacuum).

2.1.1. Generating the PSF file

The Cartesian coordinates of an isolated methane molecule are provided in the archive, *viz.* `methane.pdb`. Using `psfgen` with `top_methane.inp`, which contains a topology for methane, build the structure PSF file for the isolated solute, forming a segment named `METH`.

- **Bulk system for FEP**: Starting from the structure PSF file of methane, use the `Add Solvation Box` feature of VMD to hydrate the molecule. Make a cubic box with 30 Å sides (-15 to 15 in each dimension). The VMD plugin generates PDB and PSF files that correspond to the complete system, *i.e.* the methane molecule immersed in water.

- **Tagging water oxygen for the ABF calculation**: In VMD, mark water oxygen atoms by setting their beta value to 1:

  ```
  [atomselect top "name OH2"] set beta 1
  ```

  Then save the coordinates in PDB format as `reference.pdb`.

- **Interfacial system for ABF**: The same bulk system can be used as a starting point for the interfacial calculation simply by setting different periodic boundary conditions. If the third basis vector of the periodic lattice is set to be 60 Å along the z axis, it creates a gap between the water box and its periodic image, which will play the role of the gas phase.

Visual inspection using VMD is highly recommended prior to proceed with a short energy minimization (a few hundred steps) and thermalization (100 ps) of the molecular assemblies.

2.1.2. Preparing the NAMD configuration file

- **ABF calculation**: As has been mentioned previously, a simple and adequate choice for a reaction coordinate is `distanceZ`, the projected distance on the z-axis of two groups of atoms — `ref` corresponding to the aqueous medium and `main` to the solute.

  In practice, the adaptive biasing force is applied along ξ on methane, while its counterpart is distributed over all water molecules forming the water lamella. In the present example, `ref` gathers the indices of all oxygen atoms of the water phase.

  Since the expected free energy profile in this case is rather smooth, a bin width $d\xi$ of 0.2 Å is small enough.

```
Colvarstrajfrequency    100
Colvarsrestartfrequency 1000

colvar {
   name ProjectionZ

   width 0.2

   lowerboundary 0.0
   upperboundary 5.0

   lowerwallconstant 10.0
   upperwallconstant 10.0

   distanceZ {
      ref {
         atomsFile reference.pdb
         atomsCol B
      }
      main  {
         atomnumbers { 1 2 3 4 5 }
      }
   }
}

abf {
  colvars ProjectionZ
  fullSamples      500
}
```

To enhance the efficiency of the ABF simulation, the reaction pathway spanning approximately 25 Å — *i.e.* from the "isolated" state in the gas phase to the hydrated state in the midst of the water lamella, will be broken down into five consecutive 5-Å wide windows. Each of these five simulations should start from a configuration in which the coordinate z is in the relevant interval. A way to prepare such starting points is to run a short ABF calculation where z can span the full range (0 to 25 Å), and use VMD to select appropriate configurations from that trajectory. Chosen frames can then be saved in PDB format using the *File/Save Coordinates* option.

Standard MD options should be pasted in the NAMD configuration file for an MD simulation in the canonical ensemble. The temperature will be maintained at 300 K using Langevin dynamics. Long-range electrostatic forces will be handled by means of the particle-mesh Ewald (PME) algorithm. The `rigidBonds` option will be set to `all` and the equations of motion will be integrated with a time step of 1.0 fs.

- **FEP calculation**:The hydration free energy of methane will be determined by decoupling the solute from the aqueous medium. Within NAMD, such a free energy calculation supposes that an `alchFile` has been defined preliminarily. The latter is a clone of the PDB file, wherein the `B` (*beta*) column is set to -1 to flag atoms that are being decoupled, i.e. those of methane). Owing to the nature of the perturbation, a stratification strategy will be employed to connect the reference state of the transformation (fully coupled) to its target state (methane fully decoupled from water). Two configuration files should be prepared to conduct forward and backward transformations between reference and target states.

```
source ../FEP_tools/fep.tcl

alch                on
alchType            fep
alchFile            decouple.fep
alchCol             B
alchOutFreq         10
alchOutFile         decouple.fepout

alchVdwLambdaEnd    1.0
alchElecLambdaStart 0.5
alchVdWShiftCoeff   5.0
alchDecouple        yes

alchEquilSteps      2500
set nSteps          52500

runFEP         0.0   1.0   0.1      $nSteps
```

The above settings in the configuration file setup the forward transformation. To setup the backward transformation, create a copy of the configuration file for the forward transformation and replace the last line to:

```
runFEP         1.0   0.0  -0.1      $nSteps
```

The usual MD section of the NAMD configuration file is written for an MD run in the isobaric-isothermal ensemble. The temperature and the isotropic pressure are maintained at 300 K and 1 atm, using, respectively, Langevin dynamics and the Langevin piston method. Long-range electrostatic forces are handled by means of the particle-mesh Ewald (PME) algorithm. The `rigidBonds` options set to `all` and the equations of motion are integrated with a time step of 2.0 fs.

2.2. Running the calculations

This set of calculations is more CPU-intensive than the previous, prototypical unfolding of deca-alanine. Yet, the use of a stratification strategy, by breaking down the reaction pathway into consecutive windows — either from the perspective of alchemical free energy perturbation (FEP) or from that of thermodynamic integration along a reaction coordinate — is envisioned to improve the overall efficiency of the solvation free energy calculations.

