

27th Annual CSP Workshops on “Recent Developments in Computer Simulation Studies in Condensed Matter Physics”, CSP 2014

Absolute Free Energies for Biomolecules in Implicit or Explicit Solvent

Joshua T. Berryman, Tanja Schilling

Université du Luxembourg, Faculté des Sciences, de la Technologie et de la Communication, 162a Avenue de la Faïencerie, L-1511 Luxembourg

Abstract

Methods for absolute free energy calculation by alchemical transformation of a quantitative model to an analytically tractable one are discussed. These absolute free energy methods are placed in the context of other methods, and an attempt is made to describe the best practice for such calculations given the current state of the art. Calculations of the equilibria between the four free energy basins of the dialanine molecule and the two right- and left-twisted basins of DNA are discussed as examples.

© 2014 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of The Organizing Committee of CSP 2014 conference

Keywords: Free Energy, Thermodynamic Integration, AMBER, EMIL, TI, Einstein Molecule

PACS: 05.10.Ln, 65.20.De, 82.60.Lf, 87.10.Rt, 87.10.Tf

1. Philosophies of Free Energy Calculation

The central object for the science of computational calorimetry is the partition function Z of the system under study:

$$Z = \int d\vec{r} d\vec{p} e^{-\mathcal{H}(\vec{r}, \vec{p})/k_B T} \quad (1)$$

If the partition function is known, then the free energy is also available:

$$A = -k_B T \ln(Z). \quad (2)$$

For the purpose of this discussion we will concentrate on the Helmholtz free energy, A , which is defined at fixed (N, V, T) .

E-mail address: josh.berryman@uni.lu

The integral to find Z is over the entire configurational and momentum space of the N -particle system, however the momentum part of the integral can be factored out of the expression:

$$\mathcal{H}(\vec{r}^N, \vec{p}^N) = V(\vec{r}^N) + K(\vec{p}^N) \quad (3)$$

$$Z = \int d\vec{r}^N \exp(-\beta V(\vec{r}^N)) \int d\vec{p}^N \exp(-\beta K(\vec{p}^N)) \quad (4)$$

$$Z = \int d\vec{r}^N \exp(-\beta V(\vec{r}^N)) \prod_i \int d\vec{p}_i \exp(-\beta \vec{p}_i^2 / m_i) \quad (5)$$

$$Z = Z_r Z_p \quad (6)$$

Because Z_p is readily accessible from the equipartition theorem, the calculation of a free energy is typically considered to have been solved if the configurational term of the partition function (Z_r) is obtained.

1.1. The Strategy of Fitting a Distribution to the Partition Function

An established and well-used method to find the free energy is to fit a multimodal Gaussian distribution to Z_r (Schlitter [1993]). A “normal” simulation of the system is run, using molecular dynamics (MD) or Monte Carlo (MC) to explore the equilibrium configurational space, with whatever techniques for accelerated sampling are available. When the simulation is judged to have converged, the mass-weighted covariance matrix of the atomic coordinates is diagonalised. The eigenvectors yielded by the diagonalisation are the normal modes of the system, and the eigenvalues describe the width of the distribution of displacements along each mode. If a Gaussian is fitted to the probability distribution function (PDF) for each mode, then this corresponds to the PDF of a harmonic spring and the free energy may be written down as that of a sum of independent harmonic oscillators.

This method of normal modes (sometimes called the Schlitter method) has the advantage that the spectrum of modes can give useful physical insight into the dynamics of the system, and that the analysis can be carried out on the results of a normal molecular dynamics simulation without any need to plan any computationally demanding separate calculation.

Extensions and refinements to this method are numerous, for example addressing the problem of studying systems which do not have harmonic dynamics, such as fluids (Henchman [2007]), however it is inherent to the class of methods in which some mathematical entity must be fit to Z_r (of which I will choose to view Schlitter as the exemplar) that the basic difficulties of fitting a very-high dimensional distribution are difficult to evade. If a simple function such as the Gaussian chosen by Schlitter is used, then the entropies for some modes will be incorrectly estimated; while more complex fitting functions such as the non-parametric histogramming employed in the paper (Numata et al. [2007]) require correspondingly more data to optimise the fit. Although approaches from this “Schlitter family” are certainly useful, their inherent difficulties and limitations suggest that other avenues should also be explored.

