Free Energy Difference Calculations by Thermodynamic Integration: Difficulties in Obtaining a Precise Value

Michael J. Mitchell* and J. Andrew McCammon

Department of Chemistry, University of Houston, Houston, Texas 77204-5641

Received 28 May 1990; accepted 28 August 1990

Free energy difference calculations have been performed by the "slow growth" method of thermodynamic integration of the AMBER 3.0 molecular dynamics program for the mutation of a conformationally restricted threonine dipeptide, *N*-acetyl threonyl-*N*-methylamide, to the corresponding alanyl dipeptide. By varying the total simulation length, it has been determined that precise free energy values are obtained only for simulations of greater than 100 ps total simulation time length. By varying the starting configurations for simulations of the same length, it has been determined that averaging the free energies obtained from shorter simulations may not give precise answers. Possible reasons for this behavior are discussed.

INTRODUCTION

Thermodynamic integration calculations of free energy differences are being applied increasingly to a variety of biological questions. Such calculations are being used to study small molecule-protein binding, antibody-antigen interactions, peptide solvation, the solution stabilities of proteins, conformational transitions in small molecules, and other issues related to the designing of drugs, pesticides, and materials.[1-4] Although such methods have a rigorous foundation in statistical mechanics, their use on large systems of molecules, particularly organic and biological molecules, awaited the availability of supercomputers and minisupercomputer workstations, and is still in its infancy.[1-5] The details for successfully calculating free energy differences in large systems by thermodynamic integration have not yet been carefully elucidated.

One computational difficulty encountered in many systems that contain organic or biological molecules is associated with the existence of rotational isomers. Peptides in solution, for instance, will have many rotational isomeric states separated by energy barriers of magnitude equal to or greater than kT, the product of Boltzmann's constant and absolute temperature. The existence of these rotational isomeric states has been shown to significantly complicate free energy difference calculations.[5] Thermodynamic integration calculations often sample only one well of a rotational isomer in moving between the starting and ending molecular states or often surmount a rotational transition only infrequently.[4] This can easily lead to significant errors in calculated free energy differences.[5] Fortunately, methods for dealing with the rotational isomer problem have been suggested,[6,7] and are being tested on a variety of computational problems. One proposed method for dealing with molecules which have multiple conformational states is to restrict the allowed conformations during the thermodynamic integration,[6] and to correct the resulting free energy values to allow for the various rotational isomers. Even with a conformationally restricted system, the limited extent of the sampling of phase space during a thermodynamic integration simulation may still result in sizable statistical errors in the resulting answers. Unfortunately, many thermodynamic integration calculations are being performed in a manner which would suggest that the answers obtained might be characterized as fortuitous.

The extensive use of explicit water models in molecular dynamics simulations of biological systems also requires substantial computational resources. Even placing a small spherical cap of water around one region of a protein can result in an increase in the total number of atoms by a factor of two or three. Since the number of nonbonded atomic interactions grows by approximately the square of the number of atoms, supercomputer power is needed for any but the smallest of biological systems. Hence, relatively little experience is available in thermodynamic integration simulations using explicitly hydrated systems. Here, too, statistical error in the sampling of phase space may be a problem during free energy difference calculations because the phase space with water is much greater than without.[4] Efforts at overcoming the statistical sampling problem have included averaging the free energies obtained for forward and backward thermodynamic integrations,

*Author to whom correspondence should be addressed.

Journal of Computational Chemistry, Vol. 12, No. 2, 271–275 (1991)

CCC 0192-8651/91/020271-05$04.00

and averaging the values obtained for more than one simulation.[8,9]

Systematic errors due to the length of time required for the water to relax as a result of perturbations of the solute atoms may appear in calculations of free energy differences.[10] Thermodynamic integration studies for the mutation of hydrated neon to sodium cation indicated that total simulation times of 40 ps to 80 ps might be sufficient to allow for both statistical sampling and water relaxation;[11] however, mutations involving smaller changes in electric fields may lead to slower responses. A recent study of the formation of a methane-like cavity in water indicates that very long simulation times may be necessary for this system.[10]

The purpose of this communication is to illustrate some of the computational difficulties one can encounter in obtaining precise free energies when performing thermodynamic integration calculations on biological systems in water. Results have been obtained which indicate that free energy difference calculations may have to be performed over much longer simulation times than appears to be the norm in published calculations. It appears that the calculated free energy difference may not converge to a distinct value unless thermodynamic integration has been performed for hundreds of picoseconds.