A FEP calculation performed over 2×10 windows, with 100 ps of sampling per window, *i.e.* a total of 2 ns, represents a reasonable scheme for the purpose of this tutorial.

Similarly, ABF simulations of 0.5 ns carried out in the five adjacent windows forming the complete translocation pathway of methane across the water liquid-vapor interface constitutes an appropriate duration to yield a convincing answer. The user is, however, advised to repeat the free energy calculations with a more thorough sampling strategy to assess its convergence properties. [5]

It is unlikely that the necessary simulations, both ABF and FEP, can be completed in a single day using only one processor. Therefore, after setting up the simulations and ensuring they run, you are encouraged to examine the provided example output in lieu of running them to completion.

2.3. Analyzing FEP simulation output with ParseFEP

Once the FEP simulations of forward and reverse transformations are performed, ParseFEP plugin in VMD can be used to reconstruct the SOS or BAR free-energy differences and perform the error analysis [14]. Input of ParseFEP are two separate `alchOutFile` output files corresponding to forward and backward transformations, which are also provided in the example output.

To access ParseFEP, open a new VMD session and choose the Extensions → Analysis → Analyze FEP Simulation (Fig. 3). To obtain BAR free-energy differences, in the ParseFEP GUI set temperature to the temperature of the simulation, set Gram-Charlier order to 0, select paths to `alchOutFile` files for forward and backward transformations, and check the box for BAR-estimator. After running the ParseFEP plugin, the results in the form of the free-energy difference between reference and target states and the

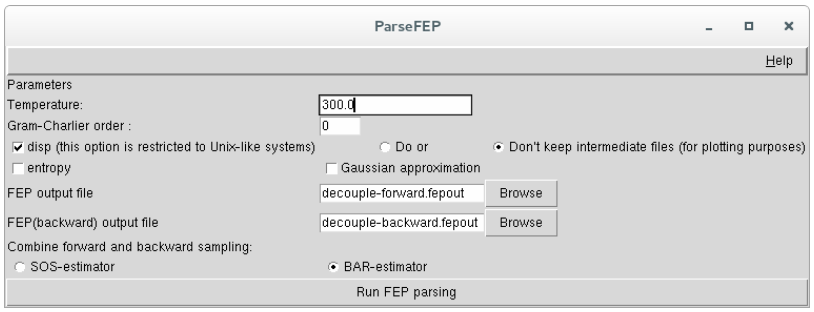


Figure 3: ParseFEP plugin GUI in VMD.

associated error will be provided in the VMD terminal window. These results and detailed analysis of FEP simulations (such as free energy profiles in dependence of λ) are recorded in the ParseFEP.log file, written out by the ParseFEP plugin.

On Unix-like systems, it is possible to plot ParseFEP results, by checking the disp option in ParseFEP GUI before submitting the ParseFEP analysis. The plotted results include the time evolution of free energy differences for each of the λ windows and the probability distribution plots (for backward and forward transformations). ParseFEP plots for the performed FEP simulations are shown in Fig. 4.

While use of ParseFEP to obtain BAR free-energy differences is encouraged, another script provided in the distribution can be used to recover SOS free-energy differences. A single `alchOutFile` file (resulting from both the “forward” and ”backward” transformations) can serve as an input to the provided script.

2.4. Analyzing the results

The results of the free energy calculations, following the two alternative routes, are depicted in Figure 5. The ABF free energy profile exhibits a shallow minimum near the Gibbs dividing surface, characteristic of favorable dispersion interaction with the interfacial environment. The hydration free energy is equal to the difference of the free energies around ξ = 0, *i.e.* when the solute is fully immersed in the water lamella, and around ξ = 25 Å, *i.e.* when it is located in the gas phase, sufficiently far from the aqueous interface. Based on the data of Figure 5, $\Delta G_{\mathrm{hydration}}$ = +2.4 kcal/mol, in excellent agreement with the value of 2.44 kcal/mol expected from the CHARMM force field, and within $k_B T$ of the experimentally