1.2. The Strategy of Making a Thermodynamic Path

Consider a piston of length $L + \lambda$, where λ can be adjusted between 0 and 1 by applying a force. The free energy difference ΔA between two different volumes of the piston is just the equilibrium work to compress or expand it:

$$\Delta A = \int_0^1 d\lambda, \left\langle \frac{\partial \mathcal{H}(\lambda, \vec{r}, \vec{p})}{\partial \lambda} \right\rangle_{NVT,\lambda} \quad (7)$$

Here, angle brackets indicate a time-average and the subscript λ indicates that each time-average is collected at a value of λ which has been held constant long enough for the system to fully equilibrate.

If we consider two systems with distinct Hamiltonians \mathcal{H}_0 and \mathcal{H}_1 , then λ can represent a path of alchemical transformation between them, such that the total Hamiltonian is defined as equal to \mathcal{H}_0 when $\lambda = 0$ and to \mathcal{H}_1 when $\lambda = 1$, but otherwise is some weighted average between them. The free energy difference between the systems represented by the two Hamiltonians can then be found in the same way as the work done to compress the piston in

equation 7. Here λ no longer represents a physical degree of freedom, but instead a path through a space having one extra degree of freedom compared to the physical systems \mathcal{H}_0 and \mathcal{H}_1 . The gradient $\frac{\partial \mathcal{H}(\lambda, \vec{r}, \vec{p})}{\partial \lambda}$ is thus referred to as a ‘generalised force’.

Instead of defining a continuous thermodynamic path, it may be preferred to compare points on the path in a pairwise way. In this case, a common approach is to find the Metropolis Monte Carlo probability for two states to exchange their value of λ :

$$e^{-\beta \Delta A} = \frac{\langle M(\beta[H(1, \vec{r}, \vec{p}) - \mathcal{H}(0, \vec{r}, \vec{p})]) \rangle_{\lambda=0, NVT}}{\langle M(\beta[H(0, \vec{r}, \vec{p}) - \mathcal{H}(1, \vec{r}, \vec{p})]) \rangle_{\lambda=1, NVT}} \quad (8)$$

where $M(v) = \min(e^{-v}, 1)$ is the Metropolis function (Bennett [1976]).

These two variants share a common essential philosophy, of sampling a path between two systems by means of a series of intermediate states, and I note the method of eqn. 8 here only as context for the remainder of the discussion which will focus on the Thermodynamic Integration (TI) approach of eqn. 7.

The strategy of thermodynamic paths has frequent difficulties in that the unphysical Hamiltonians which typically exist for intermediate values of λ can lead to numerical instabilities in the simulations, and also in that the convergence of the calculation proceeds only in proportion to the square of the length of simulation time which is employed (assuming random sampling of the configurational space). Despite these problems, TI calculations have become quite standard and are an available feature of most major MD simulation codes, such as AMBER (Case et al. [2012]) and Gromacs (Hess et al. [2008]), where they are commonly used for biochemical calculations such as free energies of ligand binding (e.g. in Genheden et al. [2011]), or to find solvation free energies (e.g. Wescott et al. [2002]).

Any discussion of path-based free energy calculations should acknowledge the ingenious non-equilibrium methods which can be used to find the equilibrium free energy (e.g. Goette and Grubmller [2009]), typically by exploiting Crooks’ theorem to relate the equilibrium work to the work done in moving dissipatively along the integration path (Crooks [1999]).

2. The Strategy of Making a Thermodynamic Path to a Tractable Model

The purpose of this paper is to discuss the strategy in which one endpoint of the mixing (by convention here: $\mathcal{H}_1 := \mathcal{H}(\lambda = 1, \vec{r}, \vec{p})$), represents a system for which the free energy is tractable and can be directly written down. This method generates absolute free energies rather than free energy differences, so is abbreviated here as ‘TIA’ (vs. ‘TIR’ for the relative method). Absolute free energies are not in themselves very useful, as it is the ΔA which determines chemical equilibria, however a set of absolute values can sometimes yield relative values more conveniently than a network or cycle of relative calculations.

