Calculating Free Energy Differences from Molecular Simulation
Hands-on notes for the exercises

Daniel Markthaler and Niels Hansen*

Institute of Thermodynamics and Thermal Process Engineering,
University of Stuttgart, D-70569 Stuttgart, Germany

Version: January 10, 2026

1 Overview

In the computational exercises the dimerization of two united-atom methane molecules will be investigated using different types of equilibrium free energy calculations. The aim of the tutorial is illustrated in the thermodynamic cycle in Figure 1. The simulations can be divided into those regarding the two methanes in an empty computational box ("vacuum simulations") and simulations conducted in a box of SPC water ("water simulations"). Two basic types of free energy calculations will be used: (1) *Umbrella Sampling* (US) representing the family of physical-path-based methods and (2) *Incremental Alchemical Decoupling* representing the family of alchemical perturbation methods. This tutorial will guide you through the individual steps for obtaining a *potential of mean force* (PMF) along a characteristic order parameter (in this case the methane-methane distance) using US as well as the determination of a hydration free enthalpy from alchemical decoupling. In addition to all the required input files and scripts, also the simulation output relevant for the analysis will be provided. Since it is assumed that you are familiar with the basic GROMACS features the instructions are represented as short as possible in a cook book-like recipe. Care was taken of writing the simulation parameter files (`.mdp`) clear and well commented and to establish an easy to understand automatized workflow using simple (rather than aesthetic) scripts. The exercises were designed with the purpose to gather freely available useful tools for the simulation setup and analysis distributed over the community which should be easy to apply to your own real world problem.

*hansen@itt.uni-stuttgart.de

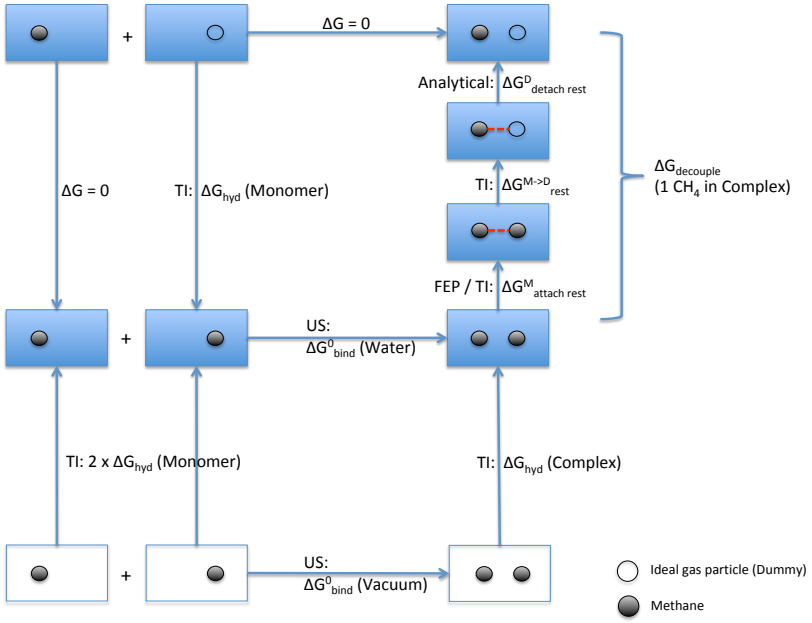


Figure 1: Underlying thermodynamic cycle for the dimerization process of two methane molecules in water and vacuum. The individual free energy terms follow either from Umbrella Sampling (US), Thermodynamic Integration (TI), Free Energy Perturbation (FEP) or analytical calculation (see main text).

2 Prerequisites

In the exercises the MD package **GROMACS 2023.5**[1] patched with the open source library **PLUMED**[2] is used. The analysis is carried out with the GROMACS built-in version for the *Weighted Histogram Analysis Method* (WHAM) as well as with external software providing a Python3 code for the *Umbrella Integration* (UI) method, **umbrella_integration.py**,[3] and the **alchemlyb**[4] library.

