A novel method for calculating free energy difference between systems

Asaf Farhi[†]

Physics Department Weizmann Institute of Science, Rehovot 76100, Israel

School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Calculating free energy differences is a topic of substantial interest and has many applications including chemical reactions which are used in organic chemistry, biochemistry and medicines. In equilibrium free energy methods that are used in molecular simulations, one molecule is transformed into another to calculate the energy difference. However, when the compared molecules have different number of atoms, these methods cannot be directly applied since the corresponding transformation involves breaking covalent bonds which will cause a phase transition and impractical sampling. Thus, Quantum Mechanical Simulations, which are significantly more demanding computationally, are usually combined to calculate free energies of chemical reactions. Here we show that the free energies can be calculated by simple classical molecular simulations followed by analytic or numerical calculations. In this method each molecule is transformed into its replica with the VDW and Coulomb terms of the different atoms relaxed in order to eliminate the partition function difference arising from these terms. Then, since each transformed system can be treated as non interacting systems, the remaining difference in the (originally highly complex) partition function can be directly calculated. Since molecular force fields can often be automatically generated and the calculations suggested here are rather simple the method can form a basis for automated free energy computation of chemical reactions.

Introduction

Free energy calculations have a variety of applications which include binding, solvation, chemical reactions and more. In equilibrium methods one molecule is transformed into another to get the free energy difference. When the goal is to calculate the free energy difference of a chemical reaction, we can directly transform between the molecules. However, if the molecules have different number of atoms, a direct transformation will involve breaking bonds and as a result phase transition and impractical sampling. Thus, Quantum Mechanical calculations are usually combined with molecular simulations in free energy calculations of chemical reactions. One way to calculate the free energy difference of a chemical reaction in the general case is to calculate the solvation free energies of the molecules using molecular simulations. Then, the free energy difference between the molecules in the gas state is calculated with Quantum Mechanical methods. Thus, using the Thermodynamic Cycle the free energy difference between the molecules in the liquid state can be calculated.[1] Alternatively, QM/MM simulations in which the relevant part of the system has QM force fields can be performed. These simulations also generate information on the dynamics of the simulated system.[2] Here we suggest that the free energy difference can be calculated by classical molecular simulations followed by analytic or numerical calculations.

A separate simulation for each molecule

The idea in the method is to transform the reactants and the products (between which the free energy difference is calculated) into molecules that have the same partition functions up to factors that can be calculated. For now we will assume that by relaxing some of the energy terms, the free energies of the transformed system (molecule) can be decomposed to the free energy of the identical part between the systems (identical sub molecule) and to the free energy of the different part (different sub molecule) that can be calculated. Then, we will justify this assumption. First, we will match reactant molecules with similar product molecules. In some chemical reactions all the reactant molecules will be matched, and in some (usually when there are more reactants than products or vice versa) there will also be unmatched molecules.

For example, given the chemical reaction

$$A + B \rightarrow C + D \tag{1}$$

We can usually match molecule A, B to molecules C, D respectively.

[†] asaf.farhi@gmail.com

Then, we can transform the systems into the systems A', B', C' and D' by relaxing the potential terms of the atoms that are different between the systems and calculate the free energy differences $\Delta F_{A' \to A'}, \Delta F_{B' \to B'}, \Delta F_{C \to C'}$ and $\Delta F_{D \to D'}$. Then, if $F_{A'}$ and $F_{C'}$ can be decomposed into free energy of the common sub molecule between A and C and free energies that can be calculated (and similarly for $F_{B'}$ and $F_{D'}$), we will be able to get the free energy difference.

Another example is the following the chemical reaction:

$$A + B \to C \tag{2}$$

We can usually match molecule A to C and B will be left unmatched. Then we can transform the systems into the systems A', B' and C' by relaxing potential terms of the atoms that are not in the common sub molecule and calculate the free energy differences $\Delta F_{A' \to A'}, \Delta F_{B' \to B'}$ and $\Delta F_{C \to C'}$. Then, if the matched molecules can be decomposed into free energy of the common sub molecule and the different part (that can be calculated) and the free energy of the transformed unmatched molecules can be calculated, we will be able to calculate the free energy difference.