The classical example of this strategy is the ‘Einstein Crystal’ method of Frenkel & Ladd (Frenkel and Ladd [1984]), in which the endpoint of the integration is a set of non-interacting harmonic oscillators, with the centre of each oscillation defined as the centre of the Wigner-Seitz cell of the original crystal. In order to treat liquids, a different choice of \mathcal{H}_1 has been advanced in which the particles are no longer bound to a single well but can exchange wells with each other in a Monte-Carlo move (Schilling and Schmid [2009]; Schmid and Schilling [2010]) giving more ‘liquid-like’ behaviour while retaining an analytically tractable partition function.

2.1. Why absolute free energies?

Calculations of absolute free energy at the time of writing are expensive and also can be difficult to parameterise, while finding the relative free energy of (for example) removing a ligand from a binding site on a protein is now a well-documented technique that can be carried out using automated workflows such as FEW (Homeyer and Gohlke [2013]). In the author’s opinion, the best occasion to use absolute free energy methods is when a significant conformational rearrangement takes place between the systems being compared. In this case it is easier to transmute a system into one which has a very different Hamiltonian (the tractable reference Hamiltonian \mathcal{H}_1) but a similar conformation than one which has a different conformation but a similar Hamiltonian.

3. Treatment of Water

When calculating absolute free energies of fluids, it may be valuable to allow a Monte-Carlo move in which the particles exchange the well to which they are attracted. This use of this move has the effect of adding to the partition function of the system a factorial term due to the indistinguishability of the particles, such that the total free energy A is larger than the sum of the free energies of separate particle-well systems. The definition and use of these wells for single-particle fluids is discussed elsewhere (Schilling and Schmid [2009]; Schmid and Schilling [2010]) however the best method to extend this approach to water, and to liquids of multi-particle components is not immediately obvious.

In the first calculation to use explicit water together with the swapping moves of Schmid and Schilling (Berryman and Schilling [2013]), all hydrogens were kept in place using the SHAKE algorithm, and of the water molecules only the oxygen atoms were therefore treated by the \mathcal{H}_1 Hamiltonian. This method is not general to molecules with more complex intramolecular degrees of freedom than water and also raises concerns in the case where the normal intermolecular forces on the hydrogens are mixed away, at values of λ approaching 1. The solution adopted for the current implementation (AMBER 14) is to create harmonic wells for the hydrogen atoms, and to position the wells in space relative to the oxygen atom (or to the first heavy atom of the chain, in the case of other solvents or melts). This approach has the odd effect of giving each water molecule a preferred orientation as it diffuses about in the system, but has the benefit of providing a numerically stable model with a well-defined free energy.

4. Correlated Sampling and Analysis of Errors

Consider two particles in separate 1-dimensional damped harmonic wells of spring constants k_1 and k_2 and friction constant γ , subject to the same sequence of pseudorandom forces, G :

$$\ddot{x}_1 = -\gamma \dot{x}_1 - k_1 x_1 + \gamma G \quad (9)$$

$$\ddot{x}_2 = -\gamma \dot{x}_2 - k_2 x_2 + \gamma G \quad (10)$$

If we are interested in the difference of the two displacements, X , then it turns out that we can write a new Langevin-like equation in this quantity:

$$X := x_1 - x_2 \quad (11)$$

$$\ddot{X} = -\gamma \dot{X} - k_1 X - (k_1 - k_2)x_2. \quad (12)$$

Viewed in this way, X is itself a damped harmonic oscillator, but with a random force that enters only indirectly through the last term $(k_1 - k_2)x_2$. This final term approaches zero as k_1 approaches k_2 , so for systems which are similar to each other we can expect that the dynamics will come to synchronise under the same sequence G , even if the initial conditions are markedly different.

