

Free Energy Calculations in Water and Vacuum

Guido Putignano^a, Lorenzo Tarricone^a

^aDepartment of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

Abstract

Free-energy calculations encompass a set of methodologies designed to estimate the differences in free energy between distinct states within a molecular system. This study addresses a specific problem related to free energy and emphasizes its role in resolving particular conditions. Various approaches, including thermodynamic integration (TI) and other pertinent methods, can be employed to achieve this goal. The focal point of this endeavor involves the meticulous preparation of perturbation topologies, the establishment of a series of TI simulations, and a comprehensive analysis of these simulations to derive accurate estimates of free-energy differences. Through this research, we aspire to contribute meaningfully to the advancement of the field, fostering an insightful discussion on the subject matter.

Keywords: Free Energy, Water, Vacuum, Simulation, Biomolecular Systems.

1. Hamiltonian contributes to $\frac{\partial H}{\partial \lambda}$

Question A: In analyzing the GROMOS output files (*.omd files) to identify the terms contributing to $\frac{\partial H}{\partial \lambda}$, it's crucial to consider the components of the Hamiltonian. The Hamiltonian, in the context of molecular dynamics simulations, represents the total energy of the system. In the formula provided in class, there are typically five main components influencing the Hamiltonian's dependence on the scaling parameter λ .

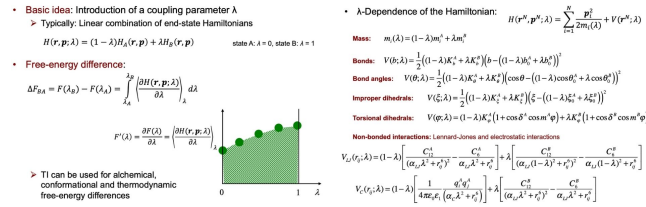


Figure 1: Hamiltonian and its dependencies

1. **Lambda (λ):** Lambda is a scaling parameter that varies between 0 and 1. It is a key factor in the Hamiltonian, governing the strength of certain interactions or constraints in the simulation. Changes in λ directly impact the potential energy terms associated with the system.
2. **Potential Energy Terms:** The potential energy component of the Hamiltonian is fundamental. It includes contributions from bonded and non-bonded interactions, and changes in λ will affect the potential energy landscape of the system.
3. **Pressure Terms:** The pressure component in the Hamiltonian is significant, especially in the context of periodic boundary conditions. The term related to pressure accounts for the simulation box dimensions and influences the system's behavior under different conditions.

4. **Kinetic Energy (Mass Terms):** Mass terms are crucial in the context of molecular dynamics simulations as they impact the kinetic energy of particles. Changes in mass or the treatment of different particle masses can affect the dynamics of the system.
5. **Other Energy Terms:** Depending on the specific details of the simulation and force field, there may be additional energy terms contributing to the Hamiltonian. These can include, for example, angle or dihedral potential terms.

It's essential to carefully examine how each of these components contributes to the overall Hamiltonian and, consequently, how the Hamiltonian varies with respect to the scaling parameter λ ($\frac{\partial H}{\partial \lambda}$). This analysis provides insights into the sensitivity of the system to changes in λ and helps in understanding the behavior of the simulated molecular system under different conditions.



Figure 2: Image describing hamiltonian dependencies in a *omd file

Question B: When considering all the plots generated, we can see the thermodynamic integration (TI)

