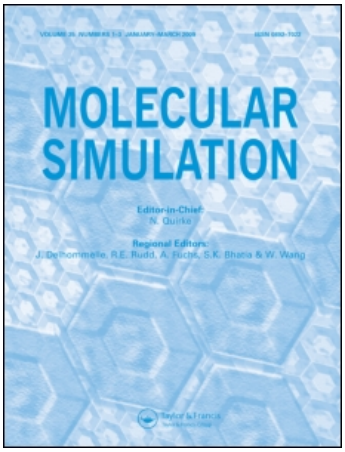

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Free Energy Calculations. The Long and Winding Gilded Road

Christophe Chipot[a]; David A. Pearlman[a]
[a] Equipe de chimie théorique, Institut nancéien de chimie moléculaire, UMR CNRS/UHP 7565, Université Henri Poincaré, Vandœuvre-lès-Nancy cedex, France

Online publication date: 26 October 2010

To cite this Article Chipot, Christophe and Pearlman, David A.(2002) 'Free Energy Calculations. The Long and Winding Gilded Road', Molecular Simulation, 28: 1, 1 — 12
To link to this Article: DOI: 10.1080/08927020211974
URL: http://dx.doi.org/10.1080/08927020211974

Molecular Simulation, 2002, Vol. 28(1–2), pp. 1 – 12

FREE ENERGY CALCULATIONS. THE LONG AND WINDING GILDED ROAD

CHRISTOPHE CHIPOT[a] and DAVID A. PEARLMAN[b,*]

[a]*Equipe de chimie théorique, Institut nancéien de chimie moléculaire, UMR CNRS/UHP 7565, Université Henri Poincaré, BP 239, 54506 Vandœuvre-lès-Nancy cedex, France;*
[b]*Vertex Pharmaceuticals Incorporated, 130 Waverly Street, Cambridge, Massachusetts 02139, USA*

(Received December 2000; accepted January 2001)

The fourth workshop on free energy calculations held at the Centre Européen de Calcul Atomique et Moléculaire (CECAM) in Lyon, last June 19 – 21, 2000, presented an opportunity for a critical look at the past successes and failures of simulations targeted at the estimation of free energies, and for a glimpse into their promising future. In an abridged historical background, the methodological milestones paving the road for today's free energy simulations are summarized. The continuing difficulties encountered when attempting to obtain accurate free energy estimates are then discussed, with an emphasis on the potential usefulness of such large-scale calculations in non-academic environments. Finally, current strategies for improving the reliability of free energy calculations, while at the same time making them more affordable (and thus more compatible with the constraints of the industrial environment) are discussed.

Keywords: Free energy calculations; Statistical simulations; Errors, of calculated free energies; Convergence, of calculated free energies

1. HISTORICAL BACKGROUND

An understanding of nearly any chemical process requires, at its core, an understanding of the underlying free energy behavior. For instance, in the field of rational, *de novo* drug design, such crucially important properties as protein-ligand association constants and membrane-water partition coefficients cannot be reliably and accurately predicted without the knowledge of

*Corresponding author.

ISSN: 0892-7022 DOI: 10.1080/08927020290004340

the associated free energy changes. The ability to predict *a priori* the physical constants associated with these processes is now a part of the modeler's toolkit – a result of relentless developments on both the software and the hardware fronts that can be traced back over the past fifty years.

In many respects, the foundations of the field of free energy calculations were laid in the seminal work of John Kirkwood [1], Robert Zwanzig [2], and Charles Bennett [3]. In 1954, Zwanzig described a protocol based on simple, classical statistical mechanics for calculating the free energy difference, ΔA, between any two states a and b of a system, in what will be referred to as free energy perturbation (FEP) method:

$$\Delta A_{a\to b} = -k_B T \ln\langle e^{-[\mathscr{H}_b(\mathbf{r}^N,\mathbf{p}^N)-\mathscr{H}_a(\mathbf{r}^N,\mathbf{p}^N)]/k_BT}\rangle_a \tag{1}$$

where k_B is the Boltzmann constant, T, the temperature, $\mathscr{H}_a(\mathbf{r}^N,\mathbf{p}^N)$, the Hamiltonian of the N-particle system in state a, and $\langle\cdots\rangle_a$, the ensemble average over configurations representative of that state. In most cases, the free energy difference between states a and b is so large that it precludes the direct use of perturbation formula (1). In its practical implementation, FEP decomposes the path that leads from a to b into a series of non-physical states connected to each other by means of a so-called "coupling parameter", λ [4]. Formally, the FEP Eq. (1) can be restated in terms of a continuous integral, leading to the well-known thermodynamic integration (TI) approach:

$$\Delta A_{a\to b} = \int_0^1 \left\langle \frac{\partial\mathscr{H}(\mathbf{r}^N,\mathbf{p}^N;\lambda)}{\partial\lambda}\right\rangle_\lambda \mathrm{d}\lambda \tag{2}$$

where λ is the coupling parameter that links state a, *i.e.*, $\lambda=0$, to state b, *i.e.*, $\lambda=1$. Unfortunately, implementation of the proposed equations for tackling sizeable systems of interest would have to wait for another thirty years from Zwanzig's initial work, to the dawn of the modern supercomputer era and the generalized availability of large computational resources.

In 1982, Postma *et al.* [5] and Arieh Warshel [6] published the papers that really ignited the modern field of free energy calculations. Using what were then considered to be fast computers, and implementing the ideas put forward by Zwanzig [2], they were able to predict solvation free energies for small systems. The remarkable agreement between theory and experiment in these pioneering articles led to an explosion of activity in the field, which was further fueled by papers presenting applications to protein-ligand

systems with similarly good results – see for instance Ref. [7]. Subsequently, free energy calculations have been applied to a plethora of chemical questions ranging from the simple estimation of the free energies of hydration [8–14] and adsorption [15] for small, organic molecules to the more complex determination of relative and absolute protein-ligand binding free energies [7, 16–21].

In retrospect, with nearly two decades of hindsight, it is clear that in some cases the amazingly good agreement between theory and experiment in these early studies must have been fortuitous. More recent investigations have demonstrated that simulations an order of magnitude or more longer than those used in the original studies are often required to obtain reliable and reproducible results [22]. As applications of free energy methods exploded within the scientific community, so did the realization that applying these novel approaches in a predictive and reliable fashion was not as straightforward as hitherto expected. The ensuing nearly two decades were, thus, spent characterizing and improving free energy methodologies to bring these calculations to the point where – in the very best cases – they are now truly about as reliable and predictive as they were thought to be back in 1982. Among these improvements, characterization of the convergence properties of Eqs. (1) and (2) was an important step forward, showing that, depending on the type of problem, one methodology should be preferred over the other [23, 22]. Improved understanding of the deleterious effects of insufficient equilibration [24, 25] and of the importance of adequately sampling slow-relaxing degrees of freedom [22] made it clear that quality free energy calculations are dependent not only on massive computational resources, but also on basic issues related to the underlying physics and on performing the computations properly. More fundamentally, from our improved understanding of free energy calculations, it is now obvious that such simulations cannot be performed as "black box" routine jobs that require little to no attention to the setup or the interpretation of the results.

Another approach – very closely related to FEP/TI calculations – that is capable of generating experimentally relevant free energy results is the "umbrella sampling" (US) method devised by Torrie and Valleau [26]. Like FEP/TI, application of the US method to complex systems of biological relevance was delayed for many years, until fast enough computers were widely available. Within US, sampling along some reaction coordinate, ξ, can be restrained to a limited portion of the configurational space by means of appropriately chosen biasing potentials, $\mathscr{V}_{\text{ext}}(\xi)$. Torrie and Valleau have demonstrated that the unbiased ensemble average of some quantity $\mathscr{A}$ can

be obtained from the biased ensemble *via*:

$$\langle \mathscr{A} \rangle = \frac{\langle \mathscr{A}\ e^{-\mathscr{V}_{\text{ext}}(\xi)/k_BT} \rangle_{\text{bias}}}{\langle e^{-\mathscr{V}_{\text{ext}}(\xi)/k_BT} \rangle_{\text{bias}}} \tag{3}$$

As a result, the free energy profile for a given ξ range, or "window" [27], can be readily derived from the probability, $\mathscr{P}_{\text{bias}}(\xi)$, of finding the system at the different values of ξ within that range, computed in the biased configurational space:

$$A(\xi) = -k_BT\ \ln \mathscr{P}_{\text{bias}}(\xi) - \mathscr{V}_{\text{ext}}(\xi) + A_0 \tag{4}$$

