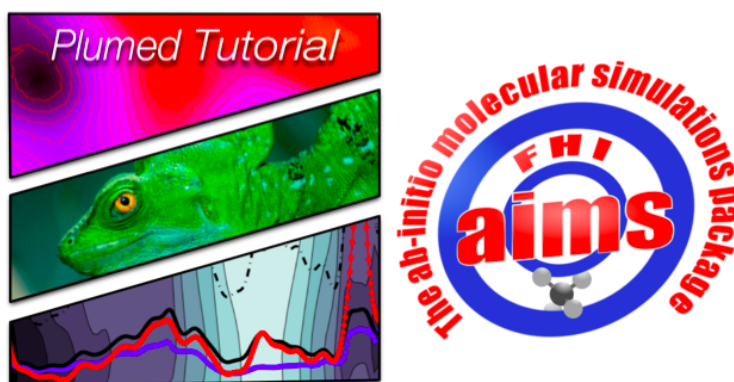# a micro PLUMED Tutorial for FHI-aims users

PLUMED:
portable plugin for free-energy calculations
with molecular dynamics

FHI-aims Developers' and Users' Meeting

Organized by Fritz Haber Institute,
CECAM-mm1p, and CECAM-SCM

Free University Berlin, Berlin, Germany
August 28 - August 31, 2012

This document and the relative computer exercises have been written by Davide Branduardi and Luca Ghiringhelli and is based on a previous tutorial held in CECAM, Lausanne in 2010 whose authors are:

Massimiliano Bonomi
Davide Branduardi
Giovanni Bussi
Francesco Gervasio
Alessandro Laio
Fabio Pietrucci

Tutorial website:
http://sites.google.com/site/plumedtutorial2010/

PLUMED website:
http://www.plumed-code.org

PLUMED users Google group:
plumed-users@googlegroups.com

PLUMED reference article:

Contents

Chapter 1

What PLUMED is and what it can do

PLUMED is a series of open source (L-GPL) routines that are now interfaced to a number of molecular dynamics engines (beyond FHI-aims: NAMD, GROMACS, LAMMPS, ACEMD, SANDER, CPMD, G09-ADMP, and more) that enable the code to perform a number of enhanced sampling calculations. Its core consists in a routine that takes the position of the atoms at each timestep and introduces forces according to the specific configuration of the system: this is the core of many enhanced sampling calculations. Since this is a very tiny link to the MD code, this can be done in a rather general way and that is on of the main reasons why many MD codes incorporates it.

The usefulness of these enhanced sampling calculations is that they allow for calculating free-energy differences associated to conformational transitions, chemical reaction and phase transitions.

In general, the free energy of a system in a specific point can be defined as

$$F(s) = -\frac{1}{\beta} \log P(s) \tag{1.1}$$

where F is the Helmoltz's free energy, s is a descriptor of the system (a bond distance, angle, or another generic order parameter of some sort), $P(s)$ is the probability that the system is in the state labelled

by the descriptor s, $\beta = 1/k_B T$ where k_B is the Boltzmann constant and T is the temperature of the system.

Therefore, building an histogram of the conformations is the easiest way one can build a free-energy profile as a function of some parameter.

Unfortunately, very often it happens that the exploration via MD is all but exhaustive. In fact the system may stay trapped in some minimum inasmuch the kinetic energy provided from the thermostat (which is $0.5 k_B T = 0.3$ kcal/mol per degree of freedom at 300 K) very often does not allow for the the overcoming of barriers within the usual time span of a MD simulations (typically tens to hundreds of ps for ab initio MD, up to several μs for classical-force-field-based MD). In this case we say that the system is in a "metastable state" which is exemplified in the Fig. 1

In order to illustrate this, let us consider a typical barrier for a chemical reaction: $\Delta U^\ddagger \simeq 8$ kcal/mol. One can estimate the rate of a reaction by using the Arrhenius equation that reads

$$\nu(T) = \nu_0 \exp^{-\frac{\Delta U^\ddagger}{k_B T}} . \tag{1.2}$$

By considering ν_0 for a typical covalent bond, $\nu_0 = 5 \ 10^9 \text{s}^{-1}$, as an upper bound, one easily get a rate of $7 \ 10^3$ events per second. Therefore, each single event can well take up to a millisecond and the observation of one single event is (very) fortuitous in the typical timescales of a DFT-based MD. Moreover, since the estimate of the probability requires converged statistics, the converged free-energy calculation can require even much more time.

What PLUMED can do is to physically displace the system from where it is so that you can measure the slope of the free-energy landscape in various ways and then reconstruct the free energy or simply make some event happen. In the next we'll show some basic functionalities.
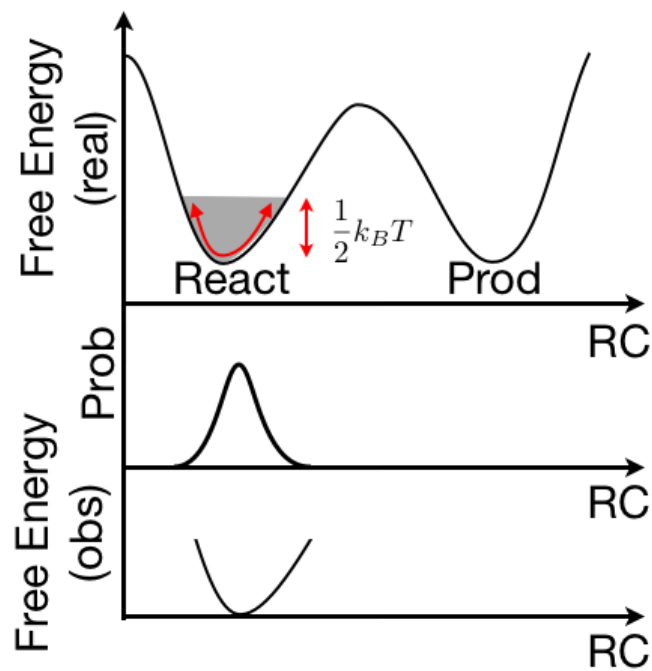
Figure 1.1: Sketch of the typical case where the dynamics is locked in a metastable state. Upper panel: the real free energy landscape is represented for a model landscape. Since the kinetic energy is low compared to the barrier, the crossing is highly unlikely and the system stays trapped. The observed probability is reported in the middle panel and in the lower panel is the computed free energy from the histogram. This is clearly different from the actual one which is reported in the upper panel.

Chapter 2

Basics: how to enable PLUMED in FHI-aims

Enabling PLUMED from FHI-aims requires a separate compilation. A separate

`Makefile.meta`

is provided (see FHI-aims manual).

