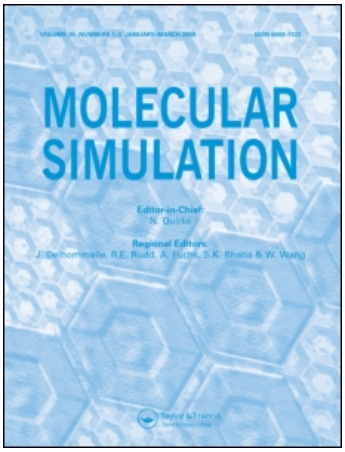

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Convergence Behavior in Free Energy Simulations

Mansoor A. S. Saqi[ab]; Julia M. Goodfellow[a]
[a] Department of Crystallography, Birkbeck College, London [b] Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, London

To cite this Article Saqi, Mansoor A. S. and Goodfellow, Julia M.(1991) 'Convergence Behavior in Free Energy Simulations', Molecular Simulation, 6: 1, 185 — 189
To link to this Article: DOI: 10.1080/08927029108022147
URL: http://dx.doi.org/10.1080/08927029108022147

Molecular Simulation, 1991, Vol. 6, pp. 185–189
Reprints available directly from the publisher
Photocopying permitted by license only

Note

CONVERGENCE BEHAVIOR IN FREE ENERGY SIMULATIONS

MANSOOR A.S. SAQI* and JULIA M. GOODFELLOW

Department of Crystallography, Birkbeck College, Malet St., London WC1E 7HX

(Received March 1990, accepted May 1990)

KEY WORDS: Free energy perturbation method, local minima, computer simulation

Free energy perturbation (FEP) simulations offer much promise for understanding site specific mutagenesis in protein engineering experiments and for rational drug design. These calculations can in principle yield the free energy difference between two states (for example a native and mutant protein) and some quite promising results have been obtained using the FEP technique. However much work remains to be done on accessing the accuracy and limitations of these calculations before they can be used on a routine basis [1]. Many of the problems occur because there is a complicated potential energy surface describing the two states in the calculation.

Previously we have studied the effect of the local environment on a given mutation, that of an OH group to a CH_3 group in the example of the mutation of the amino acid threonine to valine within peptide chains [2]. Here we present in detail the convergence behavior of a similar mutation namely that of methanol to ethane. Using this simple model, we can see how on occasion these FEP calculations can lead to incorrect results due to the presence of multiple minima in the energy surfaces for both states (reference and mutant). This leads to problems when the two states (reference and mutant) fall into unrelated minima. The problem of unrelated or non-equivalent minima is described in some detail in a paper by Haneef [3].

Statistical mechanics [4] gives the free energy difference between two states i (the reference state) and j (the mutant state) as

$$\Delta F = -kT \ln \langle \exp(-\Delta U/kT) \rangle_i$$

where $\Delta F = F_j = F_i$, $\Delta U = U_j - U_i$ and $\langle\ \rangle$ refers to averaging over the reference state, i. Generally this cannot be estimated in one step and the total free energy difference between a reference (state 0) and a mutant (state 1) system is obtained by summing the free energy changes over a number of hybrid intermediate states which are described with the aid of a coupling parameter (λ). Parameters, g, (eg non bonded energy terms and partial charges) are generated by

*Current address: Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX.

$$g(\lambda) = \lambda g_1 + (1 - \lambda)\, q_0.$$

In this study a constant volume (N, V, T) Monte Carlo FEP simulation is carried out for the interconversion of methanol to ethane i.e. $CH_3OH_{(aq)} \rightarrow C_2H_{6(aq)}$. These solutes are modelled using the OPLS force fields [5] and the solvent by the TIP4P model for water [6]. The simulations are carried out inside a cubic box of side length 18.63 Å with 214 waters and a 9 Å cutoff, using the Monte Carlo program BERNAL, developed at Birkbeck College, which employs double wide sampling.

The equilibration period is 0.5 million steps in all cases except for the window $0.125 < \lambda < 0.25$ and $0.25 < \lambda < 0.5$ where 0.8 million steps of equilibration are carried out. Results are presented for the last 0.5 millions steps for each window. The convergence behavior of the free energy is monitored in batches of 50,000 Monte Carlo steps. Thus means and standard deviations can be obtained from these batch averages for each window. From these simple statistics, it is seen that fluctuations in free energy and in solute ... solvent energy can be quite significant even for simple mutations especially if the number of thermodynamic windows used in the calculation are few.

The results for the forward (methanol ($\lambda = 0$) to ethane ($\lambda = 1$) and backward (ethane to methanol)) simulations are given in Table 1.

