ELSEVIER

Enhancing the accuracy, the efficiency and the scope of free energy simulations

Tomas Rodinger and Régis Pomès

Many different methods exist for computing free energy changes from molecular simulations. Recent advances have led to improvements in the theoretical framework underlying these calculations, as well as in the accuracy and sampling efficiency of the algorithms. Novel methods combining the advantages afforded by various existing approaches offer promising strategies and open up new perspectives to help elucidate the physical basis of important biological processes.

Addresses
Structural Biology and Biochemistry, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

Corresponding author: Pomès, Régis (pomes@sickkids.ca)

Current Opinion in Structural Biology 2005, **15**:164–170

This review comes from a themed issue on
Theory and simulation
Edited by J Andrew McCammon and Rebecca C Wade

Available online 9th March 2005

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2005.03.001

Introduction

The calculation of free energy from molecular simulations is an area of intense research activity. This is because free energy is at once one of the most central and one of the most difficult quantities to compute from atomic-level simulations. It is of paramount importance in efforts to relate microscopic details stemming from atomic interactions to measurable macroscopic quantities, and to understand the physical and structural basis of biological phenomena. Desirable biophysical applications of free energy calculations include screening for or characterizing enzyme–ligand binding, elucidating protein folding equilibria and determining free energy profiles underlying transport processes.

The principal practical difficulty facing these calculations arises from the complexity and the ruggedness of the energy landscape underlying conformational fluctuations of biological macromolecules. Many transformations of biological interest involve conformational changes that span time and/or length scales currently beyond the reach of typical molecular dynamics simulations (i.e. longer than 10^{-9} to 10^{-8} s). To work around this limitation, many different methods have been developed to calculate free energy changes. These methods are, to some extent, tailored to specific problems. They may be performed either under conditions approximating thermodynamic reversibility (equilibrium) or not; the exploration of phase/conformational space (sampling) may be either restrained or unrestrained; the transformation may be either local or non-local; and it may follow a physical or an alchemical pathway — or no pathway at all. Recent theoretical efforts have resulted in promising developments that combine some of the advantages afforded by various approaches to widen the scope of the calculations, and to reduce both systematic and statistical sampling errors — that is, to improve both the accuracy and the efficiency of the calculations. This review is an attempt to highlight some of these promising developments.

Extensive reviews of free energy calculations can be found elsewhere [1–6]. Here, we focus on some of the theoretical and methodological advances that are likely to affect the calculation of free energies from molecular simulations of biological systems. Because of space constraints, we limit the scope of this survey to classical simulations involving configurational averaging. It should be noted that advances in relevant application fields, such as free energy calculation of pK_as, molecular recognition and transport in membrane proteins, are reviewed elsewhere is this issue.

In the following discussion, free energy simulation methods have been loosely grouped into three broad classes: free energy perturbation (FEP) and other non-equilibrium work (NEW) approaches based on exponential averaging of the accumulated work; thermodynamic integration (TI); and replica exchange (RE) methods. These categories are neither mutually exclusive nor exhaustive.

Free energy perturbation and non-equilibrium work methods

In FEP [7], the relative free energy (ΔF) between two distinct thermodynamic states labeled 0 and 1 is computed from the expectation value of the Boltzmann-weighted energy difference:

$$e^{-\beta\Delta F} = \langle e^{-\beta\Delta U} \rangle_0 \quad (1)$$

where $\beta = (k_B T)^{-1}$ is the reciprocal temperature, with k_B the Boltzmann constant and T the absolute temperature, $\Delta U = U_1 - U_0$ is the potential energy of perturbed state 1 relative to reference state 0, and the angle brackets with subscript 0 denote Boltzmann-weighted ensemble averaging performed in the reference state.

FEP methods have been widely used in biomolecular free energy calculations. Recent analyses and developments have stressed the relationship of FEP to NEW methods, which rest on Jarzynski's identity [8,9]:

$$e^{-\beta\Delta F} = \langle e^{-\beta W} \rangle_0 \quad (2)$$

where the averaging is performed over independent measurements of the accumulated work (W) for transformations (or switching) starting from equilibrium state 0 [8,9]. Jarzynski's identity (Equation 2) is remarkable in light of the second law of thermodynamics, which states that free energy is a lower bound of the average of the accumulated work:

$$\Delta F \leq \langle W \rangle \quad (3)$$

If the process is infinitely slow (i.e. if it follows a reversible path), then W is the reversible work or potential of mean force (PMF), and Equations 2 and 3 give $\Delta F = W$ throughout the path. However, under non-equilibrium conditions (whereby switching takes place over a finite time), Equation (2) still holds, whereas Equation 3 is an inequality. FEP (Equation 1) is another limiting case of Equation 2, whereby the transformation takes place instantly.