The actual use of the plugin is switched on by this single line in `control.in`:

```
 plumed .true.
```

With `plumed .false.` (default) or nothing, the code would behave exactly as compiled without this plugin. It is implied that some MD scheme must be used in `control.in`, in order to see PLUMED acting. What PLUMED does, in facts, is to modify the molecular dynamics forces according to the selected scheme.

All the specific controls of the free-energy calculation are contained in the file `plumed.dat` (which must be in the working directory, together with `control.in` and `geometry.in`, if `plumed .true.` is set).

As a simple example, let's consider a methyl-chloride molecule and a chlorine ion. This is a well studied system and computationally convenient. Additionally, the chlorine anion is able to displace the second chlorine and give a SN2 (bimolecular nucleophilic substitution)
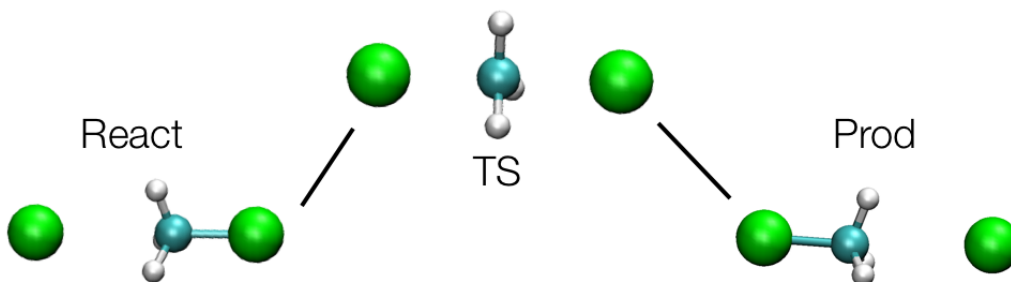
Figure 2.1: A sketch of SN2 reaction

type of reaction represented in Fig 2.1. This is a well studied system (see [1]) and yet computationally convenient.

Concerning FHI-aims setup, all the results reported below are performed with a "first tier" basis set and an `outer_grid` of 302. Runs of this size may take almost two hours on 4 proc machine for 5 ps (typical run time in these calculations). In order to fit into a tutorial time, `pw-lda` density functional and an `outer_grid` of 194 could be used. Here we report a minimal example of `plumed.dat` that just calculates a couple of distances between one atom and (the geometric center of) a group of atoms along the simulation:

```
#
# Note: distances are reported in bohr
#
DISTANCE LIST 1 <g1>
g1->
3 4 5 6
g1<-
DISTANCE LIST 2 <g1>
PRINT W_STRIDE 2
ENDMETA
```

The first line tells PLUMED that a distance should be calculated between the bound chlorine, which is atom number 1 (the first atom in FHI-aims `geometry.in` file) and the geometric center of a group

of three atom represented by the symbol `<g1>` that correspond to the
methyl group. The notation

```
g1->
3 4 5 6
g1<-
```

starts the group, puts the number of the atoms included in it, and
closes the group. The other line `DISTANCE LIST 2 <g1>` asks for
the calculation of the distance between the unbound chlorine and the
methyl group. The line

```
PRINT W_STRIDE 2
```

determines that the distance (and other relevant informations) will be
printed later every 2 timesteps in the `COLVAR` file. Every time that you
restart a simulation, the old `COLVAR` file is backed up in `COLVAR.old`
and a new `COLVAR` file is written. Beware that if you do this two times
the first `COLVAR` file will be lost.

PLUMED produces also a `PLUMED.OUT` file which is very important
for understanding whether PLUMED is really doing what you asked it
to do. We really warmly recommend everyone to give a read to it. You
can find there also useful citations for specific methods implemented.
This file is never backed up and is overwritten every time.

Please note that all the settings and examples reported for this
molecule are meant only to allow you to do some exercise within the
time of the tutorial and are NOT, BY ANY MEANS, similar to a
good production setup which is much more computational intensive.
Therefore, for the same reasons, we do not enter in the details of the
simulations but for one single exception: the calculation is a molecular
dynamics. In our case the input contains a section that looks like:

```
MD_run    5.0  NVT_parrinello 300 0.1
MD_time_step  0.001
MD_clean_rotations  .true.
MD_restart    .false.
```

8

```
output_level MD_light
MD_MB_init  300
MD_RNG_seed 12345
plumed .true.
```

which denotes a molecular dynamics setup where one specifies the ensemble `NVT_parrinello 300 0.1` (a stochastic thermostat [2] at 300K with a coupling parameter τ of value 0.1 ps), with the initial temperature set to 300 K (`MD_MB_init 300`) and a timestep of 1 fs (`MD_time_step 0.001`) which is an upper limit for hydrogen containing molecules (we would recommend 0.25 fs for production). Additionally the statement `plumed .true.` denotes the activation of plumed, therefore a plumed.dat file should be present in the same directory where `geometry.in` and `control.in` are.

When sending the calculation, one obtains a `PLUMED.OUT` file in which it is reported, among the other things:

```
:::::::::::::::: READING PLUMED INPUT ::::::::::::::::
|- GROUP FOUND: g1
|- GROUP MEMBERS:  3 4 5 6

1-DISTANCE: (1st SET: 1 ATOMS), (2nd SET: 4 ATOMS);  PBC ON
|- DISCARDING DISTANCE COMPONENTS (XYZ): 000
|- 1st SET MEMBERS:  1
|- 2nd SET MEMBERS:  3  4  5  6



2-DISTANCE: (1st SET: 1 ATOMS), (2nd SET: 4 ATOMS);  PBC ON
|- DISCARDING DISTANCE COMPONENTS (XYZ): 000
|- 1st SET MEMBERS:  2
|- 2nd SET MEMBERS:  3  4  5  6

|-PRINTING ON COLVAR FILE EVERY 2 STEPS
|-INITIAL TIME OFFSET IS 0.000000 TIME UNITS
```

`|-ANALYSIS: YOU WILL ONLY MONITOR YOUR CVs DYNAMICS`

This means that the groups are detected correctly and that the distances are correctly interpreted along with the PRINT command.

Along with it the `COLVAR` file is obtained. Its format is typically like

```
#! FIELDS time cv1 cv2
     0.0020        3.624455684        5.681056610
     0.0040        3.645635574        5.666974513
     0.0060        3.667778896        5.652366040
     0.0080        3.690936439        5.637267098
```

In the first line is a summary of the fields that one is seeing, namely the time (in ps, the time units of FHI-aims) and the two CVs (in this case two distances, reported in bohr which is the natural distance unit used internally to FHI-aims).

What we observe, by plotting this distance is that the second CV (the distance between the unbound chlorine and the methyl group) is oscillating around 5.75 bohr and never reach the distance of the other C-Cl bond which is around 3.7 bohr. As a matter of fact this represent a metastability since it never shows the other symmetric complex. The metastability is due to the fact that the chlorine atom creates a partial positive charge on the methyl group and the negative anion feels a Coulombic attraction to it.