Here, A_0 is a constant. The early, 1982, work of Northrup *et al.* [28], although premonitory of future generalized use of US for related problems, also suggested that alternative methods like FEP might be more effective in handling such calculations. Subsequent work has not, however, necessarily demonstrated this to be the case, despite the publication of some nice attempts to construct potentials of mean force [29] (PMFs) using FEP and TI [30–32]. Instead, advancements in US and related methodologies have helped hold this latter approach at the forefront. Among these significant improvements is the weighted histogram analysis method [33, 34] (WHAM) that takes advantage of all possible statistical information collected in the various windows along reaction coordinate ξ, to generate, in a self-consistent fashion, the full free energy profile. In contrast, Eq. (4) suggests a manual match of the sub-profiles, $A(\xi)$, in their overlapping regions; this is not possible for two-dimensional, and higher, reaction coordinates [35, 36].

In the realm of PMF calculations, another noteworthy step forward was made by understanding the effects of constrained, internal coordinates on calculated free energy changes. Until the late 1980's, it was widely accepted that constraining flexible internal coordinates during an MD simulation – *e.g.*, using the SHAKE algorithm [37] to remove high frequency bond vibrations, thereby allowing the use of longer time-steps – effectively removed any contributions to the calculated free energy difference from these internal coordinates. It has been subsequently shown that there is, in fact, a contribution to the net free energy from these constrained degrees of freedom, and the procedure to calculate it has been described [38, 39]. In some instances, this free energy component can be appreciable. Furthermore, the assumption that sampling the constrained phase space would yield a free energy difference identical to that evaluated in an unconstrained phase space has been shown, in a number of cases, to be somewhat erroneous.

Characterization of the necessary corrections [40, 41] can bear critical consequences in those examples where chemical bonds are grown or shrunk in the course of an alchemical transformation, like a point mutation, or when torsional angles are modified to determine conformational free energy differences.

Either Monte Carlo (MC) or molecular dynamics (MD) simulations may be employed to generate the statistical ensemble of configurations necessary to evaluate the free energy according to Eqs. (1), (2) or (4). Roughly coinciding with the first published applications of FEP to large systems appeared the first of a long series of sophisticated algorithms that allow MD simulations to be carried out in important thermodynamic ensembles [42–46]. Free energy calculations have naturally benefited from these major developments. For example, one can select whether to simulate the isothermal–isobaric or canonical ensemble, depending on whether one wishes to evaluate the Gibbs or the Helmholtz free energy. Nearly a decade later, new, alternative methods for better modeling long-range, electrostatic forces started to flourish [47–49]. This better treatment of the electrostatics *via*, for instance, Ewald sums [50] or a linearized form of the Poisson equation [51, 52], now makes it possible to calculate properly free energies for systems that incorporate ionic species [53, 54].

Last and of equal importance, empirically-based potential energy functions, object of continuous enhancements ignited some twenty years ago, have now reached a point where they can predict free energies within an accuracy of 0.5–1.0 kcal/mol for a variety of chemical systems of differing sizes and complexities [55–57]. Just as it has been progressively realized that convergence in free energy calculations might is more difficult to reach than anticipated from the unduly optimistic, early results of the eighties, so, too, have deficiencies in the underlying assumptions used in the parametrization of all-purpose, molecular mechanical force fields become clearer. For example, a particularly large improvement in the force field was effected by inclusion of the aforementioned Ewald sums [58]. Another problematic issue is the description of an N-body problem by means of a pairwise, additive potential energy function. To address this issue, development of non-additive, polarizable force fields was described as early as 1972 [59]. Just as with free energy calculations, useful implementation of polarizable force fields had to wait for the era of supercomputers before such methods could practically be applied to large systems relevant to the modeler [60, 61]. Polarizable force fields remain a work in progress, and examples of free energy simulations incorporating polarizability are still few and far between.

2. ONGOING DIFFICULTIES IN FREE ENERGY CALCULATIONS

Despite dramatic increases in computational resources, the development of new, efficient algorithms, and the improvements in general-purpose, empirically-based potential energy functions over the last twenty years, the accurate estimation of free energy changes in large molecular assemblies still constitutes a challenge for modern theoretical chemistry.