BACKGROUND

Typically, free energy difference calculations on biologically important systems are performed by either a perturbation method, sometimes termed "windowing," or a thermodynamic integration method.[1,2,12,13] The perturbation method utilizes the formally exact statistical mechanical relationship

$$\Delta A = -RT \ln \langle \exp(-\Delta H/RT) \rangle_0$$

where ΔA is the free energy difference, ΔH is the perturbation to the reference system Hamiltonian or potential energy function (e.g., corresponding to the replacement of a hydrogen atom by a methyl group), and $<>_0$ represents the time average of configurations of the reference molecular system.[1,2,12,14] The thermodynamic integration method assumes that the Hamiltonian can be defined as a function of a continuously varying paramater, λ. Then

$$\Delta A = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \approx \sum_1^N \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \Delta\lambda$$

where $\lambda = 0$ represents the starting state of the system and $\lambda = 1$ the final state (perhaps corresponding to a system containing a mutagenized protein), and $<>_\lambda$ represents an ensemble average at a particular λ value.[5,8,12] An additional, widely used approximation in the method eliminates the ensemble average at each λ value to yield the so-called slow growth formula[8,13]

$$\Delta A = \sum_1^N (H(\lambda_i) - H(\lambda_{i-1}))_{\lambda_{i-1}}$$

The accuracy of the above mentioned free energy calculation methods is strongly dependent upon an adequate sampling of the corresponding phase spaces. Statistical error estimation is more straightforward for the perturbation method than for the thermodynamic integration or "slow growth" methods.[12] In perturbation, the error associated with the ensemble average may be applied to the calculated free energy.[5,12] For thermodynamic integration methods, statistical errors are often estimated by noting convergence or hysteresis.[5] When two long simulations give the same answer, the simulation may have run long enough to sample the thermally accessible phase space and the calculation is said to have converged. Hysteresis is the sign corrected difference in free energy obtained for the same thermodynamic integration performed forwards and backwards. Some workers take half the hysteresis value as a measure of error in free energy change.[9] The absence of hysteresis or the appearance of convergence are necessary but not sufficient indicators of precision in calculated free energy. For example, short simulations of systems which include discrete water molecules as the solvent may give incorrect free energy values yet exhibit small hysteresis as the solvent may not have had time to relax in response to the changing Hamiltonian.

RESULTS AND DISCUSSION

Thermodynamic integration calculations have been performed to determine the Helmholtz free energy change for the mutation of a conformationally restricted threonine dipeptide, *N*-acetyl threonyl-*N*-methylamide, to the corresponding alanine dipeptide in a 6617.14 Å^3 box of 198 TIP3P waters. The backbone dihedral angles, ϕ and ψ were restrained by sinusoidal potentials centered at $\phi = -150°$ and $\psi = 150°$ with 394.9 kcal/mol barrier heights. The χ_1 dihedral angle was restrained with a sinusoidal potential centered at $\chi_1 = -60°$ with a 218.1 kcal/mol barrier height. All molecular dynamics equilibrations and thermodynamic integration simulations were performed under NVT conditions using 2 fs time steps and utilizing the Kollman "united atom" force field.[15] All bond lengths were constrained by the AMBER version of the SHAKE algorithm.[16] For thermodynamic integrations, the parameter λ, affecting the mutation of nonbonded repulsion, dispersion, and charge parameters, was varied linearly with time. The "slow growth" method of the GIBBS module of AMBER[16,17] as ported to the NEC SX2-400

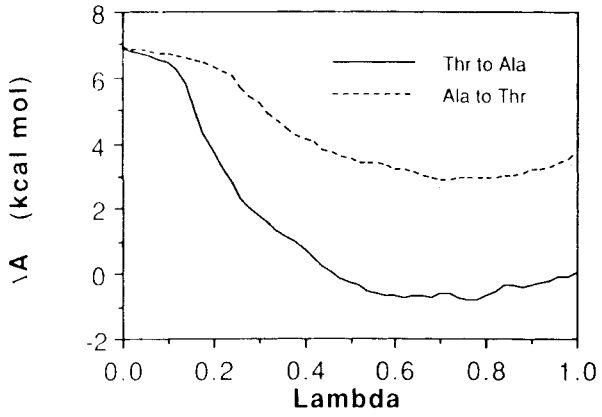

Figure 1. A plot of calculated free energy vs. λ for the AMBER "slow growth" method of thermodynamic integration of (solid line) threonine dipeptide (λ = 1) to alanine dipeptide (λ = 0), and for (dashed line) alanine dipeptide to threonine dipeptide in a periodic box of 198 TIP3P waters. The dashed line has been vertically offset by 6.88 kcal/mol so that the λ = 0 values coincide to illustrate the hysteresis.