It is worth mentioning that, since FHI-aims dumps the coordinate at each timestep, one can do this analysis a posteriori via the `driver` utility included in PLUMED distribution and by using the nice interface of VMD (Extensions-Analysis-Collective Variable analyis (PLUMED)) created by Toni Giorgino.

Chapter 3

A simple bias on the system

Now let's try to apply a simple harmonic bias, often called "harmonic potential", that stretches the bond and thus "encourages" the transition. This is the first example that shows how PLUMED biases the simulation.

First, in addition to the previous descriptor, we add another one, namely the difference between the two distances. This is a rather well known and used descriptor whenever some bond breaking/bond forming reaction is involved since, by moving from one value to its negative (in the case of such symmetric reaction), it allows to simultaneously breaking one bond while forming another one.

The `plumed.dat` file therefore will look like:

```
#
# monitor the two components
#
g1->
3 4 5 6
g1<-
DISTANCE LIST 1 <g1>
DISTANCE LIST 2 <g1>
PRINT W_STRIDE 1
# the distance between C-Cl' and C-Cl
DISTANCE LIST 1 <g1> DIFFDIST 2 <g1>
UMBRELLA CV 3 AT 0.0 KAPPA 0.01
```

```
ENDMETA
```

Note that here we have additionally put the directive:
```
UMBRELLA CV 1 AT 0.0 KAPPA 0.01
```
This enables the addition of the harmonic potential that has the form:

$$U_{bias}(x) = \frac{k}{2} \left(CV(x) - CV_0 \right)^2 \tag{3.1}$$

in which `AT` determines the CV_0, `KAPPA` determines the k value. In this case the units of k are hartee/bohr^2 so that the energy is in hartree.

The output of `PLUMED.OUT` in this case shows a line like

```
|-UMBRELLA SAMPLING OF COLVAR 3 AT 0.00: SPRING=0.010 SLOPE=0.00
```

which denotes the addition of the harmonic bias.

The output of `COLVAR` changes in something like:

```
#! FIELDS time cv1 cv2 cv3 vwall XX XX RST3 WORK3
 0.002  3.604  5.694  -2.090  0.021  RST 3  0.000   0.000
 0.004  3.668  5.644  -1.975  0.019  RST 3  0.000   0.000
 0.006  3.709  5.607  -1.897  0.018  RST 3  0.000   0.000
 0.008  3.768  5.555  -1.787  0.015  RST 3  0.000   0.000
 0.010  3.834  5.495  -1.661  0.013  RST 3  0.000   0.000
 0.012  3.898  5.434  -1.535  0.011  RST 3  0.000   0.000
```

in which it appears a new field: `vwall`. This is the potential energy calculated by the harmonic potential. The other remaining fields on the right are only relevant with a moving restraint (see below).

Now, in order to compare with the unbiased results, let's compare the two bond distances in the plain simulation and in the biased one. In Fig. 3.1 it is evident that the umbrella potential is biasing the two distance toward a set of values whose difference is centered on zero value. Since at the beginning the simulation is starting far from the zero value imposed by the umbrella potential, initially the simulation is dominated by this spring and oscillates strongly around the center of the harmonic potential. Generally this behavior disappeares after a while, when the system reaches equilibrium. For this reason it is
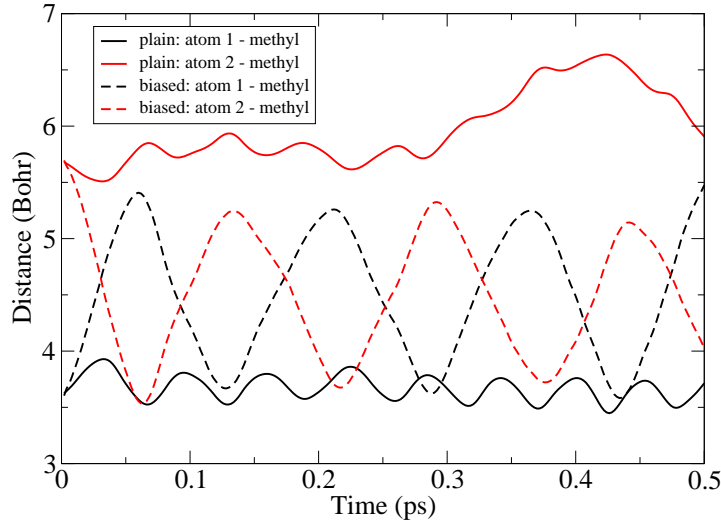
Figure 3.1: A plot of time evolution of the distances of the two chlorine atoms to the central methyl group in case of plain dynamics and biased dynamics.

generally better to use some moving restrain potential that gently drags the system in the desired position (see below).

Conceptually, this procedure is not only allowing to see a transition that otherwise will never happen in a system, but has also a thermodynamic meaning that we can explore here. Let's start from the common definition of Helmoltz free energy as function of an order parameter.

$$F(s_0) = -\frac{1}{\beta} \ln P(s_0) \tag{3.2}$$

$$= -\frac{1}{\beta} \ln \frac{\int e^{-\beta V(x)} \delta(s(x) - s_0) dx}{Q} \tag{3.3}$$

with Q being the canonical partition function:

$$Q = \int e^{-\beta V(x)} dx \ . \tag{3.4}$$

Now we can relate this free energy with the one obtained from a biased ensemble in which we use a function $V_b(s)$ (in the example was the

13

harmonic function). This would read

$$F_b(s_0) = -\frac{1}{\beta} \ln P_b(s_0) \tag{3.5}$$

$$= -\frac{1}{\beta} \ln \frac{\int e^{-\beta(V(x)+V_b(s(x)))} \delta(s(x) - s_0) dx}{Q_b} \tag{3.6}$$

with

$$Q_b = \int e^{-\beta(V(x)+V_b(s(x)))} dx \tag{3.7}$$

Since all the points selected by the delta function $\delta(s(x) - s_0)$ present the same value of $V_b(s(x))$, the exponential related to the bias can be pulled out from the integral to give:

$$F_b(s_0) = -\frac{1}{\beta} \ln e^{-\beta V_b(s_0)} \frac{\int e^{-\beta V(x)} \delta(s(x) - s_0) dx}{Q_b} \tag{3.8}$$

$$= V_b(s_0) - \frac{1}{\beta} \ln \frac{\int e^{-\beta V(x)} \delta(s(x) - s_0) dx}{Q_b} \tag{3.9}$$

$$= V_b(s_0) - \frac{1}{\beta} \ln P(s) \frac{Q}{Q_b} \tag{3.10}$$

$$= V_b(s_0) - \frac{1}{\beta} \ln P(s) + \frac{1}{\beta} \ln \frac{Q}{Q_b} \tag{3.11}$$