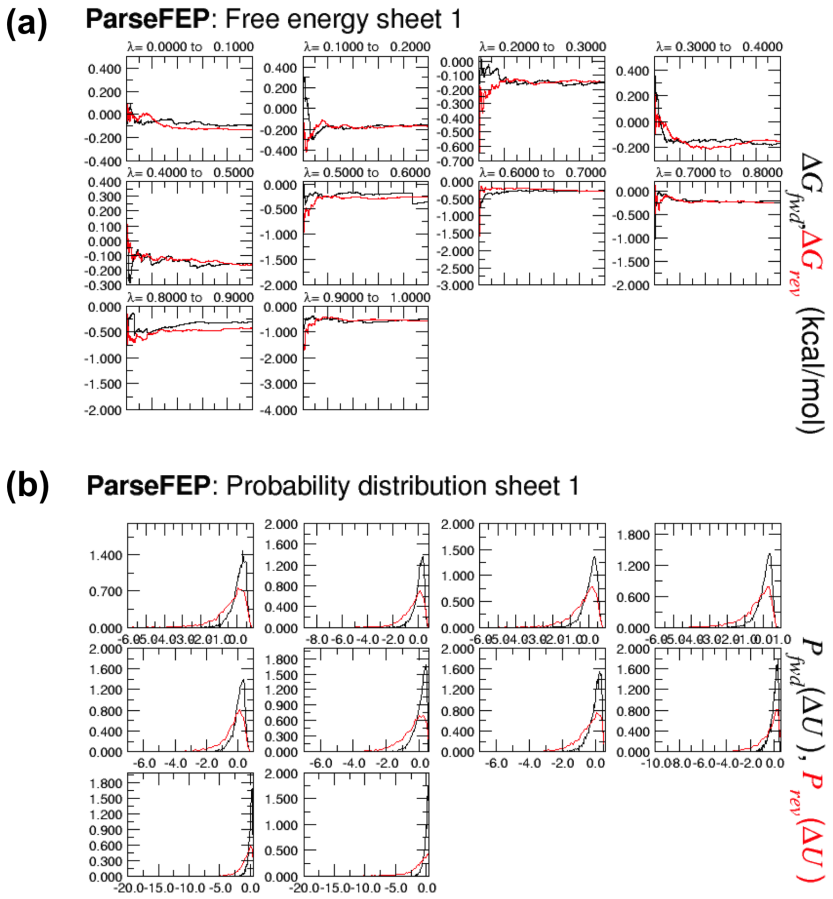


Figure 4: ParseFEP plots of FEP simulation results. (a) Time evolution of the free energy. (b) Histograms of the probability distributions.

determined value of +2.0 kcal/mol. [15] The reference CHARMM value was obtained by Shirts *et al.* on the basis of lengthy, 73.2-ns thermodynamic integration calculations.

The alternate, perturbative route yields a free energy change, $-\Delta G^1_{\text{decoupling}}$, of +2.4 kcal/mol, coinciding with the ABF estimate. This result underscores the consistency of two completely independent approaches for measuring free energy differences — *viz.* FEP *vs.* ABF.

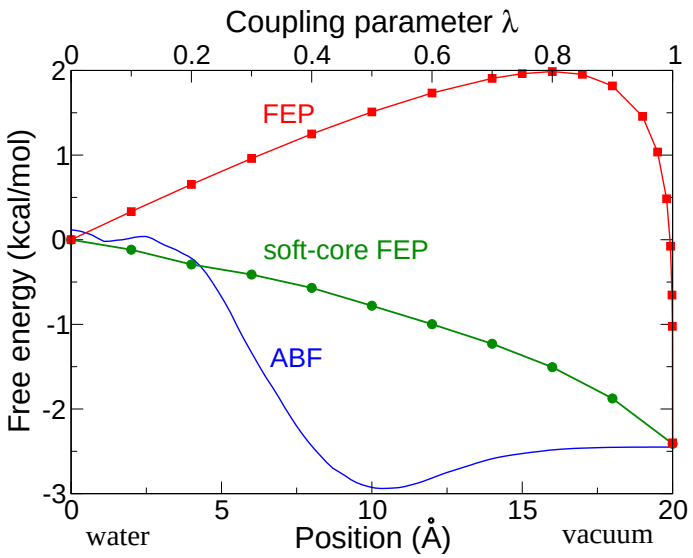


Figure 5: Free energy profile delineating the reversible translocation of a methane molecule across a water liquid-vapor interface as a measure of its hydration free energy (blue). For comparison purposes, alternative, FEP calculation wherein the solute is decoupled from the aqueous environment, with a traditional linearly scaled potential (red) and using a soft-core potential (green). The red curve exhibits a singularity at $\lambda = 1$, making convergence extremely difficult.

Probing the convergence properties of the FEP simulation

Verify that the free energy difference in each window has reached a plateau. This information can be extracted from `fepOutFile`, breaking the latter into individual files — *e.g.* `grep "FepEnergy:"` *fepOutFile name* `| split -l` *number of MD steps per window*. From this data, the overlap of configurational ensembles ought also to be checked by constructing histograms of ΔU. On Unix-like systems, ParseFEP plugin in VMD can be used to plot time evolution of free energy differences in all windows and the histograms of ΔU (shown in Fig. 4).

3. Isomerization of N–acetyl–N'–methylalanylamide

The main thrust of this section of the tutorial is to address the important issue of conformational equilibria in short peptides. To tackle this problem, use will be made of the prototypical N–acetyl–N'–methyl–L–alanylamide (NANMA) [16, 17, 18, 19], often improperly referred to as dialanine or alanine dipeptide (see Figure 6). Isomerization of this short peptide will be investigated employing classical molecular dynamics and ABF simulations. To assess the role played by the environment on the conformational

equilibrium of NANMA, the conformational preference of the latter will be examined both in vacuum and in an aqueous solution.

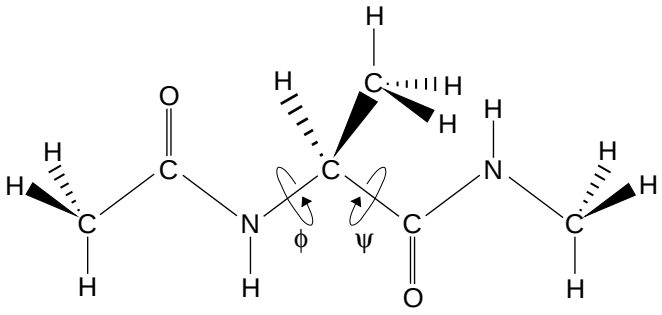


Figure 6: Blocked N–acetyl–N'–methyl–L–alanylamide (NANMA), featuring torsional angles φ and ψ

In unbiased molecular dynamics simulations, the peptide isomerizes freely, guided by the conservative forces that derive from the potential energy function. The free-energy difference between two conformations may be estimated from their respective probabilities as measured from equilibrium sampling, yet this estimate will converge very slowly, making the method inefficient here and inapplicable to slower systems. Conversely, a two-dimensional free-energy surface will be established using ABF to not only determine the free-energy differences between key conformational states of the Ramachandran map, but also estimate the barriers that separate these states.