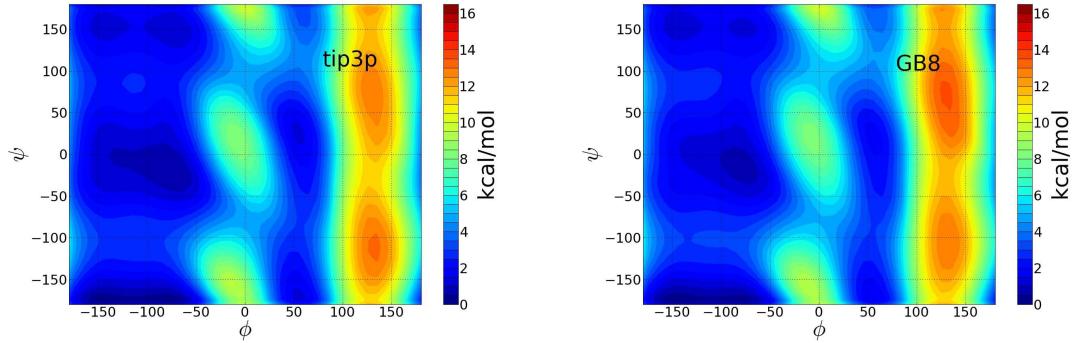
Although the strength of this effect in systems which are subject to chaotic dynamics or which are otherwise markedly anharmonic is not immediately apparent, strong benefits from this approach in calculating differences between similar systems even using other dynamical schemes than Langevin have recently been shown (Assaraf et al. [2011]).

4.1. Example Correlated Sampling Calculation Using Diananine

A test calculation comparing the four conformational wells ($C7_{eq}$, α_L , $C7_{ax}$ and α_R) of the capped diananine molecule was made. The solvent was treated using a Generalised Born model (Nguyen et al. [2013]) and the interatomic potentials were specified using the AMBER ff99SB forcefield (Hornak et al. [2006]). The four conformational wells were defined as rectangular regions using flat-bottomed restraints on the Φ and Ψ angles as defined in table 1. This is a very rough approximation to the real well shapes (fig. 1) which is however sufficient for the purposes of investigating convergence properties of the method. Different random number generator seeds were used for each value of λ (and

	C_{7eq}	α_R	C_{7ax}	α_L
$\Phi_{min}(^\circ)$	-210	-210	0	0
$\Phi_{max}(^\circ)$	20	0	120	120
$\Psi_{min}(^\circ)$	10	-120	-225	0
$\Psi_{max}(^\circ)$	260	30	0	120

Table 1. Dihedral angles boundaries used to define the four major free-energy basins of the dialanine molecule.

Fig. 1. Free energy landscapes of the dialanine molecule in explicit (TIP3P) water and implicit (GB8) dielectric continuum with respect to the dihedral angles Φ and Ψ , provided by Carlos Simmerling.

should be used for any restarts of the same trajectories), but the same set of seeds was used for each of the four separate well calculations.

Simulations were run at 9 equally-spaced values of λ on the range [0..1]. The generalised force values at each timestep were integrated using a simple Euler method to find the time series of ΔA . The averages of the ΔA values relative to the C_{7eq} well were collected after 50,000 timesteps (50ps) and the means and estimated standard errors were estimated assuming a decorrelation time $\tau = 1000$ steps. Table 2 shows that use of correlated sampling provides a modest but real reduction of the statistical error for this system. According to equation 12, the reduction available for other calculations should depend in some way on the degree of similarity between the systems being compared.

	C_{7eq}	α_R	α_L	C_{7ax}
Simmerling (implicit)	0.0	1.1	4.0	2.7
JTB, (implicit)	0.0	0.9 ($\pm .10$)	2.4 ($\pm .10$)	2.3 ($\pm .10$)
with new seeds:	0.0	0.9 ($\pm .13$)	2.4 ($\pm .14$)	2.3 ($\pm .13$)

Table 2. Free energy differences obtained from the Simmerling reference PMF (using the rough well definition of table 1) compared with values calculated using TI. The use of correlated sampling reduces the statistical error by approximately 25% for this example.

5. Screening and Softcoring

When the Lennard-Jones or Coulomb potentials are scaled by some small λ , their gradients in the region that is accessible at a given thermal energy $k_B T$ become gradually steeper, leading to severe numerical instabilities in MD calculations. It is therefore necessary to modify the interatomic potentials in some way, either by adding an additional purely repulsive term to the Hamiltonian ('screening') or by introducing a λ -dependence directly to the Lennard-Jones and Coulomb interactions ('softcoring').