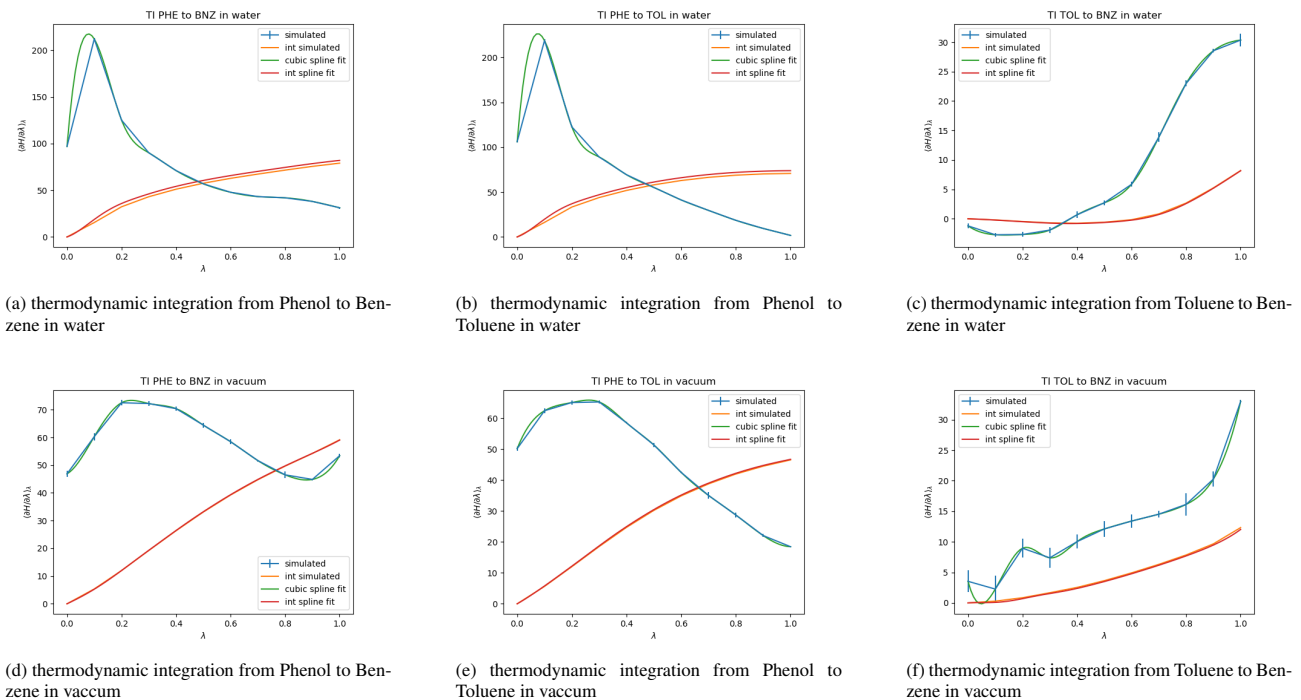


Figure 3: Thermodynamic integration under different conditions

Question C: After calculating the values (considering that it has been possible through the Colab notebook), we can revise the obtained values.

Entry	Integrated (Python)	Integrated after Cubic Spline Fitting (Python)
TI_PHE.BNZ_water	79.067693255 ± 0.403462141	82.0185464142
TI_PHE.BNZ_vac	59.156861685 ± 0.83185440425	59.08648731059075
TI_PHE.TOL_water	70.827498882 ± 0.5249082715	73.67484288423513
PHE.TOL_vac	46.48966307 ± 0.60559684515	46.66570997532655
TOL.BNZ_water	8.19493657 ± 0.5014732724	8.169056049890335
TOL.BNZ_vac	12.3333335565 ± 1.35484010715	12.061377325266035

Table 1: G values obtained via trapezoidal integration and cubic spline fitting (CSF) of the TI curves in kJ/mol

We can see the biggest difference within the case of the TI from Toluene to Benzene either in water or in vacuum. In this case, there are some differences related to the scale of the system that has much lower values. The reasons why it's present can be seen in the *.omd files. Here, the mass, energy potential, and pressure can influence the results. What is noticeable from the plots is that in the case of thermal integration of the other two examples, there was a condition where the estimated value used to increase and then decrease, with the TOL to BNZ being the opposite. There is a general trend of increase in the speed of growth at higher λ .

When considering the accuracy of the values, it's important to notice that even though the values of the TI with the TOL to BNZ system tend to be lower, the error always has the same dimension, or is even higher than in the other cases. For this reason, there are several approaches to solving this problem. We hereafter report the suggestions given in class.