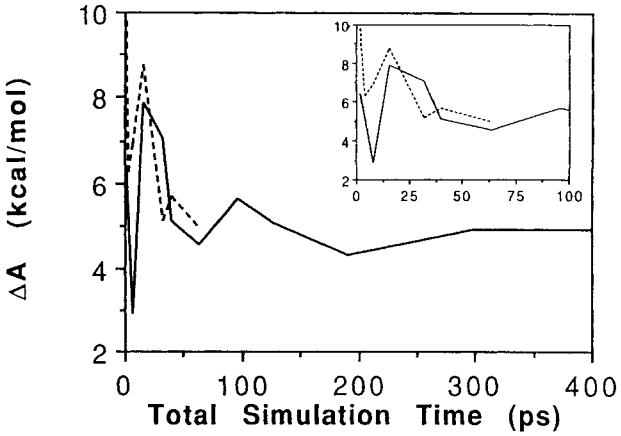

Figure 2. Calculated free energies for a series of threonine to alanine thermodynamic integrations (solid line). Conditions are the same as for Figure 1 except that the total simulation times are varied from 2 ps to 400 ps. All simulations started from the same configuration. The inset illustrates the same data for the shorter simulation times. A second series of simulations was performed (dashed line) for total simulation times varying from 2 ps to 64 ps. This second series of simulations differed from the first only in that all of them started from a configuration different from the one used for the first series.

supercomputer[18] was used in all cases. Figure 1 shows a hysteresis plot for the system where one 40 ps simulation was performed in each mutation direction. Each of the simulations was started from a thoroughly equilibrated configuration. As can be seen from Figure 1, the hysteresis is 3.2 kcal/mol or nearly half the total ΔA value observed for this simulation. Even taking half of this value as an estimate of error[9] still gives an unacceptably large error.

To determine the length of simulation necessary to obtain a precise free energy value, a series of different length free energy simulations was performed. Each simulation commenced from the same well-equilibrated configuration. The simulations varied in length from 2 ps to 400 ps. The calculated Helmholtz free energies (final ΔA values for thermodynamic integration of threonine to alanine) are shown in Figure 2 as a function of total simulation length (solid line). As can be seen, the answers vary widely for shorter simulation times. The calculated ΔA values for simulations longer than 100 ps still vary by more than 1 kcal/mol. The calculated free energies continue to vary, slightly, at longer times, though convergence appears to have been reached for the 300 ps and 400 ps simulations. In addition, a small hysteresis is obtained for the 300 ps simulation as shown in Figure 3 (ΔA = +4.93 kcal/mol for $\lambda = 1 \rightarrow 0$, and ΔA = −4.97 kcal/mol for $\lambda = 0 \rightarrow 1$). The starting molecular configuration used in the above series of simulations was further equilibrated by performing 20 ps of NVT molecular dynamics and a second series of free energy difference calculations was performed, all starting from this second molecular configuration, for total simulation times of 2 ps to 64 ps (dashed line, Figure 2). Even for simulations of 64 ps in length, the differences in the calculated free energies of the two series are on the order of 1 kcal/mol (inset, Figure 2). It is clear that free energy values obtained from short simulations can be strongly dependent upon the starting configuration. From these data it appears that at least a 300 ps thermodynamic integration simulation is necessary to obtain a precise free energy difference (i.e., to within about 0.1 kcal/mol by using the hysteresis as the error[9]) for a rotationally constrained peptide system in an explicit model, water bath. Shorter simulations suffer from random errors in phase space

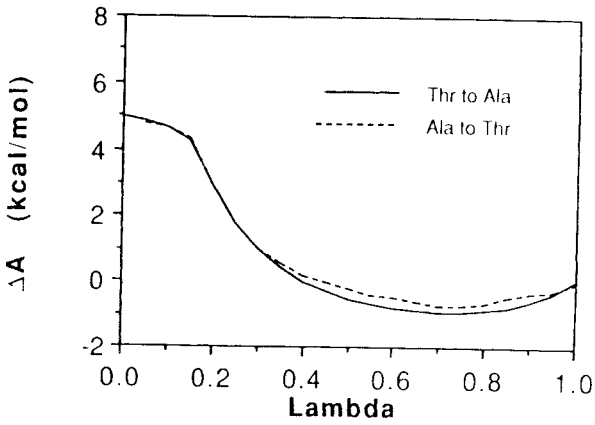

Figure 3. A plot of calculated free energies (ΔA) vs. λ for the AMBER "slow growth" method of (solid line) thermodynamic integration of threonine dipeptide (λ = 1) to alanine dipeptide (λ = 0), and for (dashed line) alanine dipeptide to threonine dipeptide. All conditions were the same as for Figure 1 except that the total simulation time in each direction was 300 ps. The dashed line is vertically offset by +4.93 kcal/mol to illustrate the hysteresis.