$$= V_b(s_0) - \frac{1}{\beta} \ln P(s) + C \tag{3.12}$$

$$F_b(s_0) = V_b(s_0) + F(s_0) + C \tag{3.13}$$

from which one has:

$$F(s_0) = F_b(s_0) - V_b(s_0) + C \tag{3.14}$$

that means that, from a biased simulation, one can retrieve the correct free energy just by knowing the bias and the histogram in the biased ensemble. So, if one can find a function that compensates the free-energy landscape, then the system is free to travel in the space of CVs and the histogram can be acquired with accuracy since all the metastabilities are removed. This is crucial for many enhanced sampling methods, e.g., Metadynamics (MetaD) [3] that relies exactly on an iterative procedure aimed to produce the compensating potential.
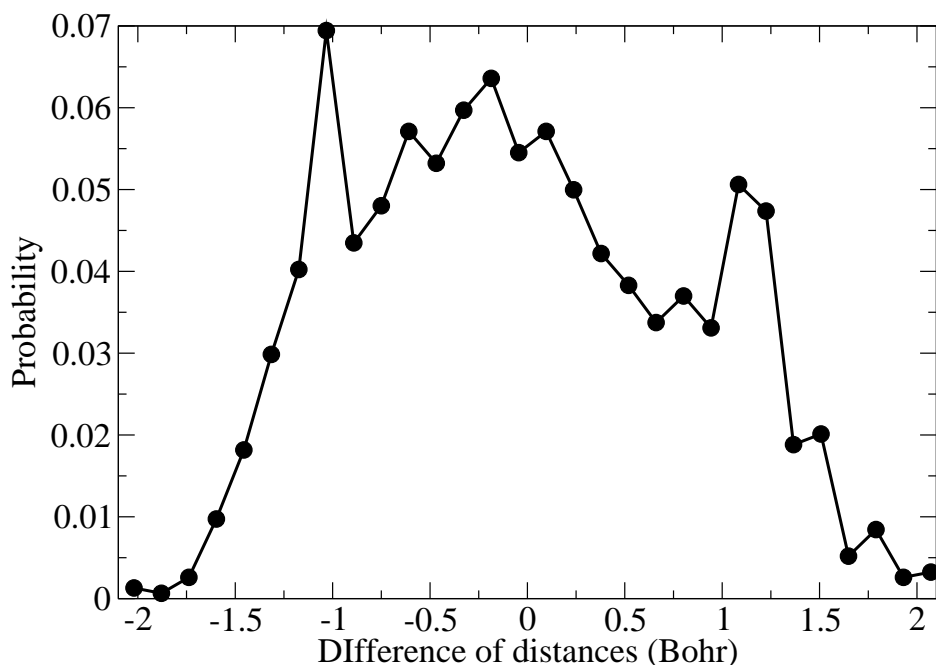
Figure 3.2: A sketch of the distribution from the umbrella potential for the SN2 reaction.

Let's try this approach on this case. Please consider that, in order to obtain a good-quality estimate of the free energy, one should have a long trajectory and check the convergence, the easiest way being the comparison of histograms from the first half of the trajectory and the second half of the trajectory. First we retrieve the probability from the trajectory. This can be done with the awk script provided.

```
grep -v FIELDS COLVAR >f1
awk '{print $1,$4}' f1 |./distribution.awk  >dist
```

then the free energy can be calculated with one liner in gnuplot

```
p "dist" u 1:(-(0.597)*log($2)-0.5*0.01*627.51*($1*$1)) w lp
```

that aim to plot the probability of the biased ensemble `-(0.597)*log($2)` where 0.597 is $k_B T$ in kcal/mol at 300K and `-0.5*0.01*627.51*($1*$1)`
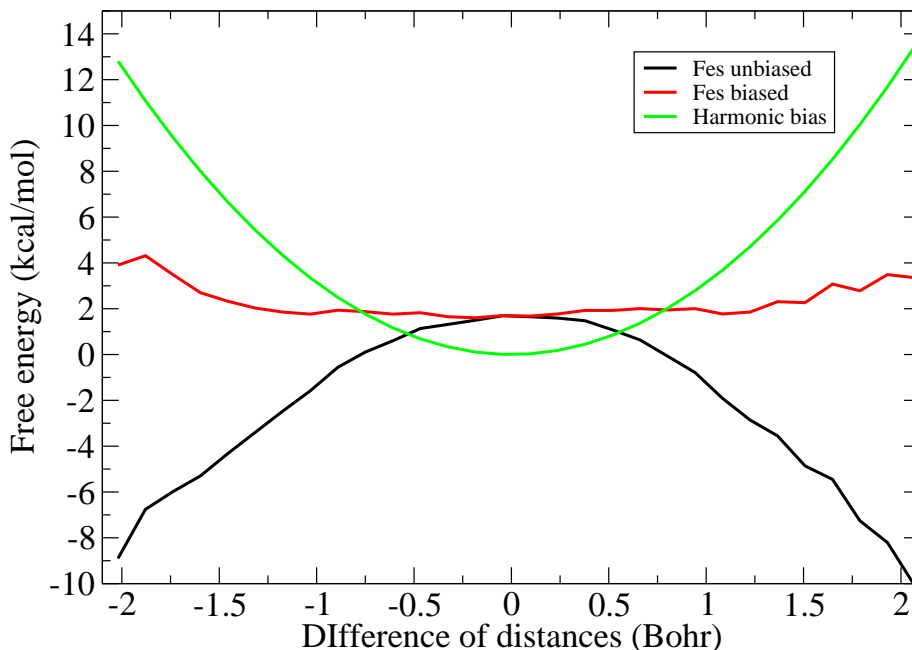
15

Figure 3.3: A sketch of the unbiasing procedure and the results for the SN2 reaction

is the harmonic bias, converted from hartee into kcal/mol (the results of this calculation can be exported through `set term table; rep` commands in gnuplot).

It is worth noting that the calculation involves a logarithm of a probability that around +/-2 bohr goes to zero. This number is actually unreliable and therefore the free energy is trustable only in the region where most of the probability is concentrated.

For this reason generally one imposes many umbrella potentials centered on different points along the reaction coordinate and then combine the probabilities only in the regions where they are reliable. This is achieved by imposing the minimization of the errors. The most applied procedure for doing this is called Weighted Histogram Analysis Method (WHAM) (see [4, 5] and [6]).

16

Another necessary word is to be spent on the plausibility of results: a plot like the one shown in Fig. 3.2 is clearly suspicious since the reaction should be symmetric while the distribution has some spikes. This is generally a sign of bad convergence. This is of course inevitable in the short running time of a tutorial but should be carefully checked in production runs and its origin should be understood. Very often it is the character of the descriptor itself that can produce various aberrations since it may compress the phase space in anisotropic way. This is not a problem as far as it is understood as an issue different from the convergence of the calculations.