3.1. System setup

3.1.1. Generating the PSF file

NANMA can be thought of as an L–alanine amino acid, blocked by an acetyl moiety, `ACE`, and an NH–methyl moiety, `CT3` instead of the standard termini used in CHARMM, *viz.* `NTER` and `CTER`. An initial conformation of alanine can be built using the molecular builder Molefacture available in VMD. Create an alanine residue with the protein builder tool and save it as a PDB file. Then, using `psfgen` and the CHARMM22 protein topology file, build a segment containing the alanine residue and capped by the `ACE` and `CT3` patches:

```
segment ALA {first ACE; last CT3; residue 1 ALA}
```

Input coordinates for alanine using `coordpdb` and build the termini from internal coordinates with `guesscoord`. Save a PSF and a PDB file, which will both be utilized in the gas phase simulations.

The Solvate plugin of VMD will next be used to prepare the simulation in an aqueous environment. Considering the size of NANMA, limiting the dimensions of the cell to $24 \times 24 \times 24$ Å^3 appears to be sufficient. These dimensions correspond to approximately 400 molecules. The structure files (PSF) and coordinate files (PDB) for the isolated and the solvated NANMA are provided in the distribution of this tutorial, so that the reader may get started rapidly with the molecular dynamics and ABF simulations.

3.1.2. *Preparing the NAMD configuration file*

It is recommended that a fast energy minimization of the structure constructed with Molefacture be run, in particular in the case of the hydrated peptide, both to refine poorly guessed coordinates and to eliminate van der Waals clashes. This step may be achieved employing a NAMD configuration file finishing with the command `minimize 500`.

Files to run four simulations (or sets of simulations) are provided. The first two involve equilibrating NANMA in vacuum and in solvent, so that its typical conformations in these environments can be observed. The latter two then map out the entire (φ,ψ) free-energy surface using both ABF and metadynamics.

Gas phase molecular dynamics simulations. In the case of isolated NANMA, a molecular dynamics simulation will be performed in the microcanonical ensemble, $(N, V, \mathcal{E})$, at 300 K, rescaling periodically the velocities of the participating atoms. The length of the trajectory should be on the order of 500 ps. The equations of motion will be integrated with a time step equal to 0.5 fs. Chemical bonds will be constrained to their respective equilibrium length, *viz.* `rigidbonds all`. Configurations will be stored every 50 to 100 steps in a `dcd`–formatted file.

Molecular dynamics simulations in water. In order to determine the extent to which the conformational equilibrium of NANMA is affected by its environment, the NANMA peptide will be immersed in a thermostated bath of water. The energy-minimized configuration will subsequently be equilibrated in the isobaric–isothermal, (N, P, T), ensemble at 300 K and 1 atm by means of Langevin dynamics and the Langevin piston algorithm. Long–range electrostatic interactions will be handled using the particle mesh Ewald (PME) approach. A 2 fs time–step will be employed to integrate the equations of motion, and all chemical bonds connecting hydrogen atoms to heavy atoms will be frozen to their equilibrium length. To

reduce the computational effort, it is strongly advised to use the multiple time-step formalism available in NAMD, with an appropriate choice of the variables `fullElectFrequency` and `nonbondedFreq`. After proper equilibration, a 500 ps production simulation will be performed, storing configurations every 50 to 100 time–steps in a `dcd` file.

ABF and metadynamics calculations. As was mentioned previously, exploration of the Ramachandran map will be conducted in the gas phase and in the aqueous environment, employing a two-dimensional model reaction coordinate formed by the (φ, ψ) variables. From a practical perspective, two collective variables colvar will be defined based on the `dihedral` component, which is a torsional angle between four groups of atoms — in this particular example, between four atoms. In the script below, conformational sampling is limited to the quadrant defined by $-180^\circ \leq \varphi \leq 0^\circ$ and $-180^\circ \leq \psi \leq 0^\circ$. Similar NAMD input files will be written for the exploration of the three other quadrants of the Ramachandran map. The atoms used to define φ are `(CY, N, CA, C)`, and those used for ψ are `(NT, C, CA, N)`, in this order. Adding the `oneSiteTotalForce` option ensures that force measurements for the two variables satisfy the orthogonality condition $\vec{v}_i \cdot \vec{\nabla}\xi_j = \delta_{ij}$, where $\vec{v}_i$ is the vector field along which the collective variable ξ_i is propagated [4].

```
colvarsTrajFrequency    500
colvarsRestartFrequency 5000

colvar {
  name phi

  width 5.0
  lowerboundary  -180
  upperboundary     0
  lowerWallConstant 0.2
  upperWallConstant 0.2

  dihedral {
    oneSiteTotalForce
    group1 {
      atomnumbers 13
    }
    group2 {
      atomnumbers 15
    }
    group3 {
      atomnumbers 17
    }
```

```
    group4 {
      atomnumbers 1
    }
  }
}

colvar {
  name psi

  width 5.0
  lowerboundary  -180
  upperboundary     0
  lowerWallConstant 0.2
  upperWallConstant 0.2

  dihedral {
    oneSiteTotalForce
    group1 {
      atomnumbers 3
    }
    group2 {
      atomnumbers 1
    }
    group3 {
      atomnumbers 17
    }
    group4 {
      atomnumbers 15
    }
  }
}

abf {
  colvars       phi psi
  fullSamples   100
}
```