3 Umbrella Sampling (US)

This section describes the workflow how to setup an Umbrella Sampling simulation for the two methane molecules in vacuum and how to obtain a potential of mean force (PMF) from the produced data with different methods (WHAM [1], UI [2][3], MBAR [4]). The transfer to the corresponding simulation in water

[1] `https://manual.gromacs.org/2023.5/download.html`
[2] `https://github.com/plumed/plumed2/`
[3] `https://github.com/M-R-Schaefer/umbrella_integration`
[4] `https://github.com/alchemistry/alchemlyb`

should be straightforward. All the required files (topology, initial configuration, simulation parameter files) are located in:
`~/Simulations/day_1_2/US_CH4Dimer_Vacuum/`
distributed over different subfolders labeled with prefixes from 0 to 4 representing the chronological order of the workflow. GROMACS specific commands for each step can be looked up in the respective `command` file within each subdirectory. Simulation parameter files (`.mdp`) are located under `mdp_files`.

3.1 Starting Configuration and Topology

`0_coord` contains the initial coordinates `start.gro` for the two methane molecules oriented along the z-axis. The folder `0_topo` contains the topology files (system topology `topol.top` and molecular topology `2CH4.itp`) and a subfolder containing position restraint files (`posre_all.itp`, `posre_us1D.itp`, `posre_us3D.itp`). The force field parameters (located in `gromos54a7.ff`) are taken from the GROMOS 54a7 force field parameter set [5]. For practical reasons the Lennard-Jones-Energy parameter ϵ for the CH_4-CH_4 interaction was increased by a factor of 20 compared to the original GROMOS 54a7 force field parameter set to make the methane model more "sticky". Since we overwrite a given force field parameter a warning is issued. Moreover, recent versions of GROMACS issue a warning when the GROMOS force field is used which is related to the fact that the GROMOS force fields was parametrized using a twin-range cutoff scheme, in which longer ranged nonbonded forces and energies are updated less frequently than shorter ranged ones. The effect of this choice on thermodynamic, structural, and dynamic properties has been discussed recently [6]. In the realm of this exercise the two warnings can be ignored which is achieved by the usage of `-maxwarn 2` for every call of the GROMACS preprocessor `grompp`. However, if you want to use the GROMOS force field within the GROMACS software in your own research project it is recommended to check whether the standard settings in GROMACS regarding treatment of nonbonded interactions affect any relevant properties.

3.2 Defining the Simulation Box

Go to the subdirectory `1_box`. A rectangular box with an uniform length of 3.2 nm can be defined using `editconf`:

```
$ gmx editconf -f ../0_coord/start.gro -o box.gro -box 3.2
```

3.3 Energy Minimization

For removing clashes between the solute molecules, an energy minimization is conducted first. Go to the subdirectory `2_min_vac` and do:

```
$ gmx grompp -f ../mdp_files/em.mdp -c ../1_box/box.gro
-p ../0_topo/topol.top -o min.tpr -maxwarn 2
$ gmx mdrun -deffnm min
```

Note that performing a simulation in GROMACS is always composed of a two-stage process: first the GROMACS pre-processor `grompp` is called. The generated binary portable tpr-file which contains everything needed for the simulation (topology, parameters, coordinates and velocities) is then used by `mdrun`. Be aware of the usage of the `maxwarn` flag during the call of `grompp` to ignore (i) the reported warning due to the conflict of the original CH4-atomtype, taken from the included GROMOS 54a7 parameter set (in `topol.top`) and the redefined one from the include topology file (`2CH4.itp`), and (ii) the warning about the GROMOS force field in general.

3.4 NVT Equilibration

Go to the subdirectory `3_eq_nvt`. For assigning particle velocities according to the temperature of interest a short NVT equilibration is performed (even though one can not expect to get instantaneous or even averaged temperature values from the simulation which are close to the target value for only two particles). The stochastic dynamics integrator is used for thermostatisation with $T_{\mathrm{ref}} = 300$ K and $\tau_{\mathrm{T}} = 0.1$ ps (see simulation parameters file `eq_nvt.mdp`). The equilibration can be performed by:

```
$ gmx grompp -f ../mdp_files/eq_nvt.mdp -c ../2_min/min.gro
-p ../0_topo/topol.top -o nvt_eq.tpr -n ../index.ndx
-maxwarn 2
$ gmx mdrun -deffnm nvt_eq -plumed ../plumed.dat
```

In the simulation parameters file (`nvt.mdp`), two non-standard energy groups (`CH4_1`, `CH4_2`) are referenced, corresponding to the two individual molecules in order to evaluate the interactions between them. These two groups have to be provided by the index file `index.ndx`. The PLUMED input file `plumed.dat` required for writing out collective variables such as the intermolecular distance is read in via the flag `-plumed`. Explanation about the PLUMED syntax is given in the next section.