Jarzynski's identity (Equation 2) establishes the relevance of non-equilibrium processes to the calculation of equilibrium properties [8–10], and provides a rigorous theoretical foundation for calculating free energy changes from single-molecule experiments [11–13] and from simulations deliberately driven out of equilibrium. Several research groups have seized this opportunity to explore applications of the approach to the calculation, using Equation 2, of free energy changes from series of simulations in which the accumulated work is computed over a finite time (switching); these efforts have been accompanied by further theoretical developments aimed at clarifying and minimizing both systematic and statistical errors in FEP and NEW calculations [14,15••,16–19,29••,21•–23•]).

Because of the non-linear nature of the Boltzmann factor in Equations 1 and 2, under non-equilibrium conditions, only a subset of important (thermally accessible) configurations contribute significantly to exponential averaging. This introduces a systematic sampling bias that can lead to substantial inaccuracies (reproducible errors) in both computational and experimental determinations of free energy changes. Thus, obtaining accurate estimates from Equation 1 requires that all important configurations of the target state are also important to the reference state, in other words, that perturbed state 1 forms a subset of reference state 0 [20••]. Because this condition is realized trivially in the limit of infinitely small perturbations (identical distributions), it is common practice to break down or stage FEP calculations over a sequence of intermediate states that differ from their neighbors by small perturbations; however, this approach is inefficient and does not directly address the systematic bias problem, it only lessens it.

A common practice in FEP and NEW calculations is to perform simulations in both forward and reverse directions (i.e. 0→1 and 1→0, respectively). This is often achieved with double-wide sampling, whereby forward and reverse perturbations are performed simultaneously at intermediate stages between 0 and 1 [24]. Because the subset requirement cannot be satisfied simultaneously by both states of interest, simple averaging of forward and reverse simulations from Equation 1 does not lead to the cancellation of systematic errors. Instead, it was recently shown by Lu *et al.* [20••] that sampling inaccuracies can be eliminated by combining the results of forward and reverse FEP calculations as follows:

$$e^{-\beta\Delta F} = \langle w(\Delta U) e^{-\beta\Delta U/2} \rangle_0 / \langle w(\Delta U) e^{+\beta\Delta U/2} \rangle_1 \quad (4)$$

where $w(\Delta U)$ is a weighting function. Equation 4 is termed overlap sampling because it introduces a virtual intermediate state in which only overlapping configurations of importance to both state 0 and state 1 are important. The choice of $w(\Delta U) = 1$ provides a simple and effective prescription to combine forward and reverse calculations. Optimal results in terms of both systematic and statistical errors are obtained with a different choice of $w(\Delta U)$; this amounts to an older method, Bennett's acceptance ratio [20••,25]. The overlap sampling approach has been generalized for NEW methods and tested with model systems [21•,22•]. Another interesting approach for improving the accuracy and efficiency of free energy calculations combines the NEW approach with transition path sampling [26••,27]. Transition path sampling ([28,29] and references therein) is a powerful technique for generating ensembles of plausible pathways between initial and final states of interest. A different way of using transition path sampling in free energy calculations has also been reported recently [30•].

An important and natural application of Jarzynski's identity (Equation 2) is the calculation of the PMF or free energy profile along a predetermined reaction coordinate. Recent work [15••,31] has shown that this can rigorously be achieved using steered molecular dynamics (SMD) simulations, in which the system is pulled by an artificial constraint, such as a harmonic spring, along a prescribed path in configuration space [32]. The approach has been applied to glycerol conduction through the aquaglyceroporin GlpF ([33]; see also the review by de Groot and Grubmüller in this issue), ammonia conduction through HisF [34] and the helix-coil transition of deca-alanine [31]. However, although Equation 2 is valid for arbitrary pulling rates, several studies comparing NEW methods with other approaches [15••,21•,23•] point out that the

approach works best when the system remains close to equilibrium.