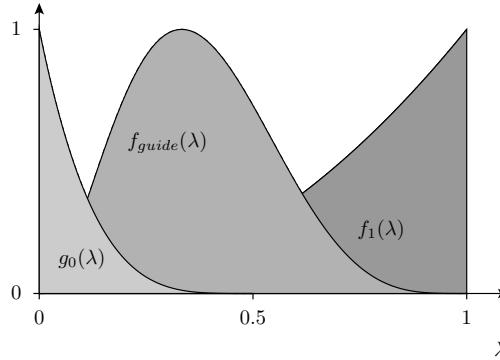


Fig. 2. Example mixing functions that were used in conjunction with a ‘screening potential’ approach to prevent numerical instabilities in treatment of the Lennard-Jones and Coulomb potentials (Berryman and Schilling [2013]).

An example screening approach that has been used (Berryman and Schilling [2013]) is to have a short-range repulsive potential of polynomial form:

$$V_{screen}(\vec{r}, \lambda) = f_{guide}(\lambda) \sum_{i,j>i} \text{MAX}[A(R_0 - r_{ij})^B, 0]. \quad (13)$$

In this approach, the screening potential is mixed out of the composite Hamiltonian at both endpoints of the calculation, using a polynomial mixing function $f_{guide}(\lambda)$ such as in figure 2, such that the total Hamiltonian is $\mathcal{H} = g_0(\lambda)\mathcal{H}_0 + V_{screen} + f_1(\lambda)\mathcal{H}_1$.

The introduction of a λ -dependence directly to the Lennard-Jones and Coulomb forces is apparently a little more elegant than the use of a simple screening potential. The method for doing this which is implemented in the popular AMBER simulation package (Case et al. [2012]; Steinbrecher et al. [2007]) is presented in eqns. 14 and 15.

$$V_{LJ}(\vec{r}, \lambda) = \sum_{i,j>i} (1 - \lambda) \left[\frac{A}{[\alpha\lambda + r_{ij}^6]^2} + \frac{B}{\alpha\lambda + r_{ij}^6} \right] \quad (14)$$

$$V_Q(\vec{r}, \lambda) = \sum_{i,j>i} (1 - \lambda) \frac{q_i q_j}{c_{ij} \sqrt{\beta\lambda + r_{ij}^2}} \quad (15)$$

where α and β are user-defined parameters. The functional forms of screened versus softcored potentials (for the summed Lennard-Jones and Coulomb interactions with unit parameters $A, B, c, \alpha, \beta, q$) are shown in figure 3.

An alternative means of softcoreing with a different functional form is implemented in the GROMACS simulation package (Hess et al. [2008]; Goette and Grubmller [2009]), however the essential difficulties and benefits of this approach are present in both cases.

It is not easy to accurately predict the phase behaviour of a system from the form of the interaction potential, so it is not surprising that the changes to the Hamiltonian made by the different softcoreing strategies can have unpredictable effects on the course of the thermodynamic integration calculation. If the change in the collective behaviour amounts to a first-order phase transition, then systematic error will arise due to trapping of metastable states at values of λ near to the transition value. Systematic error of this kind is particularly likely if a set of calculations is constructed to compare multiple free energy basins of the same Hamiltonian, for instance to find the free energy difference between liquid and solid phases. A significant hazard of this type of calculation is that the changes in the Hamiltonian with respect to λ can cause the state which is being studied to leave the metastable basin, so that the full path e.g. [solid \rightarrow tractable model] is not sampled, but instead part of the calculation jumps to e.g. the [liquid \rightarrow tractable model] path, causing hysteresis in the integration. If a chance of this type of transition exists then it is advisable if possible to add restraints to the

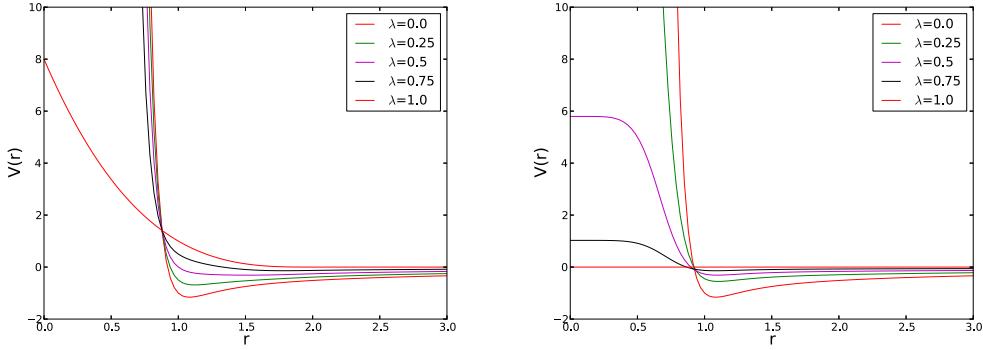


Fig. 3. Screened (left) and softcore (right) versions on the summed Lennard-Jones and Coulomb potentials.