Given this ongoing challenge, one must ask: Do free energy calculations really represent a predictive and affordable computational tool? This, in essence, is the question that determines the true worth of free energy calculations outside of a purely academic context. Even when taking advantage of parallel architectures, classical free energy simulations of interest to the pharmaceutical world can require several weeks to reach convergence, which not only precludes any routine application, but also calls into question their use in the development of new drugs. In any industrial setting, it is imperative that computational, *in silico*, experiments provide either a convincing answer faster than bench experiments would, or that they be able to rapidly supply qualitative answers that are predictive on, at least, a rank order basis – *e.g.*, protein-ligand binding free energies.

Cost-effectiveness considerations become even more crucial when one is cognizant of the intrinsic limitations of the models used, and wishes to go beyond the commonly accepted approximations introduced in classical molecular mechanics simulations, *e.g.*, pairwise additive potential energy functions, or spherical truncation of non-bonded interactions. Clearly, in a sizeable number of cases, such approximations can no longer be made without affecting the accuracy of the computed free energies. Assuming, however, that the modeler possesses a hypothetically, exact potential energy function, and that the conditions of the simulations are size-consistent, there is still no guarantee that the free energy evaluated from a computation of finite length will actually converge towards the correct answer, for free energy calculation is a slow-convergence problem [23, 22].

Ultimately, the modeler is left attempting to balance two opposing constraints. On the one hand, improvements in the accuracy of the model being used, such as the inclusion of polarization effects, increase the computational requirements of the methodology. Improvements in our understanding of convergence issues related to free energy calculations have also continually raised the bar on what is deemed a minimal, yet acceptable simulation length to achieve convergence. On the other hand, the calculations are frequently not of practical use unless they can be carried out in a

shorter amount of time than is currently required. One can only hope that advances in computer speed will allow a happy compromise to be achieved.

Even if all the above issues could be resolved, there remain recurring difficulties that – at least currently – must be considered on a case-by-case basis. Among these, the most problematic is probably that of selecting an appropriate reaction coordinate that connects those two states between which a free energy difference is to be computed. Whereas any reaction path is, in principle, acceptable, in practice, choosing a poor reaction coordinate frequently results in a simulation that has no realistic chance of converging. In some instances, the definition of an optimal reaction path is trivial, *e.g.*, point mutations in biological systems, or one-dimensional potentials of mean force. In other, more complex cases, however, the reaction path is multi-dimensional. Such is the case, for example, of protein folding, or the approach of a ligand towards a protein, for which the choice of an unequivocal reaction coordinate is clearly problematic.

In recent years, free energy simulations have, nevertheless, proven to be a powerful tool for estimating a variety of thermodynamically relevant quantities, such as association constants [62], second osmotic virial coefficients [63], and partition coefficients [64, 65]. In many cases, the free energy results compare remarkably well with experiment. The literature is particularly rich with examples of apparently successful free energy predictions for protein-ligand systems [7, 16–21]. It is worth noting that most of these simulations still reflect substantial amounts of computational and personal effort. Whereas in rare, favorable cases, such transformations can be simulated within a reasonable time, the modeler often faces quasi non-ergodic situations that result in a slow convergence of the ensemble average. This can be ascribed, among others, to slowly relaxing degrees of freedom [22].

3. WHERE WE STAND NOW

Methods to circumvent the aforementioned obstacles – to allow general use of free energy simulations as a routine, predictive tool – constituted the primary focus of a workshop entitled *Challenges in free energy calculations*, that was organized on June 19–21, 2000 at the Centre Européen de Calcul Atomique et Moléculaire (CECAM) in Lyon, France. This meeting followed a long heritage of scientific events focused on free energy calculations, made possible by the CECAM, and initiated some fifteen years ago by Herman Berendsen in Amersfoort, The Netherlands. Two ensuing workshops, held

in 1987 and 1995, in Orsay and Lyon, France, demonstrated the unceasing activity of the recently emerging community of free energy calculations.

During the June 2000 workshop, past and current applications of classical computational approaches for tackling problems relevant to the modeler were reviewed. Limitations of such approaches were delineated, and recent, cutting-edge methodological progresses in the field were discussed. Of somewhat more particular interest are the new methods targeted at the rapid computation of approximate free energy changes [22, 66]. Methods presented include those based on a Taylor expansion of the free energy with respect to a coupling parameter [67], and those based on a continuum description of the environment, generated by solving the Poisson equation or by making use of the generalized Born (GB) model [68].