As there are clearly large discrepancies for the interval $0.25 < \lambda < 0.50$, this window was repeated using double wide sampling. (This is a protocol by which the free energy difference for both a forward and backward step can be calculated at the same time e.g. $0.25 < \lambda < 0.5$ and $0.25 > \lambda > 0.125$). These values are shown in Table 1 in parenthesis. The values from the two simulations for the two windows are consistent (i.e. -2.00 ± 0.33 kcal mol^{-1} and -2.35 ± 0.27 kcal mol^{-1}, 3.68 ± 0.61 and 4.27 ± 0.44 respectively). However, the results indicate that in both cases the average value for the forward run (3.96 kcal mol^{-1}) differs significantly from that of the reverse run (-2.05 kcal mol^{-1}).

The window was then subdivided into two smaller intervals by $0.25 < \lambda < 0.4$ and $0.4 < \lambda < 0.5$. Subdividing the window in this way gives the free energy for the

Table 1

STATE i	STATE j	FREE ENERGY kcal mol^{-1}
Backward		
0.125	0.000	-2.52 ± 0.45
0.250	0.125	-2.00 ± 0.33 (-2.36 ± 0.27
0.500	0.250	-2.05 ± 0.65
0.750	0.500	-0.30 ± 0.27
1.000	0.750	$+0.60 \pm 0.34$
Forward		
0.000	0.125	(2.52)
0.125	0.250	2.38 ± 0.46
0.250	0.500	3.68 ± 0.61 (4.27 ± 0.44)
0.500	0.750	1.75 ± 0.67
0.750	1.000	0.17 ± 0.30
Subdivision		
0.400	0.250	-1.21 ± 0.13
0.400	0.500	$+0.77 \pm 0.09$

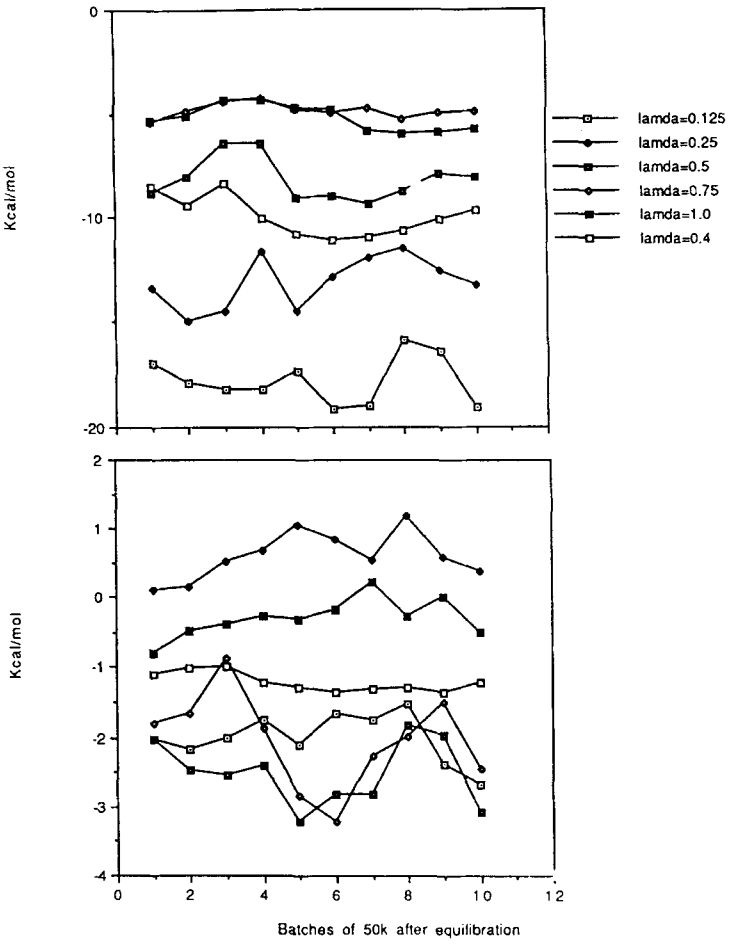

Figure 1 Solute solvent energies for various values of lambda for the methanol/ethane mutation (upper plot). Free Energy convergence profiles for steps in the backward run i.e. ethane → methanol (lower plot).

interval as 1.98 kcal mol^{-1} which now compares very favourably with that of the reverse window (-2.05 kcal mol^{-1}). The total free energy for the backward and forward simulations now becomes -6.27 kcal mol^{-1} and $+8.30$ kcal mol^{-1} (using the result 1.98 for the window 0.125 → 0.50, and assuming no hysteresis for the 0.000 to 0.125 window).