3.5 US Setup

We will employ two approaches to perform the umbrella sampling (US) simulations. The first one uses PLUMED and the second one uses the GROMACS pull code following the GROMACS tutorial by Lemkul [7].

For conducting the US simulations with PLUMED go to the subdirectory `4_US` which contains two bash scripts for setting up and starting the simulations (`prepare_sim.sh`, `run_sim.sh`). An additional script can be used for extending the simulation time of every window (`extend_sim.sh`). First a short equilibration run is performed such that the system has time to adjust the actual value of the order parameter to the desired equilibrium value of this particular window (as specified by the applied bias potential). The output data as used for the analysis are sampled within a subsequent production phase. The relevant `mdp`-files (`us_eq.mdp`, `us_prod.mdp`) for the equilibration and production phase are located under `mdp_files`. The whole US specification (definition of the order parameter, equilibrium value, type and force constant of the bias potential)

is done via PLUMED in the form of a quite simple input file which will be automatically generated by `prepare_sim.sh`:

$ bash prepare_sim.sh

It will generate a number of folders named `w_XY` each one containing a PLUMED input file `plumed.dat` where XY stands for the equilibrium value of the distance restraining bias potential. The number and the spacing of the umbrella windows can be controlled by the bash script:

```
window=($(LC_NUMERIC="C" awk 'BEGIN{ for (i=0.35; i < 1.20; i+=0.05) printf("%.2f\n", i); }'))
length=${#window[@]}
```

Using the script as it is will generate 17 equidistant windows with reference values for the order parameter ranging from 0.35 to 1.15 nm with an uniform spacing of 0.05 nm. The generated PLUMED files all have the same structure and only differ in the equilibrium value (`AT`) of the distance restraining potential:

```
#PLUMED INPUT FOR UMBRELLA SAMPLING
#Definition of Groups
CH41: GROUP ATOMS=1
CH42: GROUP ATOMS=2
#Definition of Collective Variables
Dist_CH41_CH42: DISTANCE ATOMS=CH41,CH42
Dist_CH41_CH42_comp: DISTANCE ATOMS=CH41,CH42 COMPONENTS
#Definition of Restraints
#Force Constant KAPPA, Equilibrium Value AT
disrest: RESTRAINT ARG=Dist_CH41_CH42 KAPPA=500.0 AT=0.35
#Write to File
PRINT ARG=Dist_CH41_CH42,Dist_CH41_CH42_comp.x,Dist_CH41_CH42_comp.y,Dist_CH41_CH42_comp.z,
disrest.bias STRIDE=250 FILE=COLVAR_0.35
ENDPLUMED
```

Note that PLUMED specific keywords like `GROUP`, `ATOMS`, `DISTANCE`, `COMPONENTS`, `RESTRAINT`, `PRINT` etc., called "actions", are written in capitals. In the third and fourth line, both methanes are called by their ID's and assigned to the variable names `CH41` and `CH42`. Next PLUMED calculates the radial distance between the defined groups together with the three Cartesian components of the separation vector. Each component can be addressed by the point operator like `Dist_CH41_CH42_comp.z` for instance. The 10th line defines a harmonic restraint acting on the radial distance (order parameter!) with a force constant `KAPPA` of 500 kJ mol^{-1} nm^{-2} and an equilibrium value `AT` of 0.35 nm. The quantities of interest (radial distance, three Cartesian components, value of the bias potential) are written to a file named `COLVAR_0.35` every 250'th step with the `PRINT` command. For the analysis of the US runs only the time series of the order parameter together with the applied force constant and the equilibrium value for every umbrella window is needed. Note that one could also use the GROMACS built-in Pull Code for Umbrella Sampling. However PLUMED offers you more flexibility in the definition of order parameters and the PLUMED syntax is in our opinion more intuitive.

For starting the simulations just run the second bash script `start_sim.sh`:

$ bash start_sim.sh

If you have a look at the script you will see that every window is simulated sequentially after the other. After a short equilibration run (`us_eq.mdp`) during

which the system has time to adjust the actual value of the order parameter to the desired equilibrium value of this particular window the real production run (`us_prod.mdp`) is performed. The whole simulation with the given settings should take up to a couple of minutes on a dual core machine. Note that all of the windows can be simulated independently which means that if you have access to a computer cluster they can be run in parallel. Using a cluster should require only minor modifications of the job and running script for which templates can be found in the subfolder `scripts_cluster`.