\mathcal{H}_0 Hamiltonians which can prevent such a thing from taking place, without significantly altering the dynamics of the system as long as it remains in the chosen metastable basin.

5.1. Dialanine Screening and Softcore Example

The current implementation of the AMBER software (AMBER 14) supports both screening and softcore of interactions in TIA calculations (called EMIL in the AMBER software), however softcore is supported only in conjunction with explicit water. The dialanine free energy calculation of the previous section was therefore repeated using the TIP3P (Jorgensen et al. [1983]) water model. This resulted in a system of 3193 atoms rather than the previous 22, significantly increasing the computational expense, so no attempt was made to collect an accurate energy difference. Instead, the standard deviation of the calculated energy was found over runs of length only 200,000 steps. Standard deviations are shown in table 3. The presence of a larger number of atoms leads naturally to larger energies and larger fluctuations, however it is important to note that the screened potentials have larger fluctuations than the softcored ones, which suggests that the method of screening the Lennard-Jones core (while simple to implement) is less effective than softcore at providing an energetically stable simulation.

	C_{7eq}	α_R	α_L	C_{7ax}
$\sigma(\Delta A)$, GB screen	.	5.2	5.1	5.0
$\sigma(\Delta A)$, screen	.	99.	110.	97.
$\sigma(\Delta A)$, softcore	.	87.	88.	87.

Table 3. Standard deviations of the time series of estimated free energies, with implicit solvent, explicit solvent using screened potentials, and explicit solvent using softcored potentials.

6. Estimation of Convergence Times

The formula $ESE = \sigma \sqrt{\frac{\tau}{t-\tau}}$ gives the standard error of the mean assuming that a series of uncorrelated free energy estimates are taken at intervals of τ steps (where σ is the standard deviation of the free energy). Convergence properties of MD simulations of complex molecules can benefit from more sophisticated modelling, as applied in (Berryman and Schilling [2013]), however this simple equation is used in order to get an idea of the timescales which are needed to achieve a statistically accurate measurement of the free energy. Figure 4 shows the timescales needed to achieve accuracy of 0.1 kcal/mol in each of the dialanine calculations. The approximate timescale of 10^9 timesteps

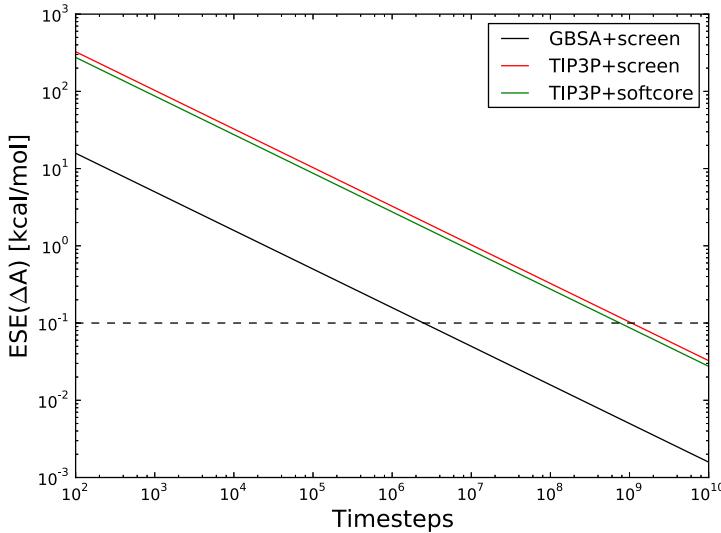


Fig. 4. Projected rates of convergence for screened and softcored calculations.

(1 μ s) for the 3193 atom systems is achievable by brute force on current hardware, but is not yet a casual calculation for the molecular dynamics amateur.