Chapter 4

A moving bias on the system

Another possibility that plumed offers is the use of time dependent biases. The simplest of them has the following form:

$$U_{bias}(x,t) = \frac{k}{2}\left(CV(x) - CV_0(t)\right)^2 \tag{4.1}$$

where

$$CV_0(t) = CV_0' + vt \tag{4.2}$$

which represents a harmonic restraint whose center is displaced linearly with time. This might be useful whenever a single umbrella is very far from a desired target configuration in CV space or, more simply, to produce nice movies in which a lot happens in a small time (but at the cost of the real physics).

The `plumed.dat` file then is:

```
g1->
3 4 5 6
g1<-
DISTANCE LIST 1 <g1>
DISTANCE LIST 2 <g1>
DISTANCE LIST 1 <g1> DIFFDIST 2 <g1>
STEER CV 3 TO 2.35 VEL 2.50 KAPPA 0.1
PRINT W_STRIDE 1
ENDMETA
```

where the added line with the directive `STEER` is gently dragging the system to the value of 2.35 bohr with a velocity of 2.5 bohr each 1000 steps. The spring constant is set 0.1 hartree/bohr^2 and the initial center CV_0' is taken as the value of the descriptor at the initial step. in a simulation where there is a time dependent restraint very often, especially for small systems, a more aggressive thermostatting (reduced period) is needed to dissipate the thermal energy externally added by the moving potential. This is true also for metadynamics.

It is important to bear in mind that such "steered MD" simulations may have very little physical meaning and are here reported just to let the reader be familiar with PLUMED and its capabilities. The reason behind their limited usefulness is that in most of the cases these are "out-of-equilibrium" simulations and they are not representative of equilibrium ensemble that is generally of interest for simulations. However, Jarzynski [7] inequality provides a connection between out-of-equilibrium trajectories and equilibrium free energy differences. This inequality consists in:

$$\Delta F = -\beta^{-1} \ln < \exp(-\beta W) > \tag{4.3}$$

where the average is calculated over the work obtained from a number of out-of-equilibrium trajectories. The value of W can be obtained via

$$W = \int_0^{t_s} dt \, \frac{\partial H_b(x,t)}{\partial t} \tag{4.4}$$

where H_b is a modified hamiltonian which contains an additional term, namely:

$$H_b(x,t) = H(x) + U_{bias}(x,t) \tag{4.5}$$

From which it comes naturally that the derivative is simply:

$$\frac{\partial H_b(x,t)}{\partial t} = -vk(CV(x) - CV_0' - vt) \tag{4.6}$$

$$= -vk(CV(x) - \lambda(t)) \tag{4.7}$$

and therefore the integral W can be obtained simply by quadrature summing up all the deviation respect to the position of the mobile

center of the harmonic restraint. In the limit of lots of pulling therefore meaningful free-energy differences can be retrieved.

The `COLVAR` output reads

```
#! FIELDS time cv1 cv2 cv3 XX XX RST3 WORK3
 0.0020 3.604  5.694 -2.090 RST 3 -2.090  0.000000000
 0.0040 3.642  5.669 -2.026 RST 3 -2.088 -0.000011999
 0.0060 3.655  5.662 -2.006 RST 3 -2.085 -0.000029506
 0.0080 3.669  5.658 -1.988 RST 3 -2.083 -0.000051119
 0.0100 3.680  5.658 -1.978 RST 3 -2.080 -0.000075653
 0.0120 3.685  5.665 -1.979 RST 3 -2.078 -0.000100674
```

where in the RST3 column it is reported the center of the moving harmonic potential and in the WORK3 the accumulated work up to that time (in hartree units).

It is rather important to understand that a calculation of such kind can be very challenging since the exponential average in Eq. 4.3 tends to give enormous weight to the trajectories that have less work. Additionally, slow steering speed and tight spring constant in which the descriptor follows closely the center of the moving harmonic potential generally deliver better results but in general, if the descriptor is wrong or the landscape is dominated by entropy and is degenerate (presents multiple pathways), generally it is not advisable to use this technique.

Additionally it is worth mentioning that forward and backward pulling can be combined via the Crooks' theorem [8].

The resulting work profile 4.2 is rather plausible if compared with [1]. The reason for this is that the system presents very few orthogonal degrees of freedom and is highly symmetrical, therefore the pathways are not degenerate and the reaction coordinate is well oriented respect to the real reaction coordinate. This is a rather unique case since in many other systems the value of work can be as large as 10-fold the expected free energy.
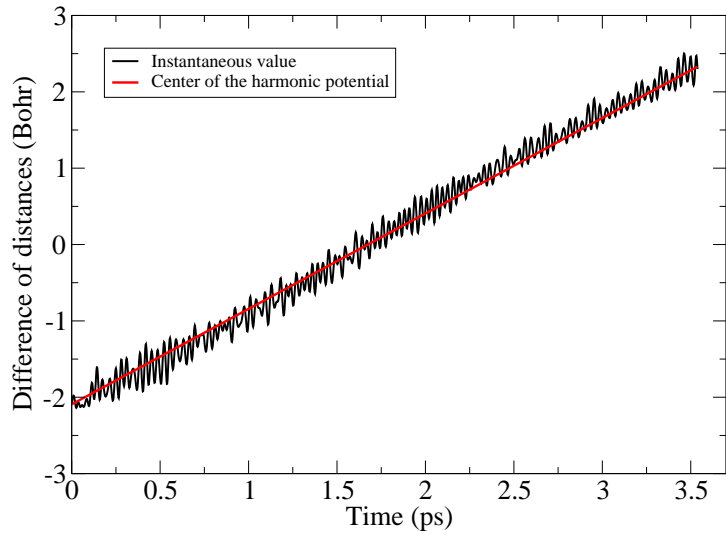
Figure 4.1: The instantaneous value of the descriptor follows closely the red line of the center of the harmonic potential.
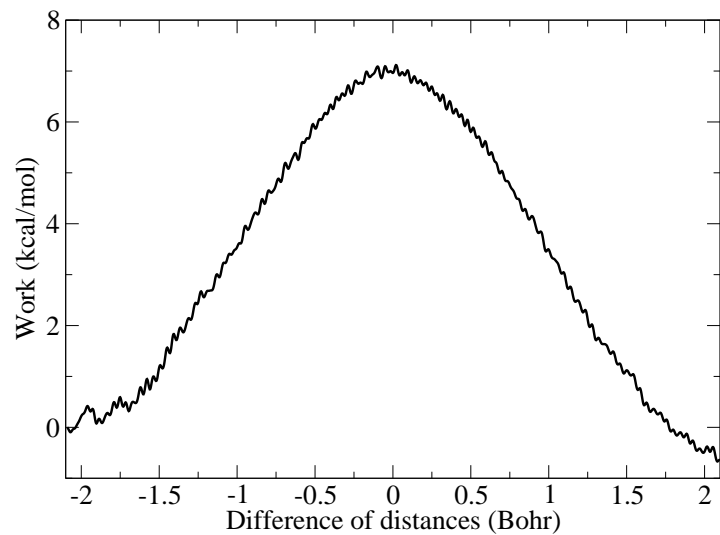

Figure 4.2: The value of work along the difference of distances.