The present special issue of Molecular Simulation gathers a number of contributions that cover these different fields, written by prominent participants in the free energy calculation community. Pursuing his analysis of constrained degrees of freedom in free energy calculations, Stefan Boresch presents an in-depth investigation on the role of non-bonded terms in "alchemical" transformations. Such transformations are also analyzed by Mihaly Mezei, emphasizing on the contribution of molecular flexibility to the estimated free energy differences. Related to the latter, Jed Pitera and Wilfred van Gunsteren compare non-bonded scaling methods in those free energy calculations that involve the creation or annihilation of particles [69].

The limiting case of Eqs. (1) and (2), referred to as "slow growth" (SG) approach [22], assumes that for an infinitesimal variation of λ, the difference in free energy between states a and b can be expressed as:

$$\Delta A_{a\to b} \simeq \sum_{i=0}^{M-1} \mathscr{H}(\mathbf{r}^N, \mathbf{p}^N; \lambda_{i+1}) - \mathscr{H}(\mathbf{r}^N, \mathbf{p}^N; \lambda_i) \tag{5}$$

where M is the number of windows utilized to connect states a and b. Hu Hao and Jan Hermans revisit this approach, often alleged to be flawed, since the continuous changes in λ mean that the system never reaches a true equilibrium, but rather lags behind the successively changing Hamiltonians [24, 70, 71]. Another article focusing on the "slow growth" approach is proposed by Gerhard Hummer. It investigates the identity that allows the computation of free energy differences for arbitrary growth rates [72]. Separating adiabatically the reaction coordinate and the remaining degrees of freedom of the system, Lula Rosso and Mark Tuckerman propose a novel scheme for generating free energy profiles for rare events. Closely related, the contribution of Eric Darve, Michael Wilson and Andrew

Pohorille tackles the intricate problem of slowly relaxing degrees of freedom in free energy calculations. Their approach, entitled "scaled-force" MD, allows energetic barriers to be lowered selectively without an *a priori* knowledge of their shape.

Moving in an alternative direction, several groups describe methods which trade off a bit of physical rigor to achieve improved sampling and speed. Wim Briels and Reinier Akkermans introduce an innovative and clever method for simulating long time-scale properties of polymer melts using coarse-grained models. Begoña Hernández, Jesus López, Modesto Orozco and Javier Luque discuss the merit of a novel approach for the fast estimation of free energy differences, based on multiple-copy MC, in conjunction with a continuum description of the surroundings.

Finally, elegant applications of free energy calculations are presented by Piotr Cieplak to study drug-enzyme and drug-DNA complexes, and by David Kombo, Jayaram, Kevin McConnell and David Beveridge to estimate the binding free energy of the λ-repressor-operator complex.

Acknowledgments

We are indebted to the Centre Européen de Calcul Atomique et Moléculaire (CECAM) for making the June 2000 workshop *Challenges in free energy calculations* possible. Michel Mareschal and Emmanuelle Crespeau are gratefully acknowledged for their help in organizing the meeting. We thank Dr. Surjit Dixit, Université Henri Poincaré, Vandœuvre-lés-Nancy, and Professor Philippe Hünenberger, ETHZ, Zürich, for helpful comments on the manuscript.

References

[1] Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures, *J. Chem. Phys.*, **3**, 300–313.
[2] Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys.*, **22**, 1420–1426.
[3] Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data, *J. Comp. Phys.*, **22**, 245–268.
[4] Beveridge, D. L. and DiCapua, F. M. (1989). Free energy *via* molecular simulation: Applications to chemical and biomolecular systems, *Annu. Rev. Biophys. Biophys.*, **18**, 431–492.
[5] Postma, J. P. M., Berendsen, H. J. C. and Haak, J. R. (1982). Thermodynamics of cavity formation in water: A molecular dynamics study, *Faraday Symp. Chem. Soc.*, **17**, 55–67.
[6] Warshel, A. (1982). Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron transfer and proton transfer reactions, *J. Phys. Chem.*, **86**, 2218–2224.
[7] Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R. and Kollman, P. A. (1987). Calculation of the relative change in binding free energy of a protein-inhibitor complex, *Science*, **235**, 574–576.