One could also imagine performing a (quasi) one-dimensional setup in which the two methanes are pulled apart on a linear path (e.g. z-axis). This can be realized for instance by using harmonic position restraints in all three directions acting on the first methane making it immobile. On the second methane only two orthogonal position restraints (x,y) are applied. The position restraints can be activated in the mdp-file by the line:

define = –DPOSRES_US1D

The respective force constants ($2.5{\cdot}10^4$ kJ mol^{-1} nm^{-2}) are read from `0_topo/posres_files/posre_us1D.itp` . In GROMACS, the reference values for the position restraints are taken from the initial coordinate file (in our case `3_eq_nvt/nvt_eq.gro`). In this 1D-setup the Cartesian z-component of the vector pointing from the first to the second methane (`Dist_CH41_CH42_comp.z`) would be used as the order parameter rather than the radial distance itself.

The US simulations with the GROMACS pull code are carried out in subdirectory `5_US_pull`. Before running the pull simulation have a look in the input file `pull.mdp` located in the `mdp_files` directory and consult ref. [7] for an explanation of the pull options. The actual pull simulation is started with

```
$ gmx grompp -f ../mdp_files/pull.mdp -c ../1_box/box.gro
-p ../0_topo/topol.top -o pull.tpr -n ../index.ndx -maxwarn 2
$ gmx mdrun -deffnm pull
```

When the simulation is complete, measure the distance between the two methane molecules over time with the `distance` command:

```
$ gmx distance -s pull.tpr -f pull.xtc -n ../index.ndx -select
'com of group 3 plus com of group 4' -oall
```

and plot the result with `xmgrace`. Now we have to select starting structures for the umbrella sampling windows. We will be sampling along the reaction coordinate from 0.35 nm to 1.15 nm in intervals of 0.05 nm. Therefore, we need to identify snapshots that have methane-methane distances that are as close as possible to the desired values. We use a Python script provided by Lemkul [7] for that purpose. A total of 17 windows define the desired order parameter. Since we simulate in vacuum we do not need to equilibrate the system at each window but can right away start with the umbrella sampling. The simulations are started sequentially using a bash script:

```
bash run_umbrella_sampling.sh
```

3.6 Potential of Mean Force (PMF)

To obtain a PMF from the generated collective variable timeseries we use three different estimators: MBAR, UI and WHAM. While MBAR and UI are used to analyze the PLUMED simulations, WHAM is used to analyze the US simulation with the GROMACS pull code. Even though these methods are related to each other they use different information from the sampled data and error estimates. It is always good practice to check if the different estimators yield the same PMF. Theoretical background information can be found in the lecture slides and the aforementioned key publications. The prepared scripts for MBAR and UI are located in `4_US/ana`. Note that the subdirectory WHAM, also present there, is only kept for legacy reasons since it works only with older GROMACS versions.

The usage of the UI Python implementation requires some transformation of the COLVAR files to the specified format: time t_i | value of order parameter at t_i | reference value of current window | respective force constant. The subfolder `UI` contains the script `prepare_UI_python.py` which will do the job for you. Before using it, one should check that the path to the folder containing all the COLVAR files is set correctly. The values for the applied force constant (`forceconst_DR`) as well as the array containing the reference values for the distance restraining (`refpos`), which should match the values and the format of the respective subfolders `w_XY` for every umbrella window, are hard-coded. Calling

```
$ python prepare_UI.py
```

will produce the inputfiles `UI_XY.out` in the specified format required for the umbrella integration program. Now the UI code can be executed via:

```
$ umbrella-integration -t 300 -i UI_*out -n 100 -o pmf_UI.dat
-pp pmf_UI.eps -pd derivatives.eps -ph histos.eps
```

The flags `-t`, `-n` and `-m` define the simulated temperature, the number of bins and the borders for the binning respectively. All available options can be displayed by calling `umbrella_integration.py` with `-h`. A plot of the PMF along with error estimates (`pmf_UI.dat`, `pmf_UI.eps`) will be produced together with a histogram plot (`histos.eps`).