We now explain how the free energy difference between the original systems and their (alcehmmically) transformed replicas can be calculated using Thermodynamic Integration. We denote the Hamiltonian with the terms that are removed in the transformation by H_r and the Hamiltonian including the other terms by H_c. The λ dependent Hamiltonian can be written as follows:

$$H_{A/B}(\lambda) = \lambda H_{A_r} + H_{A_c} \tag{3}$$

Using the following derivation:

$$\Delta F_{A \to A'} = -k_B T \left[lnZ_A(\beta, \lambda = 1) - lnZ_A(\beta, \lambda = 0) \right] = -k_B T \int_0^1 \frac{d lnZ_A(\beta, \lambda)}{d\lambda} d\lambda = \int_0^1 \langle H_r \rangle \, d\lambda \tag{4}$$

it can be seen that simulations of the system at a set of λs (in the range $[0, 1]$) that interpolate between the original system and the transformed system in which the average energies $\langle H_r \rangle$ will be calculated, will allow us to numerically integrate and get the free energy difference between these systems.

The remaining free energy difference

Now we turn to explain how the free energy of the transformed matched molecule can be decomposed into two free energies (the free energy of the different submolecule will be calculated and the free energy of the common sub moleucle will cancel out) and how the free energy of the transformed unmatched molecule can be calculated.

Molecular modeling includes covalent bond, bond angle, dihedral angle, electric and VDW potentials.[3,4] Covalent bond, bond angle and dihedral angle potential terms are composed of the coordinates of two, three and four nearest covalently linked atoms respectively. Electric and VDW potentials relate between every atom pair in the system. For reasons that will be clear later the energy terms can be separated into *uncoupling terms* - covalent bond, bond angle, dihedral angle and *coupling terms* - electric and VDW. For the purpose of this method we will associate the improper dihedral angle terms with the *coupling terms*.

Now we can switch to relative coordinates (coordinates of atoms relative to other colvalently bounded atoms) and then to spherical coordinates.

$$d\Omega = \prod_{i=1}^{n} d\mathbf{r}'_i = d\mathbf{r}'_1 \prod_{i=2}^{k} d\mathbf{r}_i \prod_{i=k+1}^{n} d\mathbf{r}_i = \prod_{i=1}^{k} d\mathbf{r}'_i \prod_{i=k+1}^{n} r_i^2 \sin\theta_i dr d\theta d\phi \tag{5}$$

where

$$r_i = r'_i - r'_{i-1} \tag{6}$$

which will be chosen as the position of atoms relative to a covalently bounded atom. k represents the last atom that is common between the compared systems and $k + 1$ represents that first atom in the different sub molecule. In the case that the molecule is unmatched $k + 1$ will be the first atom.

The decoupled molecule/submolecule is first divided into elements of standard covalent bonds, bond junctions and of more complex structures that include molecular rings. Since each of the uncoupling terms depends on one independent variable, the integration in each element is independent of the others. Thus the integrals can be performed separately and then multiplied to yield the partition function and hence the free energy difference.

We write these free energy factors explicitly:

Covalent bond

The partition function of a covalent bond in Spherical Coordinates can be written as follows:

$$Z_c = \int e^{-\frac{\beta k_c (r-d)^2}{2}} r^2 dr = \frac{\pi^2 \left[\left(2d^2\beta k_c + 1\right)\left(erf\left(d\sqrt{\beta k_c}\right) + 1\right) + 2\sqrt{\pi}e^{-\beta k_c d^2} d\sqrt{\beta k_c}\right]}{l^3 \left(\beta k_c\right)^{3/2}} \tag{7}$$

where l is an arbitrary length (l^3 cancels out in comparisons).