From thermodynamic integration to extended ensembles

In TI methods, the free energy is obtained from the expectation value of the derivative of the energy with respect to a coupling parameter (λ) [35]:

$$\partial(\Delta F)/\partial\lambda = \langle \partial U/\partial\lambda \rangle_\lambda \quad (5)$$

The dependence of U on λ defines the path of the transformation between two states of interest. Integrating Equation (5) over this path yields the free energy difference (ΔF). The transformation can be performed in two different ways. One is to change λ continuously between reference and target states [36]. This method, which may be called 'slow growth' or 'fast growth' depending on the rate of the transformation, falls into the category of NEW approaches because the system constantly lags behind the Hamiltonian [14]. The other way to integrate Equation 5, sometimes called finite-difference TI, is to perform a series of equilibrium simulations at discrete intermediate values of λ and combine the results by interpolation. In NEW, TI and staged FEP approaches, λ may be either an abstract parameter devised to couple or transform the system along an artificial pathway in so-called 'alchemical' transformations (e.g. of one chemical group into another) [37] or a physical degree of freedom, such as interatomic separation or another conformational variable. In the latter case, the calculation yields the reversible work or PMF and the right-hand side of Equation 5 is the mean force acting along the predefined variable λ in the average of all other conformational degrees of freedom.

To obtain converged statistics in free energy simulations, adequate sampling of conformational space is required. Free energy barriers dotting the rugged conformational space of biological macromolecules commonly make this difficult to achieve. Several approaches that treat λ as a dynamic reaction coordinate exist to overcome this problem. What these methods have in common is the emphasis on choosing effective pathways between thermodynamic states of interest on free energy surfaces and improving statistical sampling efficiency along these pathways. The non-equilibrium methods highlighted in the preceding section offer this possibility. Other, widely used approaches make use of umbrella sampling [38], which consists of imposing a potential energy bias to enforce uniform sampling along the reaction coordinate. Two unconstrained methods utilizing the formalism of TI, adaptive force sampling [39] and generalized ensemble methods [40••], have recently been proposed. The advantage of these unconstrained methods, as indeed of RE methods (see below), is that they allow self-optimization of sampling.

Adaptive force sampling consists of using Equation 5 without restrictions on the values taken by the reaction coordinate λ, in such a way that sampling along the selected degree(s) of freedom of λ (which may be mono- or multi-dimensional) becomes diffusive [39]. Two recent studies illustrate the effectiveness of the adaptive force approach in improving convergence rates for several molecular problems and provide critical discussions [23•,41•].

In so-called λ dynamics methods [42], the coupling parameter is replaced by a non-physical dynamic variable in an effort to improve sampling efficiency. Free energy changes can be obtained as a PMF in this extended system. Such an approach was used for competitive binding calculations [43]. Likewise, the addition of an extra non-physical ('fourth') spatial dimension was shown to provide a simple and low-barrier route to computing the free energy of inserting or extracting a molecule of interest into or out of a dense phase [44]. In a recent extension of λ dynamics, free energy differences are obtained from TI in a generalized ensemble [40••]. In this scheme, two separate systems whose respective λ values are coupled to each other are simulated at once. The procedure effectively introduces temperature scaling, which further enhances barrier crossing and results in significant improvements in statistical efficiency. Some of these traits are shared with RE methods.

Replica exchange methods

RE algorithms provide very effective ways to sample rugged energy surfaces. In this scheme, simulations of many copies or replicas of the system are performed in parallel, each with discrete values covering a range of temperatures and pressures, or with different Hamiltonians, and are periodically allowed to swap their coordinates according to a Metropolis algorithm (see [45] and references therein). Exhaustive phase space sampling makes it possible to compute relative free energies directly, without presuming a reaction coordinate, or delineating initial and final states of the system *a priori*.

The application of RE algorithms to protein folding has enabled a quantum leap in the size of peptide, up to small proteins, for which the folding-unfolding equilibrium can be sampled exhaustively. This approach has been exploited most actively by Garcia and co-workers (see [46] and references therein). Recent applications have extended the method to a 46 amino acid protein domain [47•] and into the realm of pressure as well as temperature sampling [48]. These achievements are all the more significant, from the perspective of free energy simulations, because conformational sampling is not artificially driven but instead takes place without any restraints. This makes it possible to re-examine the nature of adequate descriptors of conformational states and folding equilibria [49].

In a promising development, RE approaches were recently combined with FEP and TI techniques [50••]. Improved precision due to efficient sampling is demonstrated for the calculation of the relative hydration free energy of water and methane [50••], and of the relative affinity of halide ions for a molecular host [51].

Other developments

In the calculation of equilibrium ensemble averages (e.g. Equation 5), systematic errors arise if data from early, 'equilibration' stages of the simulation are included. The universal practice of eliminating initial data from cumulative averaging performed in the forward direction is unreliable and inefficient. Yang *et al.* [52•] introduce a commendable procedure for data analysis that is instead based on cumulative averaging in the reverse direction.