The B/Z conformational equilibrium calculation of DNA previously carried out (Berryman and Schilling [2013]) had 30,113 atoms which, assuming that the variance of the free energy measurement scales linearly with the system size, would lead to a requirement of the order 10^{10} timesteps. Fortunately, energies (and errors) for this system were quoted per base pair of the 12-bp DNA double helix (this one order of magnitude in error should save two orders of magnitude in computation time), and the acceptable error was of the order 1 kcal/mol rather than 0.1 kcal/mol (again, one order of magnitude in error should save two orders of magnitude in computation time), so the calculation was therefore of usable accuracy after simulations lasting less than 10 nanoseconds, rather than requiring runs in the 10μ s range.

Acknowledgements

This work was greatly assisted by discussions with Tanja Schilling, Charles Laughton and Sarah Anne Harris. Carlos Simmerling provided figures and reference data for the dialanine system. Joseph W. Kaus assisted with the implementation of softcore potentials in the AMBER software. Computing was provided by the HPC facility of the University of Luxembourg.

References

- Assaraf, R., Caffarel, M., Kollias, A.C., 2011. Chaotic versus nonchaotic stochastic dynamics in monte carlo simulations: A route for accurate energy differences in n -body systems. *Phys. Rev. Lett.* 106, 150601.
- Bennett, C.H., 1976. Efficient estimation of free energy differences from monte carlo data. *J. Computational Phys.* 22, 245–268.
- Berryman, J.T., Schilling, T., 2013. Free energies by thermodynamic integration relative to an exact solution, used to find the handedness-switching salt concentration for DNA. *J. Chem. Theory Comput.* 9, 679–686.
- Case, D., Darden, T., III, T.C., Simmerling, C., Wang, J., Duke, R., Luo, R., Walker, R., Zhang, W., Merz, K., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A., Kolossvary, I., Wong, K., Paesani, F., Vanicek, J., Wolf, R., Liu, J., Wu, X., Brozell, S., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.J., Cui, G., Roe, D., Mathews, D., Seetin, M., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., Kollman, P., 2012. AMBER12. University of California, San Francisco.

Crooks, G.E., 1999. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* 60, 2721–2726.

Frenkel, D., Ladd, A.J.C., 1984. New monte carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres. *J. Chem. Phys.* 81, 3188–3193.

Genheden, S., Nilsson, I., Ryde, U., 2011. Binding affinities of factor xa inhibitors estimated by thermodynamic integration and mm/gbsa. *J. Chem. Inform. Model.* 51, 947–958.

Goette, M., Grubmiller, H., 2009. Accuracy and convergence of free energy differences calculated from nonequilibrium switching processes. *Journal of Computational Chemistry* 30, 447–456.

Henchman, R.H., 2007. Free energy of liquid water from a computer simulation via cell theory. *J. Chem. Phys.* 126, 064504.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., 2008. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447.

Homeyer, N., Gohlke, H., 2013. Few: a workflow tool for free energy calculations of ligand binding. *J. Comput. Chem.* 34, 965–973.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C., 2006. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.

Nguyen, H., Roe, D.R., Simmerling, C., 2013. Improved generalized born solvent model parameters for protein simulations. *J. Chem. Theory Comput.* 9, 2020–2034.

Numata, J., Wan, M., Knapp, E.W., 2007. Conformational entropy of biomolecules: beyond the quasi-harmonic approximation. *Genome. Inform.* 18, 192–205.

Schilling, T., Schmid, F., 2009. Computing absolute free energies of disordered structures by molecular simulation. *J. Chem. Phys.* 131, 231102.

Schlitter, J., 1993. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 215, 617–621.

Schmid, F., Schilling, T., 2010. A method to compute absolute free energies or enthalpies of fluids. *Physics Procedia* 4, 131 – 143.

Steinbrecher, T., Mobley, D.L., Case, D.A., 2007. Nonlinear scaling schemes for lennard-jones interactions in free energy calculations. *J. Chem. Phys.* 127, 214108.

Wescott, J.T., Fisher, L.R., Hanna, S., 2002. Use of thermodynamic integration to calculate the hydration free energies of n-alkanes. *J. Chem. Phys.* 116, 2361–2369.