Two Bonds Junctions

When considering the case of a Linear/Bent molecular shapes, it can be seen that in most cases when varying the bond angle, the rest of the molecule moves as a rigid body. Since the spherical coordinates representation includes three independent variables, varying a bonding angle is decoupled from all the other degrees of freedom of the molecule. Hence the calculation of free energy factors, which arise from bonding potential terms that are different between the compared molecules, is straightforward. One of the most common bonding angles potentials is the following:

$$V_b\left(\theta\right) = \frac{1}{2}k_\theta \left(\theta - \theta_0\right)^2 \tag{8}$$

So the integration over the corresponding degree of freedom can be written as:

$$Z_b = \int e^{-\beta V_b} d\Omega = \int e^{-\frac{\beta k_\theta}{2}\left(\theta - \theta_0\right)^2} \sin\theta d\theta = \tag{9}$$

$$\frac{1}{2\sqrt{\beta k_\theta}} e^{-i\theta_0 - \frac{1}{2\beta k_\theta}} \sqrt{\frac{\pi}{2}}\left(i\mathrm{Erf}\left[\frac{i - \theta_0\beta k_\theta + \beta k_\theta\pi}{\sqrt{2\beta k_\theta}}\right] + \mathrm{Erfi}\left[\frac{1 + i\theta_0\beta k_\theta}{\sqrt{2\beta k_\theta}}\right] - ie^{2i\theta_0}\left(\mathrm{Erf}\left[\frac{i + \theta_0 k_\theta}{\sqrt{2\beta k_\theta}}\right] - i\mathrm{Erfi}\left[\frac{1 + i\beta k_\theta(\pi - \theta_0)}{\sqrt{2\beta k_\theta}}\right]\right)\right) \tag{10}$$

This expression is real for positive and real values of k_θ, β and θ_0.

When varying a dihedral angle, the potential term value depends on the orientation of first bond (which determines the axis from which the dihedral angle is measured). However, since the integration has to be performed over all the range $[0, 2\pi]$, varying the ϕ angle will yield a factor which is independent of the location of the first bond. Thus, the integration does not depend on the direction of the first bond and is straightforward. The commonly used dihedral angles potential is of the following type:

$$V_d\left(\phi_{ijkl}\right) = k_\phi\left(1 + cos\left(n\phi - \phi_s\right)\right) \tag{11}$$

The integration over this degree of freedom yields the following result:

$$Z_d = \int e^{-\beta V_d} d\Omega = \int e^{-\beta k_\phi(1 + cos(n\phi - \phi_s))} d\phi = 2\pi e^{-\beta k_\phi} I_0\left(\beta k_\phi\right) \tag{12}$$

where $I_0\left(\beta k_\phi\right)$ is Bessel function of the first kind at βk_ϕ which is defined as follows:

$$I_0\left(x\right) = \sum_{l=0}^{\infty} \frac{(-1)^l}{2^{2l}(l!)^2} x^{2l} \tag{13}$$

Three or more Bonds Junctions

Molecule shapes can include monomer that splits into more than one monomer. Such examples are the trigonal planar, tetrahedral trigonal pyramidal etc. These cases will necessiate numerical integraion which can be performed using the Spherical law of cosines that can be written as follows:

$$cos\left(\theta_{12}\right) = cos\left(\theta_1\right)cos\left(\theta_2\right) + sin\left(\theta_1\right)sin\left(\theta_2\right)cos\left(\Delta\phi\right) \tag{14}$$

where θ_1, θ_2 denote the bond angles of two bonds with respect to the principal bond and $\theta_{12}, \Delta\phi$ denote the bond angle and the difference in ϕ angle between these two bonds respectively. Usually in these cases there is one dihedral angle energy term which is between one of the bonds, the principal bond and a previous bond. Since the integration over the other degrees of freedom yields a factor that is independent of the value of ϕ, the integrations can be treated as decoupled. Thus, the integration for the case of one monomer that splits into two can be written as follows:

$$Z = \int e^{-\beta(V_b + V_d)} d\Omega = Z_d Z_b \tag{15}$$

3

where

$$Z_d = 2\pi e^{-\beta k_\phi} I_0\left(\beta k_\phi\right)$$

and

$$Z_b = \int e^{-\frac{\beta}{2}\left[k_1^\theta\left(\theta_1-\theta_1^0\right)^2 + k_2^\theta\left(\theta_2-\theta_2^0\right)^2 + k_{12}^\theta\left(\theta_{12}-\theta_{12}^0\right)^2\right]} sin(\theta_1) sin(\theta_2) d\theta_1 d\theta_2 d\phi_2 \tag{16}$$