Obtaining accurate quantitative estimates of solvation free energies is essential for understanding biomolecular association, as well as transfer and transport phenomena. Improved sampling fosters the wider use of explicit solvent. In recent years, it has become possible to compute absolute hydration free energies with a statistical precision comparable to that of experimental measurements themselves. This makes it possible to evaluate the accuracy of empirical force fields and simulation protocols in many cases for which the free energy can be measured experimentally, such as absolute hydration free energies of amino acids [53•,54] and water-to-hexane transfer free energies [55].

Full conformational sampling of peptides and small proteins also allows a critical assessment of force fields [46]. Detailed comparisons of results obtained for the folding of peptides from explicit solvent and from continuum models of solvation, which are used extensively in folding and binding studies, are provided by Nymeyer and García [56], and by Stultz [57] (also see the review by Feig and Brooks [58]). Because of the acute sensitivity of the folding equilibrium to hydration, these studies constitute a particularly stringent test of solvation models.

Binding free energy and entropy calculations

Biomolecular association, like protein folding, is driven by the near compensation of enthalpy and entropy. In addition to enthalpic contributions arising from specific interactions between ligand and receptor molecules, contributions involving both enthalpic and entropic changes, such as the desolvation of ligand and receptor binding site and changes in conformational free energy upon binding, also need to be addressed. The latter aspect constitutes an outstanding challenge, especially if conformational reorganization spans large length and time scales, which is generally the case with proteins. To help alleviate this problem, thermodynamic cycles [37] can be used advantageously, for example, to calculate the relative affinities of small compounds for proteins from alchemical simulations. This strategy has recently been exploited by Oostenbrink *et al.* [59,60•] to compute the binding affinities of multiple ligands from a single-step FEP simulation.

Methodological advances have recently been made in the calculation of absolute binding affinities. Absolute binding free energies were first calculated by Hermans and Shankar [61], and the theoretical basis underlying the standard-state dependence of absolute binding affinities was laid out in several studies [62–65]. More recently, Boresch *et al.* [66•] and Swanson *et al.* [67•] presented practical methods for the direct determination of contributions to association free energies that arise from the loss of translational and rotational freedom of the ligand upon binding to a receptor at the standard-state concentration. These advances make it possible to compare theoretical predictions with experimental binding affinities.

Several recent studies examine the thermodynamic basis for the hydration of protein cavities [68–70,71•,72•]. The results of these studies underline the importance of entropic contributions. Other efforts to obtain estimates of entropy from molecular simulations have focused primarily on the calculation of conformational entropy of solute molecules (for selected internal degrees of freedom) [73,74,75•] and on the extension of free energy simulation methods to the calculation of total entropy changes, which include not only solute but also solvent degrees of freedom. The latter has been discussed by Lu *et al.* [76•] and by Peter *et al.* [77•] in the context of simple solutions. Total enthalpy and entropy changes can also be obtained directly from free energy calculations performed at different temperatures, assuming that the free energy is linear over the temperature range [47•].

Conclusions

The development of free energy simulation methods is an area of continuing research activity. An outstanding difficulty is dealing with rugged energy landscapes with efficient algorithms. Many of the existing methods may perform better along specific pathways or for specific types of problems. However, the application domain of the various approaches often remains poorly defined, which underlines the need for extensive comparisons using appropriate model systems. Although a general method is not presently available, some converging trends are emerging. Particularly promising strategies are those that integrate various theoretical developments, expanding their respective scope and blurring the traditional distinctions between them. Together with ever-increasing computational power, advances in theoretical foundations and creative combinations of methods designed to enhance sampling efficiency have opened up new perspectives to improve the accuracy and precision of free energy calculations, and to help elucidate the physical basis of important biological processes, including molecular recognition, permeation and protein folding.

Acknowledgements

Funding from the Canadian Institutes of Health Research (operating grant MOP43998) is gratefully acknowledged. TR holds a scholarship from the Natural Sciences and Engineering Research Council of Canada. RP is a Canada Research Chair Program chairholder.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Beveridge DL, Dicapua FM: **Free-energy via molecular simulation - applications to chemical and biomolecular systems**. *Annu Rev Biophys Biophys Chem* 1989, **18**:431-492.