Note that in the above, the threshold defined by the keyword `fullSamples`, beyond which the adaptive biasing force is applied, is markedly smaller than in a one-dimensional free-energy calculation. This is due to the larger number of bins, *i.e.* [(`upperboundary` - `lowerboundary`) / `width`]^2 = 1,296. With a value of 100 for `fullSamples`, assuming that all the bins are filled uniformly, a minimum of 129,600 MD steps is required before biases are applied for the first time along the chosen order parameter. This issue is of less concern in the case of a metadynamics calculation, where the `abf` block is replaced by the following:

```
metadynamics {
  colvars phi psi
  hillWidth 2
  hillWeight 0.01
}
```

On account of the definition of the φ and ψ torsional angles, featuring heavy atoms bonded covalently to hydrogen atoms, the `rigidBonds` option of NAMD ought to be either turned off, or limited to water molecules (`rigidbonds water`). In the implementation of ABF in NAMD, constraint forces are not taken into account in the evaluation of the instantaneous force acting along the reaction coordinate [5]. It is, therefore, pivotal that contributions arising from holonomic constraints do not contaminate the measured force. By setting `rigidBonds` to water molecules only, a smaller integration time step will likely be needed to guarantee energy conservation of the molecular system.

3.2. Running the adaptive biasing force and metadynamics calculations

The rapid isomerization of NANMA and the decomposition of the Ramachandran map into four quadrants are expected to help achieve convergence of the free-energy calculation within a period of 2.5 ns, after thorough equilibration in each quadrant. The complete (φ, ψ) map can be built from the individual ABF simulations by means of an additional NAMD run without actual MD (zero steps), wherein the bounds of the variables are altered to embrace the full Ramachandran map, and replacing the ABF block with the following:

```
abf {
  colvars       phi psi
  inputPrefix   colvarA01o colvarB01o colvarC01o colvarD01o
}
```

For comparison between methods, the vacuum simulations provided are set up to use ABF and the solvated ones to use metadynamics. Note that in two-dimensional ABF calculations, NAMD will supply maps of the average force acting along the chosen order parameters, φ and ψ. In order to recover the free-energy landscape, an external program, `abf_integrate`, will be utilized. This program is supplied with the NAMD distribution and reads the information about the gradients found in the ABF output files `*.grad`. Employing a Monte Carlo scheme, it determines the biases that counteract exactly

the underlying gradients and from whence the free-energy surface is constructed. In metadynamics, the complete 2D PMF is provided directly by NAMD. The reader is referred to the NAMD user's guide for further details about the syntax and options of `abf_integrate` and to Hénin et al. [4] for further comparison between ABF and metadynamics.

3.3. Analyzing the results

Gas phase molecular dynamics simulations. Open the equilibrium MD trajectory (either the one that you ran or the one provided as an example) using VMD. Create a representation highlighting the intramolecular hydrogen bonds of NANMA. The time–evolution of the two dihedral angles φ and ψ can be analyzed, in terms of concomitant torsional transitions. Using the Label → Dihedrals function in VMD, label the φ and ψ angles. Under the Graphics → Labels menu, each dihedral angle can be plotted over time. The predominant conformations of NANMA in the example gas phase trajectory correspond to (φ,ψ) values of approximately (-150, 150) and (-75, 75), known as the extended C_5 and $C_{7-\mathrm{equatorial}}$ motifs, respectively [16]. Additionally, the $C_{7-\mathrm{axial}}$ motif (60, -75) is sometimes observed in gas–phase equilibrium simulations. Based on the equilibrium distributions, inferred from the time–evolution graphs as a simple count of occurrences, the relative stability of these conformers can be deduced. See the provided files `alphaL/R.pdb` and `c7ax/eq.pdb` for examples of these conformations.