Sources of Errors

- **Accuracy of calculated free-energy differences** can be affected by different sources of errors:
 - Force field inaccuracies (functional form, parameters)
 - Finite-size errors (system size, approximate electrostatics)
 - **Sampling errors** (finite simulation time, orthogonal barriers)
 - **Methodological issues**
 - Depend on the choice of pathway / reference state
 - Quadrature errors (in the case of thermodynamic integration)
 - Singularity upon atom creation/deletion (→ soft-core potentials)
 - Contribution of constraints

Figure 4: Ways of improving the accuracy of calculated free-energy differences

Question D:

For a closed system undergoing a cyclic process, the net change in any state function (like free energy) is zero. This is because a state function depends only on the initial and final states and is independent of the path taken between them. In the case of a circular pathway, the system returns to its initial state, and as a result, the overall change in free energy is zero.

Let's break down the circular pathway PHE → TOL → BNZ → PHE:

1. PHE → TOL: - A certain amount of free energy change (ΔG_1) occurs as the system goes from PHE to TOL.
2. TOL → BNZ: - Another free energy change (ΔG_2) occurs as the system transitions from TOL to BNZ.
3. BNZ → PHE: - A third free energy change (ΔG_3) occurs as the system completes the loop and returns from BNZ to PHE.

Now, if we consider the entire cyclic process (PHE → TOL → BNZ → PHE), the net change in free energy (ΔG_{net}) is the sum of these individual changes:

$$\Delta G_{\text{net}} = \Delta G_1 + \Delta G_2 + \Delta G_3$$

Since the system returns to its initial state, we can express this as:

$$\Delta G_{\text{net}} = 0$$

This result aligns with the fundamental concept that the net change in a state function for a closed loop is zero in thermodynamics, emphasizing the conservation of energy within the system.

In our case we have to divide the system in the two solutions (water and vacuum). For both cases, we can combine errors in quadrature. In a general formula, for three values A, B , and C with errors $\delta A, \delta B$, and δC , the error in their sum $S = A + B + C$ is given by:

$$\delta S = \sqrt{(\delta A)^2 + (\delta B)^2 + (\delta C)^2}$$

In the case of water:

$$\delta S = \sqrt{\delta A^2 + \delta B^2 + \delta C^2} = \sqrt{0.403462141^2 + 0.5249082715^2 + 0.5014732724^2} \approx 0.8305334644919409 \text{ kJ/mol}$$

For the cycle closure we have

$$\Delta G = 79.067693255 - 8.19493657 - 70.827498882 \approx 0.044257803 \text{ kJ/mol}$$

So the value is far below our error value.

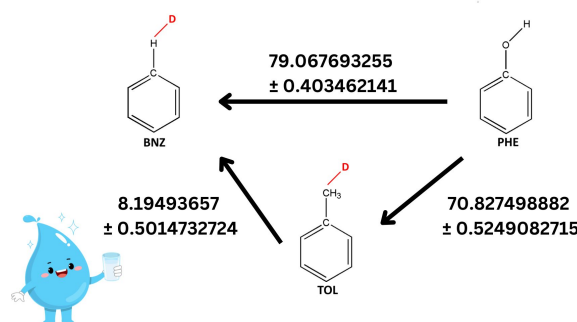


Figure 5: Cycle closure in water with unit measure kJ/mol

In the case of vacuum:

$$\delta S = \sqrt{\delta A^2 + \delta B^2 + \delta C^2} = \sqrt{0.83185440425^2 + 0.60559684515^2 + 1.35484010715^2} \approx 1.7012704090378987 \text{ kJ/mol}$$

$$\Delta G = 59.156861685 - 12.3333335565 - 46.48966307 \approx 0.333865 \text{ kJ/mol}$$

Also in this case, the value is below the error bar. The observed changes in free energy from the simulations conducted in water and vacuum are remarkably close to zero. Nonetheless, this doesn't automatically imply the achievement of thermodynamic circle closure. To begin with, there's an error margin of ± 1 to 2 kJ/mol in both scenarios, suggesting that we can't assert absolute closure in either situation. Moreover, there's a possibility that the individual free energy values were not precise and inadvertently offset each other, leading to an almost zero net free energy, which could be coincidental.