Solute-solvent energy convergence profiles and free energy convergence profiles are shown in Figure 1. The plots show the batch averages of these quantities for ten batches (0.5 million Monte Carlo steps) after equilibration. Quite large fluctations are observed for the solute ... solvent energies for those values of the coupling parameter (λ) describing the methanol-like states of the interconversion. This is due to the large partial charge on the oxygen of the methanol which gradually disappears in the course of the mutation as the molecule is changed into ethane. A much smoother behavior is seen for ethane-like hybrid states as the partial charges are now very small.

The convergence profiles for the free energy for intervals along the backward simulation show the fluctuations in the batch averages. The fluctuations are typically in the region of 0.5 kcal mol^{-1} with larger fluctuations for windows towards the methanol end of the mutation. Also shown in Figure 1 is the profile for the interval 0.40 to 0.25. This shows a much smoother behavior since it represents a smaller perturbation.

Our simulations are generally in agreement with a previous constant pressue (N, P, T) study of this system [7] but we find that even for rather small mutations the effect of the so-called multiple minima problem can be significant and care must be taken in carrying out free energy calculations. The observation that the free energies for the intervals 0.25–0.50 and 0.50–0.25 converge to very different answers highlights this problem. The fact that breaking down the forward interval, 0.25–0.50 into two smaller steps greatly improves the result suggests that the distribution of energies centered about 0.25 and about 0.50 are very asymmetric, with good overlap in one case not implying good overlap in the other. It also suggests that different step sizes might be needed for the forward backward runs.

Similar problems are encountered when the molecular dynamics slow growth algorithm is used to estimate free energy differences as there are no well-defined rules for the choice of λ during a simulation. Although ΔF is independent of the choice of path and hence λ, choice of path which involved high energy barriers will be inefficient and so one will generally choose λ so that the initial growth of an atom can proceed slowly [8]. Pearlman and Kollman [9] have recently introduced a more sophisticated approach involving dynamically modified windows for use with the slow growth molecular dynamics approach but this has only been tested on relatively simple systems such as neon to argon. Electrostatic decoupling, in which the coupling parameter λ is changed differently for the electrostatic terms compared with the other terms in the potential, may also help to reduce high energy barriers. However, the importance of using a windowing technique, compared with slow growth, is that we can monitor the precision of the free energy differences. ΔF_{ij}, at different stages in the simulation and we also obtain more realistic estimates for the error of the total forward or backward ΔF.

Throughout the simulation we have kept the solute fixed at the centre of the water box. However it is unlikely that this is what causes inadequate sampling of phase space. The convergence to different values suggests that unless the perturbation is very small, the mutant and reference systems do not fall into related minima and even very long runs are not guaranteed to overcome these barriers.

Acknowledgements

These studies have been carried out with support from the Science and Engineering Research Council from project grants GR/D/86683 and GR/D/47025. JMG wishes to thank the Wellcome Trust for a Research Leave Fellowship.

References

[1] Pearlman, D.A. and Kollman, P.A. (1989) "Free Energy Perturbation Calculations: Problems and Pitfalls along the Gilded Road" in 'Computer Simulation for Biomolecular Systems', (Eds Van Gunsteren WF and Weiner P), Escom Science Publishers, Leiden, pp 101–119.

[2] Saqi, M.A.S. and Goodfellow, J.M. (1990) "Free Energy Changes associated with Amino Acid Substitution in Proteins", *Prot. Engineering* 3, 419–423.

[3] Haneef, I. (1990) "Calculation of binding energies using a robust molecular mechanics technique: Application to an antibody-antigen complexes", *J. Mol. Graphics*, **8**, 45–51.
[4] Mezi, M. and Beveridge, D.L. (1986) "Free Energy Simulations", *Ann. NY Acad. Sci.*, **482**, 1–23.
[5] Jorgensen, W.L. and Tirado-Rives, J. (1988) "The OPLS potential functions for proteins", *J. Am. Chem. Soc.* **110**, 1657–1666.
[6] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1985) "Comparisons of simple potential functions for simulating liquid water", *J. Chem. Phys.*, **79**,926–935.
[7] Jorgensen, W.L. and Ravimohan, C. (1985) "Monte Carlo Simulation of differences in free energies of hydration", *J. Chem Phys.*, **83**, 3050–3054.
[8] van Gunsteren, W.F. (1989) "Methods for calculation of free energies and binding constants: Successes and problems" in 'Computer Simulation of Biomolecular Systems' (Eds van Gunsteren, WF and Weiner, PK), ESCOM, Leiden, pp. 27–59.
[9] Pearlman, D.A. and Kollman, P.A. (1989) *J. Chem. Phys.* **90**, 2460–2470.