The script (`umbrella_sampling_MBAR.py`) which calls the Python implementation of the MBAR estimator (`alchemlyb` has to be installed via `pip install alchemlyb`) is located in the subfolder `MBAR`. It should be checked that the path to the COLVAR files is set correctly. Values for temperature, force constant, number of bins and the equilibrium values of the umbrella windows are hard-coded in the script. Calling it via:

```
$ python umbrella_sampling_MBAR.py
```

yields the resulting PMF (`pmf_mbar_histo.out`) along with error estimates as well as the PMF (`pmf_mbar_kde.out`) employing a kernel density approximation. The details on the use of MBAR to calculate free energy surfaces are

discussed by Shirts and Ferguson [8].

The analysis of the GROMACS pull simulations with WHAM takes place in `5_US_pull`. First, we have to write a list of files that is needed by `gmx wham`. For that purpose run the Python script `write_wham_files.py`:

```
$ python write_wham_files.py
```

Next, `gmx wham` can be executed:

```
$ gmx wham -it tpr-files.dat -if pullf-files.dat -bsprof -bsres
-nBootstrap 200
```

More details on the program options can be found in the program description (`gmx wham -h`) or in the literature [9].

Comparison between the PMF profiles calculated by the different approaches is shown in Figure 2. It can be seen that the three analysis methods yield the same PMF within error bars. Now calculate the PMF for two methane molecules solvated in water. The procedure is very similar but contains some additional steps. Note that before starting the US simulations you have to generate an index file yourself using `gmx make_ndx`. In the directory `7_US` you will find bash scripts that can be used to simulate each window as a separate job on a cluster. Take a look at the `README_EULER.txt` file. Also note that for the pull simulations in water, the configuration at each umbrella window should be equilibrated before starting the actual umbrella sampling simulation (`8_US_pull`).

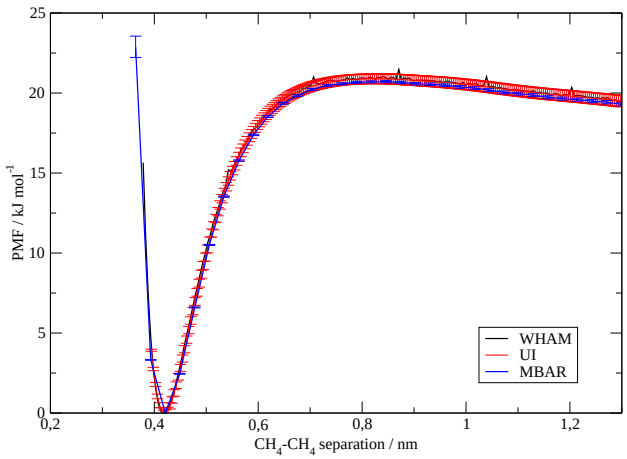


Figure 2: Comparison of calculated PMF's for two methanes in vacuum from different estimators (WHAM, UI, MBAR). Each window was simulated for 10 ns at 300 K.

4 Alchemical Decoupling

This section describes the workflow how to obtain the hydration free enthalpy for the example of a single methane molecule using incremental alchemical decoupling. All the required files are located in:
`~/Simulations/day_3/Solvation_1CH4/`
The other decoupling simulations, delivering the remaining free energy contributions of the thermodynamic cycle in Figure 1 (hydration free enthalpy for the dimerized methane complex, decoupling of one methane within the complex), are prepared in an identical way and the transfer should be straightforward.

4.1 Preparing the Simulation

The topology files and starting coordinates are located in the subfolders `0_topo` and `0_coord` respectively. Simulation parameter files (`.mdp`) together with the GROMACS specific command files for solvation, energy minimization, NVT- and NpT-equilibration (steps 2 to 5) are contained in the corresponding subfolders and should be self-explaining for MD-experienced users.