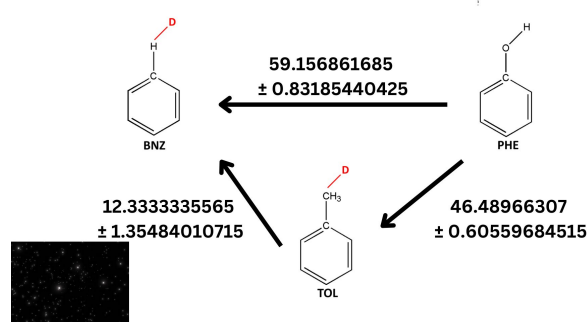


Figure 6: Cycle closure in vacuum with unit measure kJ/mol

Question E: In order to calculate the hydration free energy we can know that

$$\Delta\Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{BNZ} \rightarrow \text{PHE}} = \Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{PHE}} - \Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{BNZ}} = \Delta G_{\text{wat}}^{\text{BNZ} \rightarrow \text{PHE}} - \Delta G_{\text{vac}}^{\text{BNZ} \rightarrow \text{PHE}}$$

To calculate the hydration free energies of Phenol we have:

$$\begin{aligned} \Delta\Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{BNZ} \rightarrow \text{PHE}} &= \Delta G_{\text{wat}}^{\text{BNZ} \rightarrow \text{PHE}} - \Delta G_{\text{vac}}^{\text{BNZ} \rightarrow \text{PHE}} = \\ &= -(79.067693255 \pm 0.403462141 - 59.156861685 \pm 0.83185440425) = -19.91083157 \pm 0.9245341795 \text{ kJ mol}^{-1} \end{aligned}$$

For the case of Toluene

$$\begin{aligned} \Delta\Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{BNZ} \rightarrow \text{TOL}} &= \Delta G_{\text{wat}}^{\text{BNZ} \rightarrow \text{TOL}} - \Delta G_{\text{vac}}^{\text{BNZ} \rightarrow \text{TOL}} = \\ &= -(8.19493657 \pm 0.5014732724 - 12.3333335565 \pm 1.35484010715) = 4.138396987 \pm 1.43012776436 \text{ kJ mol}^{-1} \end{aligned}$$

In the case of the hydration-free energy of PHE relative to TOL,

$$\begin{aligned} \Delta\Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{TOL} \rightarrow \text{PHE}} &= \Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{PHE}} - \Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{TOL}} = \\ &= -(70.827498882 \pm 0.5249082715 - 46.48966307 \pm 0.60559684515) = -24.33783581 \pm 0.7978929134 \text{ kJ mol}^{-1} \end{aligned}$$

$$\begin{aligned} \Delta\Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{TOL} \rightarrow \text{PHE}} &= \Delta\Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{BNZ} \rightarrow \text{PHE}} - \Delta\Delta G_{\text{vac} \rightarrow \text{wat}}^{\text{BNZ} \rightarrow \text{TOL}} = \\ &= -19.91083157 \pm 0.9245341795 - 4.138396987 \pm 1.43012776436 = -24.04922856 \pm 1.702947114 \text{ kJ mol}^{-1} \end{aligned}$$

The value is compatible with what calculated above.

Question F: We always find a bias between the experimental values and the one obtained in the point above of around 4 kJ/mol . This shift from the given values in the table does make sense chemically because:

- The relative hydration free energy from phenol to benzene (according to the thermodynamic cycle) corresponds to $\Delta G_{\text{PHE} \rightarrow \text{BNZ}}^{\text{wat}} - \Delta G_{\text{PHE} \rightarrow \text{BNZ}}^{\text{vac}}$. We would expect the second factor of this difference to be slightly higher than the first one because benzene is expected to be a more stable compound than phenol, due to the fact that phenol is able to engage in many more reactions with its hydroxyl group
- analogous reasoning can be done for the hydration-free energy of toluene relative to benzene. Here the second term is much higher as benzene is a much more stable compound than toluene, due to the better distribution of charges along the aromatic ring.

In summary, even if the dummy atom doesn't have any non-bonded interaction. The arbitrary choice of its mass, bond length and charge will affect the simulation results.