Chapter 5

A metadynamics example

Metadynamics adds an adaptive potential to the simulation that grows with simulation time. Intuitively its functioning relies on Eq. 3.14 and, in order to find the best possible approximation to the perfectly compensating bias it adopts a history dependent potential. This bias potential acts on a restricted number of degrees of freedom of the system $\boldsymbol{s}(x) = (s_1(x), ..., s_d(x))$ often referred to as collective variables or CVs. The metadynamics potential $V(s, t)$ varies with time t and is constructed as a sum of Gaussian functions, or hills, deposited during the simulation:

$$V(\boldsymbol{s}, t) = \int_0^t dt' \omega \exp\left(-\sum_{i=1}^d \frac{(s_i(x) - S_i(x(t')))^2}{2\sigma_i^2}\right), \qquad (5.1)$$

where σ_i is the Gaussian width corresponding to the i-th CV and ω the rate at which the bias grows. In the practice, Gaussians of height equal to W are deposited every τ MD steps, so that $\omega = W/\tau$.

The function of these Gaussian potentials is to discourage the system to revisit the previously visited points in CV space and reach the point in which all the points in CV space are equally sampled in time that delivers a flat histogram which has no contribution on the free energy. Under this assumption the Helmoltz free energy results to be the negative of the deposited bias.

In order to perform a metadynamics simulation we have to:

- Choose wisely the set of CVs to address the problem. This is a long story. To cut it short, CVs a) should clearly distinguish

between the initial state, the final state and the intermediates, b) should describe all the slow events that are relevant to the process of interest, c) their number should not be too large, otherwise it will take a very long time to fill the free energy surface.

- Choose the Gaussian height W and the deposition stride τ. These two variables determine the rate of energy added to your simulation. If this is too large, the free-energy surface will be explored at a fast pace, but the reconstructed profile will be affected by large errors. If the rate is small, the reconstruction will be accurate, but it will take a longer time. The error on the reconstructed FES depends on the ratio W/τ, not on the two parameters alone [9].

- Choose the Gaussian width σ_i. This parameter determines the resolution of the reconstructed FES. The sum of Gaussians reproduces efficiently (*i.e.* in a finite simulation time) features of the FES on a scale larger than σ_i. A practical rule is to choose the width as a fraction (half or one third) of the CV fluctuations in an unbiased simulation. This is not a golden rule, since the value of the fluctuations is not universal but usually depends on the position in the CV space (see Fig. 5.1 for the calculation of the fluctuation on the CVs used for tracking the unbiased dynamics). In particular, although intuitively a larger σ could bring to a faster filling, very often, when the choice is not clear as in 5.1 it is much better to adopt the smaller one since it guarantees a better resolution and ability to fill up both narrow and large wells.

To activate a metadynamics calculation in `PLUMED` you have to use the directive `HILLS`. The deposition stride τ is specified in unit of time steps by the keyword `W_STRIDE`, the height W by `HEIGHT` in internal units of energy of the MD code used. The Gaussian width σ_i must be specified on the line of each CVs with the keyword `SIGMA`.
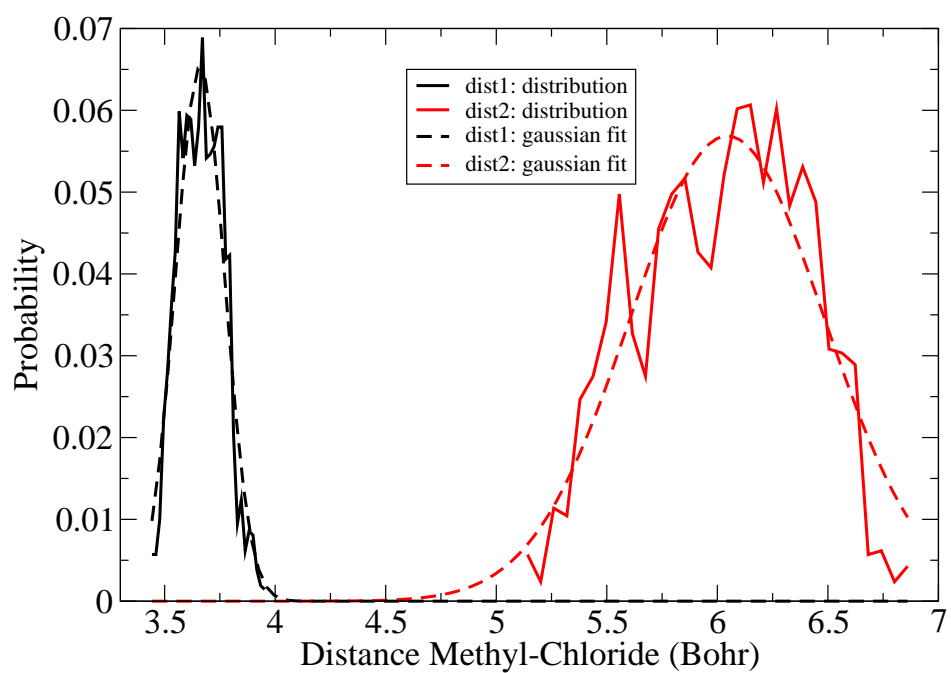
Figure 5.1: The fluctuation calculated during the unbiased dynamics for the two methyl-chloride distances. A gaussian fit is overlapped with the distribution calculated. The $\sigma=0.11$ bohr for the black profile and $\sigma=0.440$ bohr for the red profile.

Generally metadynamics is to be preferred for multidimensional problems that require no more than 3/4 CVs. The reason is that in metadynamics one does not need to explore all the space in CV to have an estimate of the free-energy landscape in a region of interest since metadynamics start exploring and filling with potential the regions at low free energy, which are the most interesting. In other methods, like WHAM combined with e.g. umbrella sampling, one has to define a priori the grid and the systematic exploration becomes cumbersome in a high-dimensional space.

For the usual case of the SN2 reaction one could think that, instead of exploring the difference of distances, one could be interested in exploring independently the two methyl-chloride distances (so that we show how to set up a multidimensional calculation). Nevertheless, we expect that we are not really interested in unbound states where the methyl-Cl distance is above 6.5 Bohr. This require to introduce an artificial soft wall to prevent the system from visiting nonreactive regions which might be trivially low in energy.