4.2 Performing Decoupling

We will perform the decoupling within 9 λ-points by perturbing the solvated methane molecule (initial state A) into an ideal gas particle (end state B). The folder `6_Decoupling` contains a Perl script `write_sh.pl` for the automated generation of a set of job scripts (`job_0.sh`, ..., `job_8.sh`), each for every λ-point based upon the template `job.sh`:

$ perl write_sh.pl job.sh

Note that in principal you should only have to adapt the paths to the binaries (PLUMED, GROMACS) in `job.sh`:

```
#!/bin/bash
source $HOME/YOUR_PATH/programs/plumed2/sourceme.sh
source $HOME/YOUR_PATH/programs/gromacs-2023.2_plumed2_INSTALL/bin/GMXRC
```

If you have a look at the job script you will notice that after generating a new directory `Lambda_XY` for every λ-point it will perform again a short energy minimization, followed by two equilibrations (NVT, NpT) before the real production starts. The two additional post-processing scripts `run_BAR.sh` and `collect_dHdl.sh` are for the analysis of the results using the GROMACS implementation of the BAR method as well as various estimators implemented in the alchemlyb library. The folder `mdp_files` contains the simulation parameter files for each stage (`EM`, `NVT`, `NPT`, `Production_MD`) for every of the 9 λ-points. Since we are doing free energy calculations, the mdp-files contain an additional free energy calculation block which is important to have a look at. For example `mdp_files/Production_MD/md_3.mdp`:

```
; Free energy control stuff
free_energy             = yes
init_lambda_state       = 0
delta_lambda            = 0
calc_lambda_neighbors   = -1   ; Calculate Potential Energy Differences between ALL neighbors
; Vectors of lambda specified here
; init_lambda_state          0    1    2    3    4    5    6    7    8
vdw_lambdas             = 0.00 0.20 0.38 0.48 0.55 0.62 0.70 0.85 1.00
```

```
coul_lambdas            = 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
; Options for the decoupling
sc-alpha                = 0.5
sc-coul                 = no    ; linear interpolation of Coulomb
sc-power                = 1
sc-sigma                = 0.3
couple-moltype          = CH4   ; name of moleculetype to decouple
couple-lambda0          = vdw   ; state A: only van der Waals interactions
couple-lambda1          = none  ; state B: turn off everything
;couple-intramol        = no
nstdhdl                 = 500
```

For detailed explanations of each option one should look at the GROMACS user manual or at the mdp-options of the online User Guide[5]. Just to comment a few: the lambdas (Van-Der-Waals (VDW)-interactions, electrostatics, restraints, ...) have to be written as an array starting from 0.0 and ending at 1.0 with arbitrary spacing which does not have to be equidistant as it is the case above. Here only the VDW-interactions are perturbed since there are no charges on the methane and the mass stays the same over all lambdas. The assigned value to `init_lambda_state` stands for the index of the lambda array and thus defines which λ state we are simulating. In this case it is 3 (the corresponding mdp-file is designated appropriately as `md_3.mdp`) and corresponds to `lambda[3] = 0.48`. `init_lambda_state = 0` means the initial state A (fully interacting methane). Since MBAR requires the potential energy differences between all states (not only neighboring ones as for BAR) for input, one have to set `calc_lambda_neighbors = -1` (!). `couple-moltype` is the name of the `[moleculetype]` looked up in the topology file to be decoupled. `couple-lambda0` and `couple-lambda1` define the activated non-bonded interactions for both of the end states. Here only VDW-interactions are present in the initial state ($\lambda = 0$) whereas in the fully decoupled state ($\lambda = 1$) all the interactions with the surrounding are switched off. `nstdhdl` sets the frequency for writing out $\partial U/\partial\lambda$ and ΔU (U stands for the system's potential energy) to the outputfile `dhdl.xvg`. These `dhdl.xvg` files are needed as input for the analysis with BAR and MBAR.

After generating the job files for every λ-point as described above one can run the simulations via:

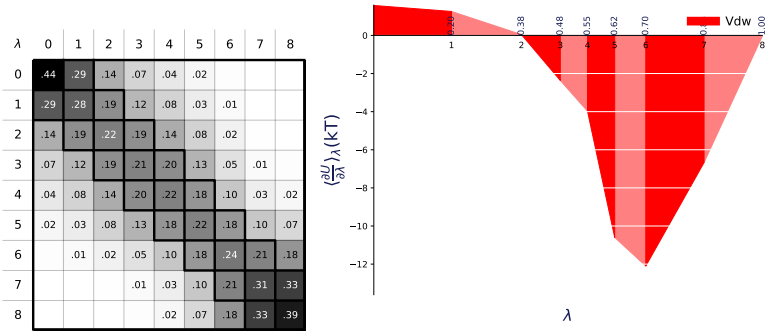
```
$ bash start_sim.sh
```

Each λ-point will be simulated sequentially after the other. However, like in the US simulations, the λ-points can be run in parallel on a computer cluster since they are independent.