2. Thinking questions

TQ1: In order to pass from the value of relative hydration-free energy (of toluene with respect to benzene) to the absolute one (of benzene) we will need to have the value of the absolute hydration-free energy of the toluene. This is due to the relation $\Delta\Delta G_{\text{TOL, BNZ}} = \Delta G_{\text{BNZ}}^{\text{vac} \rightarrow \text{wat}} - \Delta G_{\text{PHE}}^{\text{vac} \rightarrow \text{wat}}$. This will have in theory to be obtained experimentally (for example through concentration measurements made on two-phase systems). There is also the possibility of doing a simulation to obtain this value, but this could be very susceptible to the path parameterized by λ and would require advanced sampling techniques to overcome energy barriers

TQ2: I would adopt a different way of calculating the relative hydration-free energies in the first place, by using for example Enveloping Sampling Distribution instead of Thermodynamics Integration. This method is more suited to calculate the relative free energies for many different conditions. Then I will have to pass from a relative free energy value to a free energy one (see point above).

TQ3: Calculating reaction enthalpies using alchemical perturbations involves simulating the transformation of one molecule into another by gradually changing the potential energy function. However, when performing such calculations,

especially in different environments like vacuum vs. water, several factors can introduce inaccuracies. These can be due to the treatment of solvent effects, Polarization and charge transfer effects, Cavity formation and disruption, etc... Calculating the differences in reaction enthalpies in different environments can be more reliable because they help cancel out some of the inaccuracies introduced by the simulation methods.

TQ4: The final result in terms of hydration-free energy wouldn't change as dummy atoms are present in the same position in both the conformation of benzene and toluene. The problem arises when we need to calculate thermodynamic cycles. We have seen in class how to perform an alchemical transformation we need the number of particles to be identical for the two molecules. Here we have that the "arriving" physical states of $PHE \rightarrow TOL$ and $PHE \rightarrow BNZ$ have dummy atoms with some mass (influencing the kinetic energy of the Hamiltonian and therefore the lambda dynamics), while the alchemical change $TOL \rightarrow BNZ$ is done without dummy atoms. This inconsistency between the arrival point of one alchemical transformation and the initial point of the following transformation along the cycle might generate the non-perfect closure of the thermodynamic cycle. In theory though, being hydration-free energy a state function, no differences should be present in changing the "implementation" of these intermediate steps.

TQ5: In Section 1, we stated that 'the choice of intramolecular covalent interactions involving the dummy atom is arbitrary' and that it does not really matter 'because all these effects cancel out when we compare the free energies calculated in water to those calculated in vacuum.' However, this assumption may not hold universally. A counterexample can be formulated by considering a scenario where the dummy atom participates in a conformational change, and the choice of covalent interactions significantly affects the free energy, particularly when the dummy atom is involved in a flexible dihedral angle.

To formulate a more accurate condition for the allowed form of covalent interactions for dummy atoms, we should consider the following:

1. **Covalent Interactions Consistency:** Ensure that the choice of covalent interactions for dummy atoms is consistent across different environments (e.g., water and vacuum). Consistency helps avoid artifacts that might arise from changes in the dummy atom interactions.
2. **Path Independence:** The choice of covalent interactions should not introduce strong path dependence in the calculated free energies. Different paths taken during a simulation or calculation should yield consistent results.
3. **Energy Minima Consistency:** Confirm that the chosen covalent interactions do not introduce significant differences in the energy minima of relevant conformational states. If the dummy atom is involved in conformational changes, the energy landscape should be qualitatively similar with different choices of covalent interactions.

In practice, this means that we should avoid using dummy atoms to represent atoms that are involved in strong intramolecular covalent bonds.

Additionally, it's important to note that free energy components cannot be easily decomposed in the same way as energy components. The mathematical coherence in decomposing free energy is challenging due to the coupled nature of entropy

TQ6: Alternatively to the TI approach we could have used a number of other techniques, among which we have: λ -dynamics (or λ -LEUS), Enveloping Distribution Sampling (or Replica Exchange EDS) and Free-Energy Perturbation

Method	Pros	Cons
TI	Conceptually simple and widely used, easily parallelizable	Path dependent and many unphysical configurations are sampled
λ -LEUS	Created for sampling many configurations equally	Requires a good choice of the w parameter (width of the plateaus) and might require interpolation of the bias function
RE-EDS	Created for sampling many configurations equally and easily adaptable to different conformers	Introduces a bias of the Hamiltonian and setting the correct values for the s parameters is a non-trivial task
FEP	Uses a recursive update	The step size should be small enough

Table 2: Table with 4 Entries

3. Appendix

Summary 1

You can find the list for the Google Colab Jupyter here.