A possible input therefore reads:

```
DISTANCE LIST 1 <g1> SIGMA 0.11
g1->
3 4 5 6
g1<-
DISTANCE LIST 2 <g1> SIGMA 0.11
UWALL CV 1 LIMIT 7. KAPPA 0.5
UWALL CV 2 LIMIT 7. KAPPA 0.5
HILLS HEIGHT 0.00047 W_STRIDE 50
PRINT W_STRIDE 2
ENDMETA
```

which present a `SIGMA` value for each of the CVs on which the metadynamics is active (it could be a subset of all the CVs) and a specific line starting with `HILLS` that specify the height and the deposition time. Additionally we included the confining potential, denoted by `UWALL` (upper wall) which impose a quartic wall in order to prevent

the system to escape from the reactive region.

The typical `COLVAR` files that one obtains is similar to this:

```
#! FIELDS time cv1 cv2 vbias vwall
 0.8240  3.723740474  6.868711550 0.000029924  0.000000000
 0.8280  3.691187802  6.931429351 0.000013739  0.000000000
 0.8320  3.639273674  6.998560142 0.000004823  0.000000000
 0.8360  3.558399168  7.083118628 0.000000000  0.000023865
 0.8400  3.478637841  7.159044915 0.000000000  0.000319926
 0.8440  3.424739803  7.208958444 0.000000000  0.000953256
 0.8480  3.428390769  7.207133491 0.000000000  0.000920389
 0.8520  3.493585587  7.155388196 0.000000000  0.000291502
 0.8560  3.601809461  7.072873461 0.000000000  0.000014101
 0.8600  3.707646538  6.990032405 0.000002236  0.000000000
 0.8640  3.779941165  6.922057261 0.000001894  0.000000000
 0.8680  3.815026536  6.863783677 0.000004640  0.000000000
 0.8840  3.708208013  6.668182930 0.000348923  0.000000000
```

in which two additional fields are present `vbias` which is the potential energy due to the hills potentials of metadynamics and `vwall` which is the potential exerted by the quartic wall (in fact is acting only for values of the second CV larger than 7).

In addition to this, `PLUMED` produces a `HILLS`. For two-dimensional Gaussian potentials each line consists of time, the position of the centers in the two dimension, the width of the Gaussian potential in the two dimensions (in bohr since distances are used as CVs), the height (in hartree) and a multiplicative factor that is only useful when well-temperered metadynamics[14] is enabled.

```
 0.100 3.797  5.972  0.110  0.110  0.00047  0.000
 0.200 3.731  6.191  0.110  0.110  0.00047  0.000
 0.300 3.668  5.823  0.110  0.110  0.00047  0.000
 0.400 3.769  5.619  0.110  0.110  0.00047  0.000
```
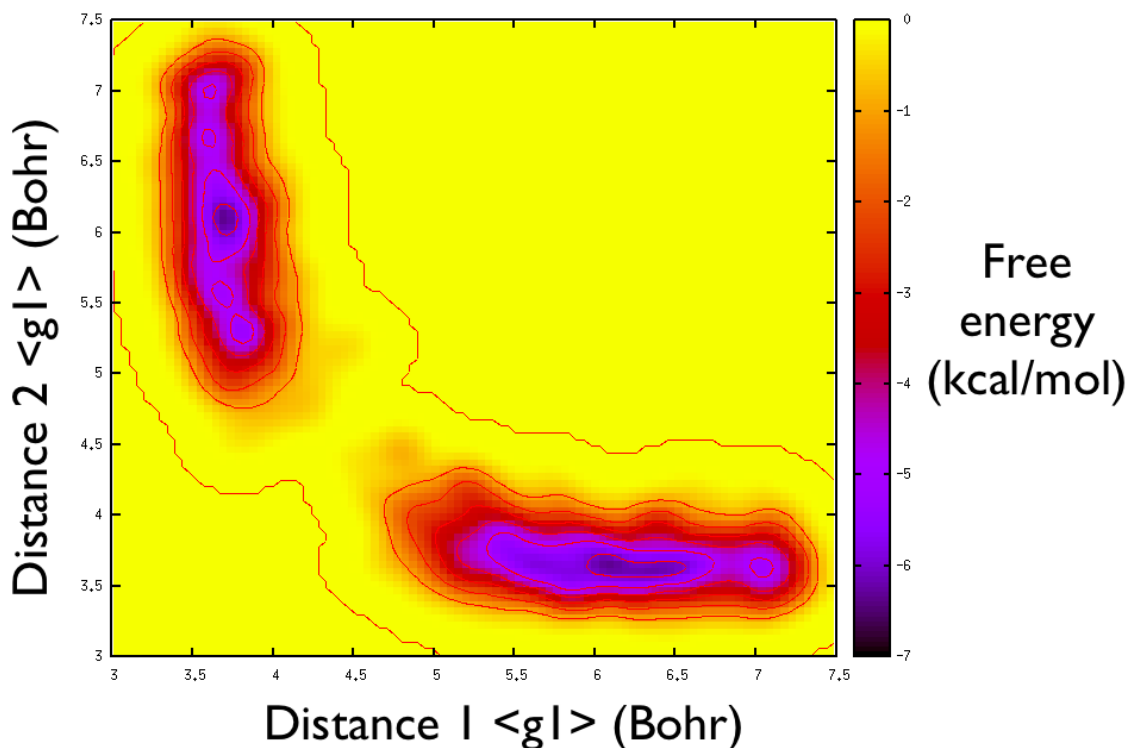
Figure 5.2: The partial 2-dimensional free energy surface after 15 ps of metadynamics. Each isosurface represents kcal/mol.

During the simulation one can inspect the instantaneous estimate of the free energy by using the `sum_hills` program included in PLUMED. which integrate the HILLS file over a domain and produce a surface called `fes.dat`. This is invoked with

```
sum_hills -ndim 2 -ndw 1 2 -file HILLS
```

Note that, after 15 ps of metadynamics, the system has still not crossed the transition state that should be located at 4.5 bohr in both dimensions. This reflects the intrinsic difficulty of a multidimensional calculation that should be taken into account whenever setting a multidimensional free-energy calculation (the same holds also for methods other than metadynamics, as WHAM). In this case a better choice of CVs, like the difference of distances shown before, allow to reduce the

dimensionality of the problem and allow a faster exploration.

Another important fact that should be taken into account when running a metadynamics run is that it allows for a "prior", which means that at the beginning the free-energy landscape is assumed flat (no hills). With the evolution of the bias potential, the estimate of free energy landscape is progressively revised and improved. Nevertheless, those regions for which the potential is zero means that they are not sampled by metadynamics and therefore should not be interpreted. For other methods based on the logarithm of probabilities this would turn into the negative logarithm of zero which is infinite and therefore cannot be interpreted.