sampling. Longer simulations may reveal that convergence has not been reached even at a total simulation time of 400 ps.

To assess the validity of averaging the answers of short thermodynamic integrations to obtain a precise free energy value, a series of 40 ps thermodynamic integration simulations was performed for both threonine to alanine and alanine to threonine mutations. The simulations were started from configurations that differed by 10 ps of molecular dynamics equilibration. It is implicitly assumed that the average of the results for the simulations in one direction will constitute an approximately correct answer, i.e., that phase space sampling errors will be averaged by this process. Figures 4 and 5 show plots of ΔA vs. λ for the several simulations (dashed lines) as well as the average values of ΔA vs. λ for the simulations in each direction (solid line). When the averages are compared, there is still a substantial hysteresis between simulations ($\Delta A = +6.2$ kcal/mol for $\lambda = 1 \rightarrow 0$, and $\Delta A = -3.8$ kcal/mol for $1 = 0 \rightarrow 1$). This large hysteresis may represent the systematic error due to incomplete relaxation of the explicit water surrounding the solvated dipeptide.[10] In any case, these data strongly suggest that for a solvated peptide or protein, a 40 ps thermodynamic integration will not generally be long enough to obtain a precise value. It also appears that one long free energy simulation is superior to the average of several shorter free energy simulations.

The wide range of ΔA values obtained for the simulations (Figs. 4 and 5) calls into question the unusually good precision attributed to reported free energy difference values that are derived from the average of only one or two relatively short thermodynamic integration simulations in each direction.[8,9] By averaging the results of only a single simulation in each direction, the data of Figures 4 and 5 could have, by chance, yielded answers varying from +3.88 kcal/mol to +6.39 kcal/mol for the threonine dipeptide to alanine dipeptide mutation. It appears that, for the AMBER method of thermodynamic integration at least, an error of no less than 1.5 to 2.0 kcal/mol should be reported for similar free energy difference calculations performed in a similar manner.

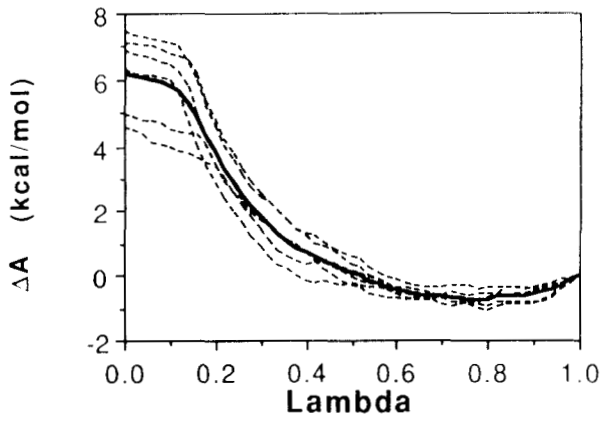

Figure 4. A plot of calculated free energies (ΔA) vs. λ for the AMBER "slow growth" method of thermodynamic integration of threonine dipeptide ($\lambda = 1$) to alanine dipeptide ($\lambda = 0$). Conditions were the same as for Figure 1 except that each simulation (dashed line) was started from a different configuratin. The numerical average (solid line) of the simulations is also shown.

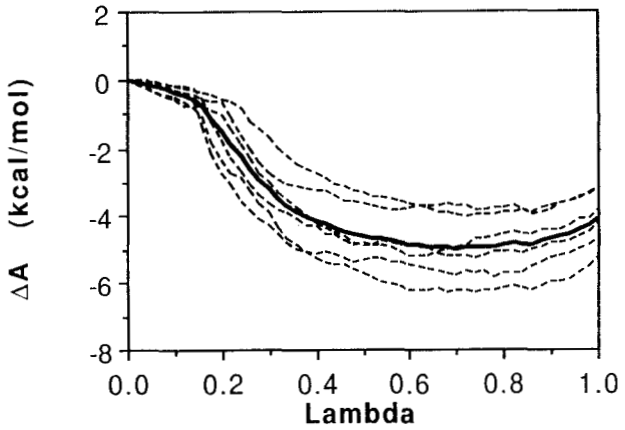

Figure 5. Same as Figure 4 except that alanine dipeptide ($\lambda = 0$) is converted to threonine dipeptide ($\lambda = 1$) during the simulations.