For the general case it can be written as follows:

$$Z_b = \int \prod_i e^{-\frac{\beta}{2}k_i^\theta\left(\theta_i-\theta_i^0\right)^2} \prod_{i>j} e^{-\frac{\beta}{2}k_{ij}^\theta\left(\theta_{ij}-\theta_{ij}^0\right)^2} \prod_i sin\theta_i d\theta_i \prod_{i\geq 2} d\phi_i \tag{17}$$

where θ_{ij} can be calculated from (14). In case there are any energy terms that introduce complexity they can be relaxed in the transformation.

In addition, the partition function of complex structures at $\lambda = 0$ can often be calculated numerically. Other internal bonding energy terms can also be included in these numerical integrations (and also not be multiplied by λ). In many cases the common sub systems include the complex structures, eliminating the need for these calculations.

These free energy factors can be substituted in:

$$\Delta F_{B'\rightarrow A'} = k_B T\left[\sum_i lnZ_{B_{c_i}} + \sum_i lnZ_{B_{b_i}} + \sum_i lnZ_{B_{d_i}} - \left(\sum_i lnZ_{A_{c_i}} + \sum_i lnZ_{A_{b_i}} + \sum_i lnZ_{A_{d_i}}\right)\right] \tag{18}$$

to give the remaining free energy difference (where A' and B' are transformed matched molecules).

Thus, we can write in terms of the partition functions:

$$Z \rightarrow Z_{\text{common int}} \prod_{i=1}^{l} Z_{c_i} \prod_{i=1}^{m} Z_{d_i} \prod_{i=1}^{p} Z_{2b_i} \prod_{i=1}^{q} Z_{3b_i} \prod_{i=1}^{r} Z_{\text{complex}_i}$$

where $Z_{\text{common int}}$ represents the partition function of the common part between the compared molecules that is interacting with the environment and Z_{c_i} and Z_{d_i} represent the ith covalent bond and dihedral angle partition function respectively. Z_{2b_i} and Z_{3b_i} represent the ith two bond and three or more bond junctions respectively and Z_{complex_i} represents the ith complex structure partition function. The arrow represents the transformation $\lambda = 1 \rightarrow 0$.

When the molecule is unmatched, we will relax all the coupling terms and calculate the free energy in a similar manner. This calculation can be written as follows:

$$Z \rightarrow Z_{\text{free particle}} \prod_{i=1}^{l} Z_{c_i} \prod_{i=1}^{m} Z_{d_i} \prod_{i=1}^{p} Z_{2b_i} \prod_{i=1}^{q} Z_{3b_i} \prod_{i=1}^{r} Z_{\text{complex}_i}$$

where $Z_{\text{free particle}} = V$ (see[5] for typical values of V).

Thus, comparison between any group of reactants and product can be performed by transforming each molecule in a separate simulation, followed by calculation of the free energies associated with the different sub molecules. Since the relaxed energy terms involve diverging terms at $r \rightarrow 0$, even at $\lambda \rightarrow 0$ these terms will still be dominant. Thus, in order for the calculations to be legitimate, soft core potentails have to be used (see for example[6]). In the case of rugged energy landscape, sampling techniques such as H-REMD have to be used in another λ dimension. In this free energy calculation method this sampling technique can be used in the same λ dimension.[6] The method has been demonstrated and compared with ThI for the calculation of free energy difference between two molecules of two atoms in a spherical potential (see Appendix for more details).