After more than 20 ps of metadynamics the transition state is crossed and the system spends other more than 20 ps to fill the second basin. After that point the system spends considerable less time for each transition and the transition state is repeatedly recrossed. Within this regime the metadynamics can be considered to provide a valid representation of the free energy. Additionally it is worth remembering that metadynamics samples the lowest basin first, therefore it scales with a computational effort which is less then a power law of the dimensionality used in the metadynamics itself if the landscape is not too complex.
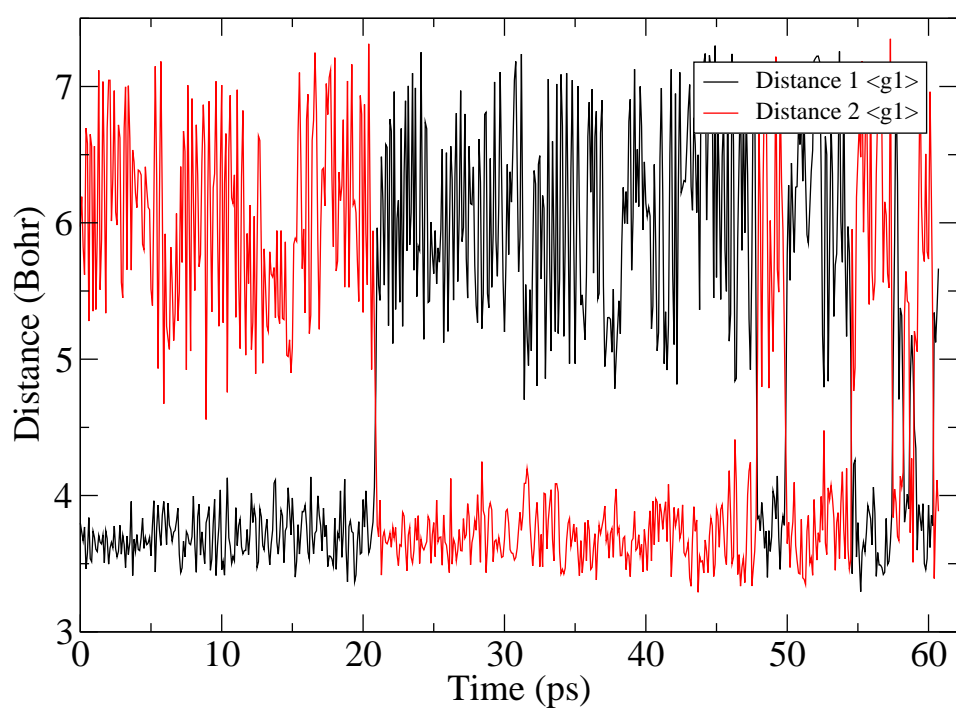
Figure 5.3: The timeline of the two distances during metadynamics. The two main basins are filled and then the dynamics proceeds in an almost diffusive regime.

Chapter 6

Other features and conclusions

This brief summary is intended only to give a first approach to the free-energy calculation problem with `PLUMED`. We recommend, before embarking any serious project, to devote time to read the literature and in particular the classic books [10, 11, 12] to acquire a solid background in both simulation techniques and statistical mechanics. `PLUMED` itself includes many more techniques than those outlined here. Among all it is worth citing multiple walkers metadynamics [13], well-tempered metadynamics [14], d-AFED [15],well-tempered ensemble[16] reconnaissance metadynamics [17] to name a few that are implemented an ready to use in conjuction with FHI-aims (other replica-exchange-based methods will be available soon on a script based method as the plain replica exchange MD already available for FHI-aims, see manual). Additionally, flexible steering schedules are available via `STEERPLAN` directives. More than this, `PLUMED` allows to have plenty of CVs available: from angles to torsion to coordination number up to path collective variables [18]. The investigation of the optimal space for projection of the free energy is a very active field indeed.

This is rather central every time one wants to perform a free-energy calculation and should be carefully explored and analyzed. Literature very often helps a lot in this respect.

Bibliography

[1] S. Yang, P. Fleurat-Lessard, I. Hristov, T. Ziegler, Free energy profiles for the identity s(n)2 reactions cl-+ch3cl and nh3+h3bnh3: A constraint ab initio molecular dynamics study, J Phys Chem A 108 (43) (2004) 9461–9468.

[2] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, Journal Of Chemical Physics 126 (1).

[3] A. Laio, M. Parrinello, Escaping free energy minima, Proc. Natl. Acad. Sci. USA 99 (2002) 12562–12566.

[4] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, Multidimensional free-energy calculations using the weighted histogram analysis method, J. Comput. Chem. 16 (1995) 1339–1350.

[5] B. Roux, The calculation of the potential of mean force using computer-simulations, Comput. Phys. Comm. 91 (1995) 275–282.

[6] A. Grossfield, WHAM.
`http://membrane.urmc.rochester.edu/Software/WHAM/WHAM.html`.

[7] C. Jarzynski, Nonequilibrium equality for free energy differences, Phys. Rev. Lett. 78 (1997) 2690–2693.

[8] G. Crooks, Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems, J. Stat. Phys. 90 (5-6) (1998) 1481–1487.

[9] A. Laio, A. Fortea-Rodriguez, F.Gervasio, M. Ceccarelli, M. Parrinello, Assessing the accuracy of metadynamics, J. Phys. Chem. B 109 (2005) 6714–6721.

[10] M. P. Allen, D. J. Tildesley, Computer Simulation of Liquids, Oxford University Press, New York, 1987.

[11] D. Frenkel, B. Smit, Understanding molecular simulation, Academic Press, 1996.

[12] D. Chandler, Introduction to modern statistical mechanics, Oxford University Press, 1987.

[13] P. Raiteri, A. Laio, F. Gervasio, C. Micheletti, M. Parrinello, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, J. Phys. Chem. B 110 (2006) 3533–3539.

[14] A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method, Phys. Rev. Lett. 100 (2) (2008) 020603.

[15] J. B. Abrams, M. E. Tuckerman, Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations, J. Phys. Chem. B 112 (2008) 15742–15757.

[16] M. Bonomi, M. Parrinello, Enhanced sampling in the well-tempered ensemble, Phys. Rev. Lett. 104 (2010) 190601.

[17] G. A. Tribello, J. Cuny, H. Eshet, M. Parrinello, Exploring the free energy surfaces of clusters using reconnaissance metadynamics, J. Chem. Phys 135 (11) (2011) 114109.

[18] D. Branduardi, F. L. Gervasio, M. Parrinello, From A to B in free energy space, J. Chem. Phys. 126 (5) (2007) 054103.