Great care should be taken when performing thermodynamic integration, free energy difference calculations on biological systems such as proteins or even small peptides in solution. Even for the relatively constrained system used for these studies (the only unconstrained dihedral angle is χ_2, which corresponds to rotations around the C—OH bond in the threonine side chain), and for reasonably long simulations (up to about 100 ps in length), one obtains answers that can vary by as much or more than 1 kcal/mol. The free energies obtained in such simulations are often additively combined, through the concept of closed thermodynamic cycles,[1,3,5] to obtain calculated free energies of the relative binding of substrates by enzymes or antigens by antibodies, etc. whose total free energy differences are often less than 2 to 5 kcal/mol. Standard error propagation algorithms applied to these calculations and using the hysteresis of short simulations as a measure of σ, will yield errors that are greater than the free energies sought unless the error in each free energy simulation is of the order of 1 kcal/mol. Using these criteria, it appears that many reported literature values for free energy simulation results may not be known with any reasonable degree of precision. In general, one should carefully analyze the convergence and hysteresis of the calculated results to ascertain when long enough thermodynamic integration simulations have been performed. One may resort to nonlinear variation of the mutation parameter λ to gain effective simulation length.[12,19]

In the particular case presented here, one might ideally use most of the simulation time for sampling in the interval $0.1 < \lambda < 0.4$, where the free energy varies most rapidly.

We are grateful to Peter Kollman for supplying the original AMBER programs and to Bill Swope of IBM and Jim Caldwell of U. C., San Francisco for supplying the IBM versions. We are also grateful to T.P. Straatsma for helpful discussions. This work has been supported in part by NSF, the Robert A. Welch Foundation, the Texas Advanced Research Program, the Houston Advanced Research Center, and HNSX Supercomputers. J.A.M. is the recipient of the George H. Hitchings Award from the Burroughs Wellcome Fund.

References

1. J.A. McCammon and S.C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, England, 1987.
2. M.P. Allen and D.J. Tildesly, *Computer Simulations of Liquids*, Clarendon Press-Oxford, Oxford, England, 1987.
3. C.L. Brooks, III, M. Karplus, and B.M. Pettitt, *Proteins, A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Wiley, New York, 1988.
4. P.F.W. Stouten and B.P. van Eijck, *Molecular Simulation*, **4,** 193 (1989).
5. D.L. Beveridge and F.M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.*, **18,** 431 (1989) and references therein.
6. T.P. Straatsma and J.A. McCammon, *J. Chem. Phys.*, **90,** 3300.
7. T.P. Straatsma and J.A. McCammon, *J. Chem. Phys.*, **91,** 3631 (1989).
8. U.C. Singh, F.K. Brown, P.A. Bash, and P.A. Kollman, *J. Am. Chem. Soc.*, **109,** 1607 (1987).
9. L.X. Dang, K.M. Mertz, and P.A. Kollman, *J. Am. Chem. Soc.*, **111,** 8505 (1989).
10. M. Mazor and B. Pettitt, *Molecular Simulation*, in press.
11. T.P. Straatsma and H.J.C. Berendsen, *J. Chem. Phys.*, **89,** 5876 (1988).
12. T.P. Straatsma, Ph.D. Thesis, University of Groningen, Groningen, The Netherlands, 1987.
13. P.A. Bash, U.C. Singh, R. Langridge, and Kollman, P.A. *Science* **236,** 564 (1987).
14. R.W. Zwanzig, *J. Chem. Phys.*, **22,** 1420 (1954).
15. S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C., Ghio, G. Alagona, S.P. Profeta, Jr. and P.K. Weiner, *J. Am. Chem. Soc.*, **106,** 765 (1984).
16. P.K. Weiner, and P.A. Kollman, *J. Comp. Chem.*, **2,** 287 (1981).
17. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *J. Comp. Chem.*, **7,** 230 (1986).
18. M.J. Mitchell and J.A. McCammon, *Computers & Chemistry*, in press.
19. A.J. Cross, *Ann. NY Acad Sci.*, **482,** 89 (1986).