[8] Tembe, B. L. and McCammon, J. A. (1984). Ligand-receptor interactions, *Comp. Chem.*, **8**, 281–283.
[9] Jorgensen, W. L. and Ravimohan, C. (1985). Monte Carlo simulation of differences in free energies of hydration, *J. Chem. Phys.*, **83**, 3050–3054.
[10] Bash, P. A., Singh, U. C., Langridge, R. and Kollman, P. A. (1987). Free energy calculations by computer simulation, *Science*, **236**, 564–568.
[11] Straatsma, T. P. and Berendsen, H. J. C. (1988). Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations, *J. Chem. Phys.*, **89**, 5876–5886.
[12] Lee, F. S. and Warshel, A. (1992). A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations, *J. Chem. Phys.*, **97**, 3100–3107.
[13] Carlson, H. A., Nguyen, T. B., Orozco, M. and Jorgensen, W. L. (1993). Accuracy of free energies of hydration for organic molecules from 6-31G*-derived partial charges, *J. Comput. Chem.*, **14**, 1240–1249.
[14] Chipot, C., Millot, C., Maigret, B. and Kollman, P. A. (1994). Molecular dynamics free energy perturbation calculations. Influence of the truncation of long-range nonbonded interactions on the free energy of hydration of polar species, *J. Chem. Phys.*, **101**, 7953–7962.
[15] Pohorille, A. and Benjamin, I. (1991). Molecular dynamics of phenol at the liquid–vapor interface of water, *J. Chem. Phys.*, **94**, 5599–5605.
[16] Merz, K. M. and Kollman, P. A. (1989). Free energy calculations on protein stability: The Thr 157 to Val 157 mutation of T4 lysozyme, *J. Am. Chem. Soc.*, **111**, 5649–5658.
[17] Hirono, S. and Kollman, P. A. (1990). Calculation of the relative binding free energy of 2'GMP and 2'AMP to ribonuclease T1 using molecular dynamics/free energy perturbation approaches, *J. Mol. Biol.*, **212**, 197–209.
[18] Fleischman, S. H. and Brooks III, C. L. (1990). Protein-drug interactions: Characterization of inhibitor binding in complexes of dhfr with trimethoprim and related derivatives, *Proteins*, **7**, 52–61.
[19] Cummins, P. L., Ramnarayan, K., Singh, U. C. and Gready, J. E. (1991). Molecular dynamics/free energy perturbation study on the relative affinities of the binding of reduced and oxidized NADP to dihydrofolate reductase, *J. Am. Chem. Soc.*, **113**, 8247–8256.
[20] Reddy, M. R., Viswanadhan, V. N. and Weinstein, J. N. (1991). Relative differences in the binding free energies of human immunodeficiency virus 1 protease inhibitors: A thermodynamic cycle-perturbation approach, *Proc. Natl. Acad. Sci. USA*, **88**, 10287–10291.
[21] Miyamoto, S. and Kollman, P. A. (1993). What determines the strength of noncovalent association of ligands to proteins in aqueous solution?, *Proc. Natl. Acad. Sci. USA*, **90**, 8402–8406.
[22] Mark, A. E., Free Energy Perturbation Calculations. In: *Encyclopedia of Computational Chemistry*, Schleyer, P. V. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F. and Schreiner, P. R. (Eds.), Vol. 2. Wiley and Sons, Chichester, 1998, pp. 1070–1083.
[23] King, P. M., Free energy *via* molecular simulation: A primer. In: *Computer Simulation of Biomolecular Systems*: *Theoretical and Experimental Applications*, Van Gunsteren, W. F., Weiner, P. K. and Wilkinson, A. J. (Eds.), Vol. 2, ESCOM, Leiden, 1993, pp. 267–314.
[24] Pearlman, D. A. and Kollman, P. A. (1989). The lag between the Hamiltonian and the system configuration in free energy perturbation calculations, *J. Chem. Phys.*, **91**, 7831–7839.
[25] Berendsen, H. J. C., Incomplete equilibration: A source of error in free energy calculations. In: *Proteins*, *Structure*, *Dynamics and Design*, Renugopalakrishnan, V., Carey, P. R., Smith, I. C. P., Huang, S. G. and Storer, A. C. (Eds.), ESCOM, Leiden, 1991, pp. 384–392.
[26] Torrie, G. M. and Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling, *J. Comput. Phys.*, **23**, 187–199.
[27] Valleau, J. P. and Card, D. N. (1972). Monte Carlo estimation of the free energy by multistage sampling, *J. Chem. Phys.*, **57**, 5457–5462.
[28] Northrup, S. H., Pear, M. R., Lee, C. Y., McCammon, J. A. and Karplus, M. (1982). Dynamical theory of activated processes in globular proteins, *Proc. Natl. Acad. Sci. USA*, **79**, 4035–4039.