2. Straatsma TP, McCammon JA: **Computational alchemy**. *Annu Rev Phys Chem* 1992, **43**:407-435.

3. Chipot C, Pearlman DA: **Free energy calculations. The long and winding gilded road**. *Mol Simul* 2002, **28**:1-12.

4. van Gunsteren WF, Daura X, Mark AE: **Computation of free energy**. *Helv Chim Acta* 2002, **85**:3113-3129.

5. Frenkel D, Smit B: *Understanding Molecular Simulation: From Algorithms To Applications*, edn 2. San Diego: Academic Press; 2002.

6. Simonson T, Archontis G, Karplus M: **Free energy simulations come of age: protein-ligand recognition**. *Acc Chem Res* 2002, **35**:430-437.

7. Zwanzig RW: **High-temperature equation of state by a perturbation method 1. Nonpolar gases**. *J Chem Phys* 1954, **22**:1420-1426.

8. Jarzynski C: **Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach**. *Phys Rev E* 1997, **56**:5018-5035.

9. Jarzynski C: **Nonequilibrium equality for free energy differences**. *Phys Rev Lett* 1997, **78**:2690-2693.

10. Crooks GE: **Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems**. *J Stat Phys* 1998, **90**:1481-1487.

11. Hummer G, Szabo A: **Free energy reconstruction from nonequilibrium single-molecule pulling experiments**. *Proc Natl Acad Sci USA* 2001, **98**:3658-3661.

12. Liphardt J, Dumont S, Smith SB, Tinoco I, Bustamante C: **Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality**. *Science* 2002, **296**:1832-1835.

13. Gore J, Ritort F, Bustamante C: **Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements**. *Proc Natl Acad Sci USA* 2003, **100**:12564-12569.

14. Hummer G: **Fast-growth thermodynamic integration: error and efficiency analysis**. *J Chem Phys* 2001, **114**:7330-7337.

15. •• Park S, Schulten K: **Calculating potentials of mean force from steered molecular dynamics simulations**. *J Chem Phys* 2004, **120**:5946-5961.
This paper reviews and extends the theoretical foundations of non-equilibrium methods. SMD is used to irreversibly transform the system from the initial to the final state over a predefined configurational pathway. Results for the unfolding of a deca-alanine peptide by pulling with stiff springs clearly indicate that CPU resources are used most efficiently when fewer non-equilibrium trajectories are used, and when each is transformed slower and thus closer to thermodynamic equilibrium.

16. Zuckerman DM, Woolf TB: **Systematic finite-sampling inaccuracy in free energy differences and other nonlinear quantities**. *J Stat Phys* 2004, **114**:1303-1323.

17. Zuckerman DM, Woolf TB: **Theory of a systematic computational error in free energy differences**. *Phys Rev Lett* 2002, **89**:180602.

18. Zuckerman DM, Woolf TB: **Overcoming finite-sampling errors in fast-switching free-energy estimates: extrapolative analysis of a molecular system**. *Chem Phys Lett* 2002, **351**:445-453.

19. Lu N, Adhikari J, Kofke DA: **Variational formula for the free energy based on incomplete sampling in a molecular simulation**. *Phys Rev E* 2003, **68**:026122.

20. •• Lu N, Singh JK, Kofke DA: **Appropriate methods to combine forward and reverse free-energy perturbation averages**. *J Chem Phys* 2003, **118**:2977-2984.
Lu *et al*. develop overlap sampling, a general method for eliminating the systematic sampling bias arising from exponential averaging in FEP calculations. In its simplest form, the approach is easy to apply and is formally equivalent to introducing in the perturbation a virtual intermediate state consisting of a subset of configurations common to both the reference and target states; however, it only requires two (forward and reverse) simulations between these. Bennett's acceptance ratio is shown to correspond to an optimal version of overlap sampling for averaging forward and reverse simulations, in terms of both accuracy and precision (statistical error).

21. • Lu N, Woolf TB: **Overlap perturbation methods for computing alchemical free energy changes: variants, generalizations and evaluations**. *Mol Phys* 2004, **102**:173-181.
Lu and Woolf provide a detailed comparison of the results obtained from various protocols for averaging forward and reverse FEP and NEW simulations, including overlap perturbation (see also [20••]).

22. • Lu N, Wu D, Woolf TB, Kofke DA: **Using overlap and funnel sampling to obtain accurate free energies from nonequilibrium work measurements**. *Phys Rev E* 2004, **69**:05772.
This paper extends and generalizes the overlap sampling method developed for FEP simulations to the more general context of NEW methods.