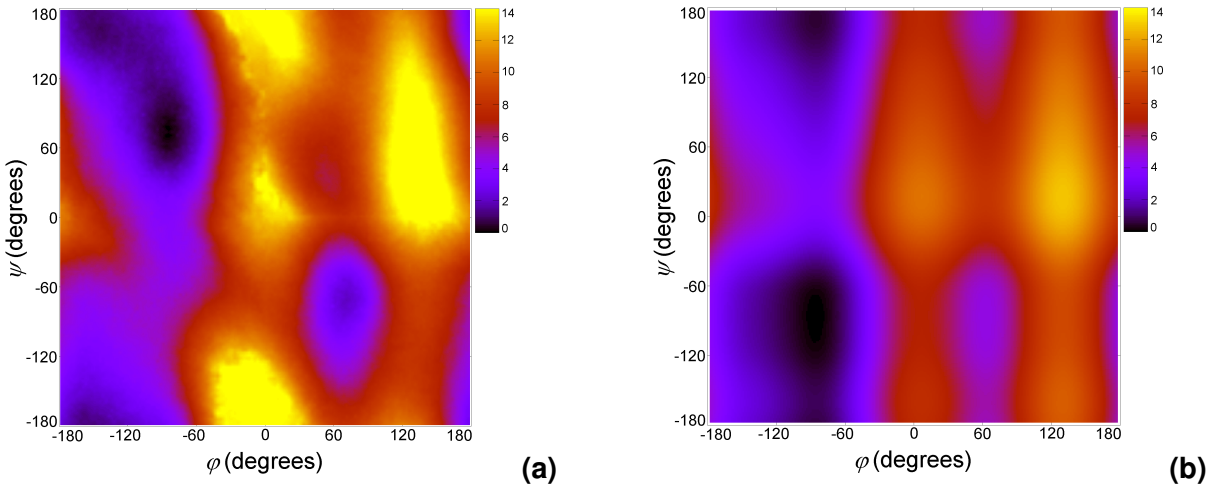


Figure 7: Free-energy surface characterizing the conformational equilibrium of N–acetyl–N'–methylalanylamide in the gas phase (a) and in bulk water (b), derived from a two-dimensional ABF simulation that uses the φ and ψ torsional angles as the reaction coordinate.

Molecular dynamics simulations in water. In a similar fashion, isomerization of NANMA in the aqueous

environment can be examined from the variation of the φ and ψ torsional angles monitored along the MD trajectory. Of particular interest, analysis of the hydration properties of the terminally blocked peptide reveals a stringent competition in the formation of intra– and intermolecular hydrogen bonds, by and large at the expense of the former. As a result, the population of the $C_{7-\mathrm{axial}}$ and $C_{7-\mathrm{equatorial}}$ conformations, and to a lesser extent extended C_5, diminish dramatically, while that of the right-handed α–helix — or α_R (-75, -60), and, less commonly, β–strand (-120, 120) emerge and become predominant. On the basis of the Boltzmann weights of the observed conformers, the free-energy difference between the α_R and β states can be determined.

ABF and metadynamics calculations. Compared to the rudimentary MD simulations described previously, either in the gas phase or in an aqueous solution, ABF and metadynamics calculations provide a detailed picture of the free-energy landscape of NANMA, from which not only relative stabilities can be inferred, but also the barriers that separate the free-energy minima. In the gas phase, the free-energy surface features the three key conformations, *viz.* C_5, $C_{7-\mathrm{axial}}$ and $C_{7-\mathrm{equatorial}}$, highlighted in the aforementioned MD simulations. Figure 7 confirms the assertion made in the previous paragraph — the water environment disrupts the intramolecular hydrogen bonds that stabilize the $C_{7-\mathrm{axial}}$ and $C_{7-\mathrm{equatorial}}$ conformations. Compare your PMFs, or the example PMFs provided, to those in Figure 7 and that in the paper Hénin et al. [4].

Conformational equilibria are affected by the environment. In the particular instance of NANMA, disruption by the aqueous surroundings of the intramolecular hydrogen bonds that stabilize the $C_{7-\mathrm{axial}}$ and $C_{7-\mathrm{equatorial}}$ motifs in vacuum results in the emergence in water of the α_R and β, otherwise unfavorable in the gas phase. Conformational preference of α_R and β, with respect to C_5, $C_{7-\mathrm{axial}}$ and $C_{7-\mathrm{equatorial}}$ can be also rationalized by additional considerations, on the basis that the electrostatic contribution to the free energy varies as μ^2, the square of the molecular dipole moment [20]. Show that the dipole moment of α_R and β favors their stabilization in water, and that isomerization between the latter motifs is accompanied by a transition from parallel to antiparallel of the bond dipole moments [16, 17].

Estimating the population of the α_R and β states requires integration over basins of the two-dimensional free-energy map:

$$\Delta G_{\alpha_R \longrightarrow \beta} = -k_B T \ \ln \frac{\displaystyle\int_{\beta} \mathcal{P}(\varphi,\psi)\, d\varphi\, d\psi}{\displaystyle\int_{\alpha_R} \mathcal{P}(\varphi,\psi)\, d\varphi\, d\psi}$$

4. Ion pairing in aqueous solution

In the last section of this tutorial, another important problem of molecular modeling, ion association, will be considered through the prototypical example of the sodium-chloride ion pair. The main objective of this test case is to recover the potential of mean force delineating ion pairing, *i.e.* the reversible work required to bring the anion and the cation from quasi-infinite separation to a contact distance. The ABF approach will be employed to describe the reversible association with an interatomic distance as the most natural choice for the reaction coordinate.

4.1. System setup

The molecular system consists of the ion pair immersed in a bath of water molecules. To avoid periodicity induced artifacts, [21] the dimensions of the primary cell ought to be large enough. A box length of 30 Å, *viz.* approximately 800 water molecules, represents a reasonable compromise with computational cost. The distance separating the ions in the primary and the adjacent cells being equal to 30 Å, the interaction energy reduces to $q^2/4\pi\epsilon_0\epsilon_1 r$ = 0.1 kcal/mol with an ideal macroscopic permittivity of 78.4 for bulk water.

4.1.1. Generating the PSF file

The first step consists in creating a PDB file containing the Cartesian coordinates of the ion pair. The initial separation of the ions will be set arbitrarily to 3 Å. Next, employing `psfgen` with `top_all22_prot.inp`, in which the topology of the sodium and the chloride ions are defined, build the structure PSF file for the isolated solute.

Use the `Add Solvation Box` feature of VMD to hydrate the ion pair and create PDB and PSF files corresponding to the complete system.