Acknowledgements

Appendix

Demonstration and comparison

In order to demonstrate the method it was used with all its ingredients in MC simulation to calculate the free energy difference between the systems *A* and *B* and this value was compared to the one calculated by numerical integration. Then, in order to asses the efficiency of the method, the free energy difference between the systems was calculated using ThI combined with H-PT (in MC simulation) and the running time of the two methods was compared. It is emphasized that the use of ThI is feasible only since we compare one molecule to another and since the molecules have the same number of atoms.

The compared systems are composed of a molecule of two atoms in which one atom is fixed to the origin and the second one is bound to the first by a covalent bond. The second atom in each system is in a θ dependent potential (in spherical coordinates), containing θ^{-12} term to represent the VDW repulsive term used in molecular modeling. The potential barrier was chosen to be of typical value of systems with tens of atoms, having rugged energy landscape. The covalent bond length difference was chosen to represent systems with few different atom lengths- see the next section for more details (the values of the pairs of spring constant and covalent bond length were taken from molecular simulation software). The following potential and parameters were used:

$$V_A = \frac{1}{2}k\left(r_{12} - r_{eq_A}\right)^2 - 5.5k_bT \cdot \left[sin\left(4\theta\right) - \frac{10^{-8}}{\left(\theta - \frac{3\pi}{8}\right)^{12}}\right] \tag{19}$$

$$V_B = \frac{1}{2}k\left(r_{12} - r_{eq_B}\right)^2 + 5k_bT \cdot \left[sin\left(4\theta\right) + \frac{10^{-8}}{\left(\theta - \frac{5\pi}{8}\right)^{12}}\right] \tag{20}$$

$$r_{eq_A} = 2.1\,\text{Å}, r_{eq_B} = 1.3\,\text{Å}, k_A = 123\,kCal/Mol\text{Å}^2$$

$$k_B = 428\,kCal/Mol\text{Å}^2, E_{cut\,off} = 7\,kCal/Mol$$

Here we used the partition function of two atoms with a covalent bond term represented by a harmonic potential that can be written as follows:

$$Z\left(\beta\right) = \frac{1}{l^3}\int_{-\infty}^{\infty}e^{-\beta k(r-d)^2}d^3x = \frac{4\pi}{l^3}\int_0^{\infty}e^{-\beta k(r-d)^2}r^2dr = \tag{21}$$

$$\frac{\pi\left[\left(2d^2\beta k + 1\right)\left(erf\left(d\sqrt{\beta k}\right) + 1\right) + 2\sqrt{\pi}e^{-\beta kd^2}d\sqrt{\beta k}\right]}{4l^3\left(\beta k\right)^{3/2}} \tag{22}$$

where l is an arbitrary length. This was used to calculate the free energy difference between the systems with the coupling terms relaxed.

The comparison of the method to the numerical integration yielded exactly the same results. The running time of the calculation of the free energy difference performed by the two methods was compared and yielded a factor of ~ 30 in favor of the novel method. This factor originates from the extra sampling dimension and the larger number of intermediates needed in ThI combined with H-PT. In Fig 1 the functions integrated in the two methods are plotted. It can be seen that the difference in
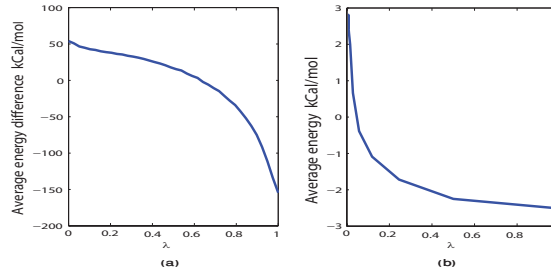

Figure 1: Integrated functions as a function of λ (a) Thermodynamic Integration (b) The novel method

magnitude of the integrated function in the novel method is much smaller than this in ThI (factor of ~ 40). In Fig 2 a scheme of the systems simulated in the two methods is presented (each point represents a simulation).