[29] Roux, B. (1995). The calculation of the potential of mean force using computer simulations, *Comput. Phys. Comm.*, **91**, 275–282.
[30] Tobias, D. J. and Brooks III, C. L. (1987). Calculation of free energy surfaces using the methods of thermodynamic perturbation theory, *Chem. Phys. Lett.*, **142**, 472–476.
[31] Wang, C. X., Liu, H. Y., Shi, Y. Y. and Huang, F. H. (1991). Calculations of relative free energy surfaces in configuration space using an integration method, *Chem. Phys. Lett.*, **179**, 475–478.
[32] Chipot, C., Kollman, P. A. and Pearlman, D. A. (1996). Alternative approaches to potential of mean force calculations: Free energy perturbation *versus* thermodynamic integration. Case study of some representative nonpolar interactions, *J. Comput. Chem.*, **17**, 1112–1131.
[33] Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. and Rosenberg, J. M. (1992). The weighted histogram analysis method for free energy calculations on biomolecules. I. The method, *J. Comput. Chem.*, **13**, 1011–1021.
[34] Boczko, E. M. and Brooks III, C. L. (1993). Constant-temperature free energy surfaces for physical and chemical processes, *J. Phys. Chem.*, **97**, 4509–4513.
[35] Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. and Kollman, P. A. (1995). Multidimensional free-energy calculations using the weighted histogram analysis method, *J. Comput. Chem.*, **16**, 1339–1350.
[36] Chipot, C. and Pohorille, A. (1998). Conformational equilibria of terminally blocked single amino acids at the water–hexane interface. A molecular dynamics study, *J. Phys. Chem. B*, **102**, 281–290.
[37] Ryckaert, J., Ciccotti, G. and Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion for a system with constraints: Molecular dynamics of n-alkanes, *J. Comput. Phys.*, **23**, 327–341.
[38] Pearlman, D. A. and Kollman, P. A. (1991). The overlooked bond-stretching contribution in free energy perturbation calculations, *J. Chem. Phys.*, **94**, 4532–4545.
[39] Straatsma, T. P., Zacharias, M. and McCammon, J. A. (1992). Holonomic constraint contributions to free energy differences from thermodynamic integration molecular dynamics simulations, *Chem. Phys. Lett.*, **196**, 297–302.
[40] Boresch, S. and Karplus, M. (1996). The Jacobian factor in free energy simulations, *J. Comp. Chem.*, **105**, 5145–5154.
[41] den Otter, W. K. and Briels, W. J. (1998). The calculation of free-energy differences by constrained molecular dynamics simulations, *J. Chem. Phys.*, **109**, 4139–4146.
[42] Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature, *J. Chem. Phys.*, **72**, 2384–2393.
[43] Nosé, S. and Klein, M. (1983). Constant pressure dynamics for molecular systems, *Mol. Phys.*, **50**, 1055–1076.
[44] Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys.*, **52**, 255–268.
[45] Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase–space distributions, *Phys. Rev.*, **A31**, 1695–1697.
[46] Martyna, G. J., Tobias, D. J. and Klein, M. L. (1994). Constant pressure molecular dynamics algorithms, *J. Chem. Phys.*, **101**, 4177–4189.
[47] Neumann, M. and Steinhauser, O. (1980). The influence of boundary conditions used in machine simulations on the structure of polar systems, *Mol. Phys.*, **39**, 437–54.
[48] York, D. M., Wlodawer, A., Pedersen, L. G. and Darden, T. A. (1994). Atomic-level accuracy in simulations of large protein crystals, *Proc. Natl. Acad. Sci. USA*, **91**, 8715–8718.
[49] Toukmaji, A. Y. and Board, J. A. Jr. (1996). Ewald summation techniques in perspective: A survey, *Comput. Phys. Comm.*, **95**, 73–92.
[50] Ewald, P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Ann. Phys.*, **64**, 253–287.
[51] Barker, J. A. (1994). Reaction field, screening, and long-range interactions in simulations of ionic and dipolar systems, *Mol. Phys.*, **83**, 1057–1064.
[52] Tironi, I. G., Sperb, R., Smith, P. E. and van Gunsteren, W. F. (1995). A generalized reaction field method for molecular dynamics simulations, *J. Chem. Phys.*, **102**, 5451–5459.