Prior to proceeding with the free energy calculation, it is recommended that the molecular assembly be properly energy-minimized and equilibrated in the isobaric-isothermal ensemble, at 300 K and 1 atm. Refer to the NAMD user's guide to set these simulations up. In the course of the latter, the user may wish

to freeze the position of the ions. This can easily be achieved using the `fixedAtoms` options of NAMD.

4.1.2. *Preparing the NAMD configuration file*

As mentioned previously, the variable that will use as a reaction coordinate is `distance`, with `group1` set to correspond to the sodium ion and `group2` to the chloride ion. The reaction pathway will be extended up to approximately half of the length of the simulation cell — *viz.* here, 14 Å. At the largest separation of the ions, the interaction energy is equal to $q_1 q_2 / 4\pi\epsilon_0\epsilon_1 r = -0.3$ kcal/mol, which is reasonably small, albeit not strictly zero. The potential of mean force, therefore, ought to be anchored at this value, rather than setting $\Delta G(\xi_{\mathrm{max}}) = 0$. Here we assume that at such a distance, the free energy profile is dominated by direct Coulombic interaction.

```
colvarsTrajFrequency    500
colvarsRestartFrequency 5000

colvar {
   name IonDistance

   width 0.1
   lowerboundary  2.0
   upperboundary 14.0
   lowerwallconstant 10.0
   upperwallconstant 10.0

   distance {
      group1 {
         atomnumbers { 1 }
      }
      group2  {
         atomnumbers { 2 }
      }
   }
}

abf {
  colvars IonDistance
  fullSamples      500
  hideJacobian
}
```

A standard threshold, `fullSamples`, of 500 samples per bin is proposed here prior to application of the ABF onto the reaction coordinate. Given the simplicity of the system and the virtual absence of other, slowly relaxing degrees of freedom directly coupled to ξ and which could hamper diffusion along the latter, a lesser number of force samples could be considered. This option ought to be tested in the light of the resolution of $\varrho(F_\xi)$, keeping in mind that the average force, $\langle F_\xi \rangle_\xi$, should be properly converged before the adaptive bias is enforced.

The standard MD section of the NAMD configuration file should be written for an MD run in the isobaric-isothermal ensemble. The temperature and the isotropic pressure will be maintained at 300 K and 1 atm, using, respectively, Langevin dynamics and the Langevin piston method. Long-range electrostatic forces will be handled by means of the particle-mesh Ewald (PME) algorithm. The `rigidBonds` option will be set to `all` and the equations of motion will be integrated with a time step of 2.0 fs.

4.2. Running the adaptive biasing force calculation

The cost of this calculation is primarily burdened by the evaluation of water-water interactions. Compared with a lengthy reference free energy calculation, the main features of the potential of mean force can be recovered with a reasonable accuracy within 0.5 ns. We expect that the region of the free energy profile corresponding to the contact minimum (CM), where the two ionic species are paired, should rapidly coincide with that of the reference curve. This can be easily explained in terms of lower entropy of the ion pairs and the solvent molecules around it, and, hence, a better convergence of the configurational averages at low values of ξ. In sharp contrast, separation of the ions yields a far greater number of possible spatial arrangements of the neighboring water molecules, and, as a result, a larger configurational entropy. Convergence of the free energy calculation from the solvent-separated minimum (SSM) and beyond is, therefore, anticipated to be significantly slower. Although a description of the entire reaction pathway by means of a 0.5-ns simulation length constitutes a reasonable scheme in the context of this tutorial, an alternative, more exhaustive sampling strategy is advisable in cases where a more thoroughly converged PMF is In the spirit of the preceding example, wherein a methane molecule was translocated across a water liquid-vapor interface, breaking the reaction pathway into consecutive windows forming is expected to improve both the efficiency and the accuracy of the free energy calculations. In particular, since the CM region of the potential of mean force converges quickly, a lesser amount of sampling could be invested in the latter, at the benefit of those regions characterized by a greater configurational entropy.

4.3. Analyzing the results

The free energy profile delineating the reversible association of a sodium and a chloride ion is depicted in Figure 8. From the onset, it can be seen that this potential of mean force possesses a narrow CM, located around 2.7 Å and 1.8 kcal/mol deep, and a shallower SSM, centered about 5.1 Å and 1.0 kcal/mol deep. This free energy profile is very similar to that obtained recently by Masunov and Lazaridis, [22] on the basis of umbrella sampling [23] simulations with the CHARMM force field.

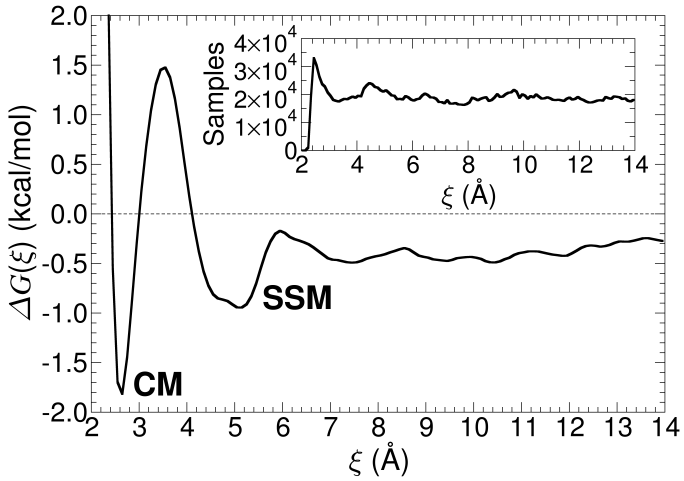


Figure 8: Free energy profile delineating the reversible association of a sodium and a chloride ion in an aqueous medium. CM and SSM are the contact the solvent-separated minima of the potential of mean force. Inset: Number of force samples accrued along the reaction coordinate corresponding to the distance that separates the two ions.