4.3 Analysis of the Results

After finishing the simulation of each λ-point the first analysis for obtaining the hydration free enthalpy ΔG_{hyd} is performed in the current folder `6_Decoupling` via the bash script `run_BAR.sh`. It calls the built-in GROMACS implementation of BAR `gmx bar` and hence only needs GROMACS to be installed. The program needs the `dhdl` files located in `Lambda_XY/Production_MD/` as input. Running `gmx bar` can be done via:

[5] `https://manual.gromacs.org/documentation/2023.2/user-guide/mdp-options.html`



(a) Overlap matrix (b) Slope $\langle \partial U/\partial \lambda \rangle_\lambda$ versus λ

Figure 3: Generated graphical output from `alchemlyb`. Each λ-point was simulated for 10 ns at 300 K.

```
$ gmx bar -f Lambda*/Production_MD/dhdl.xvg -o -oi -oh
```

or by simply executing the prepared bash script:

```
$ bash run_BAR.sh
```

Further analysis is carried out with a Jupyter notebook that uses the `alchemlyb` library. The notebook expects the input files `dhdl.xvg` to be stored in a central directory. Therefore, before executing the notebook run the prepared bash script:

```
$ bash collect_dhdl.sh
```

For a simulation of 10 ns per λ-point `gmx bar` shall give you a value around 8.49 $\pm$ 0.02 kJ mol^{-1} for ΔG_{hyd} quite close to the experimental value of 8.4 kJ mol^{-1} at 298 K [10]. Note that this is the negative (!) of the direct output since a decoupling was performed in our case above (the transfer from the liquid into the gas phase rather than the other way around). `gmx bar` will produce three additional output files (`.xvg`): `bar.xvg` which contains relative free enthalpy differences for each λ-interval, `barint.xvg` containing the cumulative ΔG as a function of λ and `histogram.xvg` with the histograms converted from the time series of $\partial U/\partial \lambda$ taken from the `dhdl` files. Within the notebook for the MBAR analysis several graphical representations of the results are generated `O_MBAR.pdf` contains the overlap matrix (Figure 3 (a)), `dhdl_TI.pdf` shows the average of the slope $\partial U/\partial \lambda$ versus λ (Figure 3 (b)) and `dF_state.pdf` is a bar plot for the free energy differences between neighboring λ points calculated with different estimators (Figure 4). Finally, `dF_t.pdf` shows the convergence of the forward and reverse free energy change with the simulation time Figure 5.

Next, we repeat the decoupling procedure for the bound complex consisting of two methane molecules. The nonbonded interactions should be strong enough such that the two molecules stick together during the decoupling procedure.

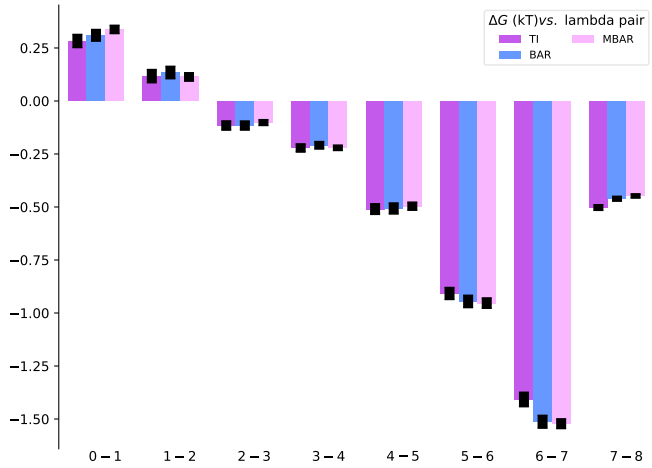


Figure 4: Generated graphical output from `alchemlyb`. Each λ-point was simulated for 10 ns at 300 K.