The dissimilarity between the systems that grows with the number of different particles, increases *both* the number of intermediates (due to a much larger difference in magnitude) and the number of simulation steps (increased variance) as compared with these in the novel method. The difference in the covalent bond description, that reduces the correlation between
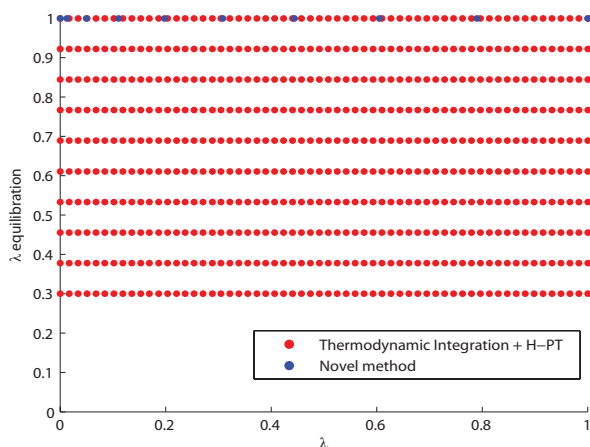
Figure 2: Systems simulated in both methods

the systems significantly (the penalties are also not bounded by the capping energy) and has the most dominant effect, has a completely negligible computational cost in the novel method. Thus, the efficiency is increased in 3 multiplicative dimensions. It is here to remind that while the method is highly efficient, its biggest advantage is that it enables comparisons of molecules with different number of atoms in the same environment and that it does not require to transform a molecule to another.

Correspondence between the toy model and realistic systems in the comparison to Thermodynamic Integration

In the case of realistic compared systems in which the molecules differ in the covalent bond length of one atom (usually when comparing two systems with different connectivity there are many such differences), when the coupling terms of the atoms that are different between the molecules are relaxed (disregarding the equilibration procedure both in the existing methods and in the novel method for simplicity), the comparison of the methods will yield results that are very similar to the toy model. This is since $\langle H_B - H_A \rangle (\lambda)$ in Thermodynamic Integration will be mainly affected by the changes between the systems. Thus, the functions integrated over in the toy model give good estimation to the ones in the comparison of the realistic systems mentioned above. Since the value of the functions integrated by Thermodynamic Integration is dominated by the covalent bond change (the rest of the difference in the energy terms is negligible as compared with it) it can be written:

$$\langle H_B - H_A \rangle (\lambda)|_{realistic} \cong \langle H_B - H_A \rangle (\lambda)|_{toy\,model} \tag{23}$$

Now we turn to show the correspondence in the novel method. We denote the energy terms of the atoms that are different between the realistic compared systems by H_{A_d/B_d}. Since these terms are the only ones in the integrated function it can be written:

$$\langle H_{A/B} \rangle (\lambda)|_{realistic} = \langle H_{A_d/B_d} \rangle (\lambda)|_{realistic} \sim \langle H_{A/B} \rangle (\lambda)|_{toy\,model} \tag{24}$$

In this case there will be similar values since the energy values of the non covalent energy terms are bounded by E_{cap}[6] and thus the average value is of typical value of $E_{capping}$. Here the covalent bond energy term is notf included in H_{A_d/B_d}. The parameters in the toy model for the covalent bond lengths and strengths are realistic and were taken from simulation software.

References

[1] William L Jorgensen, James M Briggs, and Jiali Gao. A priori calculations of pka's for organic compounds in water. the pka of ethane. *Journal of the American Chemical Society*, 109(22):6857–6858, 1987.

[2] Arieh Warshel and Michael Levitt. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology*, 103(2):227–249, 1976.

[3] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[4] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.

[5] Michael R Shirts. Best practices in free energy calculations for drug design. In *Computational Drug Discovery and Design*, pages 425–467. Springer, 2012.

[6] Asaf Farhi. A simple, robust and efficient method to calculate relative free energies. *Arxiv preprint cond-mat/1310.2112*, 2013.