Direct comparison of the theoretical results with experiment is feasible through the estimation of the association constant from the free energy profile: [24]

$$K_a = 4\pi \int_0^{\Xi} \xi^2 \, \exp[-\beta \Delta G(\xi)] \, \mathrm{d}\xi \tag{2}$$

where Ξ delineates the limit of association.

Based on equation (2), what is the value of K_a for the sodium-chloride ion pair? What is the unit of this constant, and the implicit standard state? Adjust the constant for a 1 mol/L standard state.

Acknowledgments

Development of this tutorial was supported by the National Institutes of Health (P41-RR005969 - Resource for Macromolecular Modeling and Bioinformatics).

References

[1] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, L.; Schulten, K., Scalable molecular dynamics with NAMD, *J. Comput. Chem.* **2005**, *26*, 1781–1802.

[2] Bhandarkar, M.; Brunner, R.; Chipot, C.; Dalke, A.; Dixit, S.; Grayson, P.; Gullingsrud, J.; Gursoy, A.; Hardy, D.; Hénin, J.; Humphrey, W.; Hurwitz, D.; Krawetz, N.; Kumar, S.; Nelson, M.; Phillips, J.; Shinozaki, A.; Zheng, G.; Zhu, F. NAMD *user's guide, version 2.6b1*. Theoretical biophysics group, University of Illinois and Beckman Institute, 405 North Mathews, Urbana, Illinois 61801, July 2005.

[3] Darve, E.; Rodrííguez-Gómez, D.; Pohorille, A., Adaptive biasing force method for scalar and vector free energy calculations, *J. Chem. Phys.* **2008**, *128*, 144120.

[4] Hénin, J.; Forin, G.; Chipot, C.; Klein, M. L., Exploring multidimensional free energy landscapes using time-dependent biases on collective variables, *J. Chem. Theor. Comput.* **2010**, *6*, 35–47.

[5] Hénin, J.; Chipot, C., Overcoming free energy barriers using unconstrained molecular dynamics simulations, *J. Chem. Phys.* **2004**, *121*, 2904–2914.

[6] Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the Cartesian equations of motion for a system with constraints: Molecular dynamics of n–alkanes, *J. Comput. Phys.* **1977**, *23*, 327–341.

[7] Andersen, H. C., Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations, *J. Comput. Phys.* **1983**, *52*, 24–34.

[8] Fiorin, Giacomo; Klein, Michael L.; Hénin, Jerome, Using collective variables to drive molecular dynamics simulations, *Molecular Physics* **June 2013**, *111*, 3345–3362.

[9] Chipot, C.; Hénin, J., Exploring the free energy landscape of a short peptide using an average force, *J. Chem. Phys.* **2005**, *123*, 244906.

[10] Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K., Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality, *J. Chem. Phys.* **2003**, *119*, 3559–3566.

[11] Isralewitz, B.; Gao, M.; Schulten, K., Steered molecular dynamics and mechanical functions of proteins, *Curr. Opin. Struct. Biol.* **2001**, *11*, 224–230.

[12] Zwanzig, R. W., High–temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys.* **1954**, *22*, 1420–1426.

[13] Chipot, C.; Pohorille, A., Eds., *Free energy calculations. Theory and applications in chemistry and biology*, Springer Verlag, 2007.

[14] Liu, P.; Dehez, F.; Cai, W.; Chipot, C., A Toolkit for the Analysis of Free-Energy Perturbation Calculations, *J. Chem. Theo. Comp.* **2012**, *8*, 2606–2616.

[15] Ben-Naim, A.; Marcus, Y., Solvation thermodynamics of nonionic solutes, *J. Chem. Phys.* **1984**, *81*, 2016–2027.

[16] Tobias, D. J.; Brooks, C. L., Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results, *J. Phys. Chem.* **1992**, *96*, 3864–3870.

[17] Chipot, C.; Pohorille, A., Conformational equilibria of terminally blocked single amino acids at the water–hexane interface. A molecular dynamics study, *J. Phys. Chem. B* **1998**, *102*, 281–290.

[18] Smith, P. E., The alanine dipeptide free energy surface in solution, *J. Chem. Phys.* **1999**, *111*, 5568–5579.

[19] Jang, H.; Woolf, T. B, Multiple pathways in conformational transitions of the alanine dipeptide: an application of dynamic importance sampling, *J. Comput. Chem.* **2006**, *27*, 1136–1141.

[20] Onsager, L., Electric moments of molecules in liquids, *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

[21] Hünenberger, P. H.; McCammon, J. A., Ewald artifacts in computer simulations of ionic solvation and ion–ion interactions: A continuum electrostatics study, *J. Chem. Phys.* **1999**, *110*, 1856–1872.

[22] Masunov, A.; Lazaridis, T., Potentials of mean force between ionizable amino acids side chains in water, *J. Am. Chem. Soc.* **2003**, *125*, 1722–1730.

[23] Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling, *J. Comput. Phys.* **1977**, *23*, 187–199.

[24] Shoup, D.; Szabo, A., Role of diffusion in ligand binding to macromolecules and cell–bound receptors, *Biophys. J.* **1982**, *40*, 33–39.