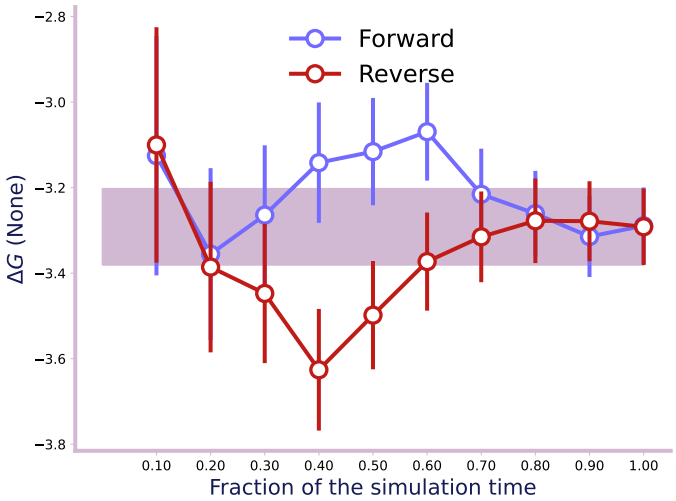


Figure 5: Generated graphical output from `alchemlyb`. Each λ-point was simulated for 10 ns at 300 K.

This can be checked by looking at the `COLVAR` files at each λ-point. The program may also crash if the complex dissociates during decoupling. If you notice that the two molecules dissociate an auxiliary bond can be defined that keeps the molecules together. Such a bond is already defined in the `topol.top` file but not activated by default. The contribution of this bond to the calculated free energy has to be accounted for as is discussed during the course.

Finally, a single methane molecule is decoupled from the complex. The files are prepared in the directory `day_4`. During the decoupling procedure we have to ensure that the decoupled molecule does not diffuse away. Therefore, a bond between the two molecules is established as the non-bonded interactions are turned off.

4.4 Cycle Closure [TO BE REVISED]

Attaching the distance restraint to the methane dimer in vacuum, $\Delta G^{\mathrm{D_2}}_{\mathrm{attach\,rest}}$, is calculated numerically by a free energy perturbation approach (FEP) derived from the Zwanzig-formula [11]:

$$\Delta G^{\mathrm{D_2}}_{\mathrm{attach\,rest}} = RT \ln\langle e^{V_r/RT}\rangle_{V_r} = -RT \ln\langle e^{-V_r/RT}\rangle \tag{1}$$

where $\langle ...\rangle$ and $\langle ...\rangle_{V_r}$ denote an ensemble average for which the restraining potential V_r is switched off and on respectively. From the above formula it follows that one basically has to calculate the exponential average of the time series of the bias potential which can be calculated from the sampled order parameter in the `COLVAR` file. Since we are regarding sticky methane, the free energy penalty of the applied distance restraint is quite low in the bound state and both identities of the FEP formula should yield almost the same (low) value. The contribution for releasing the distance restraint $\Delta G^{\mathrm{DM}}_{\mathrm{detach\,rest}}$ between one (fully interacting) methane and a dummy can be calculated via:

$$\Delta G^{\mathrm{DM}}_{\mathrm{detach\,rest}} = -RT \ln \frac{V^\circ}{V_{\mathrm{eff}}} \tag{2}$$

References

[1] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free energy calculations on biomolecules. i. the method. *J. Comput. Chem.*, 13:1011–1021, 1992.

[2] J. Kästner and W. Thiel. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "umbrella integration". *J. Chem. Phys.*, 123:144104, 2005.

[3] J. Kästner and W. Thiel. Analysis of the statistical error in umbrella sampling simulations by umbrella integration. *J. Chem. Phys.*, 124:234106, 2006.

[4] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129:124105, 2008.

[5] N. Schmid, A. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren. Definition and testing of the GROMOS force-field versions: 54A7 and 54B7. *Eur. Biophys. J.*, 40:843–856, 2011.

[6] M. Diem and C. Oostenbrink. The effect of using a twin-range cutoff scheme for nonbonded interactions: Implications for force-field parametrization? *J. Chem. Theory Comput.*, 16:5985–5990, 2020.

[7] Justin A. Lemkul. Introductory tutorials for simulating protein dynamics with GROMACS. *J. Phys. Chem. B*, 128:9418–9435, 2024.

[8] M. Shirts and A. L. Ferguson. Statistically optimal continuous free energy surfaces from biased simulations and multistate reweighting. *J. Chem. Theory Comput.*, 16:4107–4125, 2020.

[9] J. S. Hub, B. L. De Groot, and D. Van Der Spoel. g_wham-a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.*, 6:3713–3720, 2010.

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

[11] R. W. Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *J. Chem. Phys.*, 22:1420–1426, 1954.