[53] Hünenberger, P. H. and McCammon, J. A. (1999). Ewald artifacts in computer simulations of ionic solvation and ion-ion interactions: A continuum electrostatics study, *J. Chem. Phys.*, **110**, 1856–1872.
[54] Rozanska, X. and Chipot, C. (2000). Modeling ion-ion interaction in proteins: A molecular dynamics free energy calculation of the guanidinium-acetate association, *J. Chem. Phys.*, **112**, 9691–9694.
[55] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. C. and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.*, **117**, 5179–5197.
[56] MacKerell, A. D. Jr., Bashford, D., Bellott, M., Dunbrack, R. L. Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher III, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D. and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B*, **102**, 3586–3616.
[57] Scott, W. R. P., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P. and van Gunsteren, W. F. (1999). The GROMOS biomolecular simulation program package, *J. Phys. Chem. A*, **103**, 3596–3607.
[58] Cheatham III, T. E., Miller, J. L., Fox, T., Darden, T. A. and Kollman, P. A. (1995). Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins, *J. Am. Chem. Soc.*, **117**, 4193–4194.
[59] Berendsen, H. J. C. (Ed.), *MD and MC on water* (Centre Européen de Calcul Atomique et Moléculaire: Orsay, France, 1972), CECAM, workshop.
[60] New, M. H. and Berne, B. J. (1995). Molecular Dynamics Calculation of the Effect of Solvent Polarizability on the Hydrophobic Interaction, *J. Am. Chem. Soc.*, **117**, 7172–7179.
[61] Soetens, J. C., Millot, C., Chipot, C., Jansen, G., Ángyán, J. G. and Maigret, B. (1997). Effect of polarizability on the calculation of potential of mean force. The guanidinium–guanidinium ion pair in water, *J. Phys. Chem. B*, **101**, 10910–10917.
[62] Duffy, E. M., Kowalczyk, P. J. and Jorgensen, W. L. (1993). Do denaturants interact with aromatic hydrocarbons in water?, *J. Am. Chem. Soc.*, **115**, 9271–9275.
[63] Jorgensen, W. L. and Severance, D. L. (1990). Aromatic–aromatic interactions: Free energy profiles for the benzene dimer in water, chloroform and liquid benzene, *J. Am. Chem. Soc.*, **112**, 4768–4774.
[64] Daura, X., Hünenberger, P. H., Mark, A. E., Querol, E., Avilès, F. X. and van Gunsteren, W. F. (1996). Free energies of transfer of Trp analogs from chloroform to water: Comparison of theory and experiment and the importance of adequate treatment of electrostatic and internal interactions, *J. Am. Chem. Soc.*, **118**, 6285–6294.
[65] Chipot, C., Wilson, M. A. and Pohorille, A. (1997). Interactions of anesthetics with the water–hexane interface. A molecular dynamics study, *J. Phys. Chem. B*, **101**, 782–791.
[66] Liu, S. Y., Mark, A. E. and van Gunsteren, W. F. (1996). Estimating the relative free energy of different molecular states with respect to a single reference state, *J. Phys. Chem.*, **9485–9494**, 1749.
[67] Smith, P. E. and van Gunsteren, W. F. (1994). Predictions of free energy differences from a single simulation of the initial state, *J. Chem. Phys.*, **100**, 577–585.
[68] Jayaram, B., Sprous, D. and Beveridge, D. L. (1998). Solvation free energy of biomacromolecules: Parameters for a modified generalized born model consistent with the AMBER force field, *J. Phys. Chem. B*, **102**, 9571–9576.
[69] Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R. and van Gunsteren, W. F. (1994). Avoiding singularities and neumerical instabilities in free energy calculations based on molecular simulations, *Chem. Phys. Lett.*, **222**, 529–539.
[70] Wood, R. H. (1991). Estimation of errors in free energy calculations due to the lag between the Hamiltonian and the system configuration, *J. Phys. Chem.*, **95**, 4838–4842.
[71] Hermans, J. (1991). Simple analysis of noise and hysteresis in (slow-growth) free energy simulations, *J. Phys. Chem.*, **95**, 9029–9032.
[72] Jarzynski, C. (1997). Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.*, **78**, 2690–2693.