23. • Rodriguez-Gomez D, Darve E, Pohorille A: **Assessing the efficiency of free energy calculation methods**. *J Chem Phys* 2004, **120**:3563-3578.
The authors compare Jarzynski's identity method with the adaptive biasing force method of Darve and Pohorille [39], and find that the latter works better when the system is easily driven away from equilibrium.

24. Jorgensen WL, Ravimohan C: **Monte Carlo simulation of differences in free energies of hydration**. *J Chem Phys* 1985, **83**:3050-3054.

25. Bennett CH: **Efficient estimation of free-energy differences from Monte-Carlo data**. *J Comp Phys* 1976, **22**:245-268.

26. •• Ytreberg FM, Zuckerman DM: **Single-ensemble nonequilibrium path-sampling estimates of free energy differences**. *J Chem Phys* 2004, **120**:10876-10879.
In this paper, the NEW method for a transformation of interest is optimized by iteratively refining reaction pathways using transition path sampling. Applications considered here and in [27] show a very substantial improvement in sampling efficiency.

27. Ytreberg FM, Zuckerman DM: **Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems**. *J Comput Chem* 2004, **25**:1749-1759.

28. Pratt LR: **A statistical-method for identifying transition-states in high dimensional problems**. *J Chem Phys* 1986, **85**:5045-5048.

29. Dellago C, Bolhuis PG, Geissler PL: **Transition path sampling**. In *Advances in Chemical Physics*, vol 123. John Wiley & Sons; 2002:1-78.

30. • Radhakrishnan R, Schlick T: **Biomolecular free energy profiles by a shooting/umbrella sampling protocol "BOLAS"**. *J Chem Phys* 2004, **121**:2436-2444.
A novel approach combining transition path sampling with umbrella sampling is presented and applied to a sugar puckering transition.

31. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K: **Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality**. *J Chem Phys* 2003, **119**:3559-3566.

32. Isralewitz B, Gao M, Schulten K: **Steered molecular dynamics and mechanical functions of proteins**. *Curr Opin Struct Biol* 2001, **11**:224-230.

33. Jensen MO, Park S, Tajkhorshid E, Schulten K: **Energetics of glycerol conduction through aquaglyceroporin GlpF**. *Proc Natl Acad Sci USA* 2002, **99**:6731-6736.

34. Amaro R, Tajkhorshid E, Luthey-Schulten Z: **Developing an energy landscape for the novel function of a (beta/alpha)(8) barrel: ammonia conduction through HisF**. *Proc Natl Acad Sci USA* 2003, **100**:7599-7604.

35. Kirkwood JG: **Statistical mechanics of fluid mixtures**. *J Chem Phys* 1935, **3**:300-313.

36. Mitchell MJ, McCammon JA: **Free-energy difference calculations by thermodynamic integration - difficulties in obtaining a precise value**. *J Comput Chem* 1991, **12**:271-275.

37. Tembe BL, McCammon JA: **Ligand receptor Interactions**. *Comput Chem* 1984, **8**:281-283.

38. Torrie GM, Valleau JP: **Non-physical sampling distributions in Monte-Carlo free-energy estimation - umbrella sampling**. *J Comput Phys* 1977, **23**:187-199.

39. Darve E, Pohorille A: **Calculating free energies using average force**. *J Chem Phys* 2001, **115**:9169-9183.

40. •• Bitetti-Putzer R, Yang W, Karplus M: **Generalized ensembles serve to improve the convergence of free energy simulations**. *Chem Phys Lett* 2003, **377**:633-641.
To improve the statistical convergence of free energy calculations, a generalized ensemble is introduced in which the integration parameter is treated as a dynamic variable. Free energy differences are obtained from TI. The availability of extra degrees of freedom, together with the artificial temperature scaling resulting from conjugate momenta of the integration variable, helps overcome kinetic traps. Reduced energy barriers lead to enhanced conformational sampling of physical space. The rate of convergence of the hydration free energy of histidine improves by an order of magnitude compared to standard TI.

41. • Hénin J, Chipot C: **Overcoming free energy barriers using unconstrained molecular dynamics simulations**. *J Chem Phys* 2004, **121**:2904-2914.
The authors combine unconstrained simulations with adaptive biasing forces for homogeneous sampling of phase space along a spatial reaction coordinate. They show how the method can be used effectively in calculating PMFs for the unfolding of a deca-alanine peptide *in vacuo*, the association of guanidinium and acetate, the association of two methane molecules in water and the migration of methane across the water–vapor interface.

42. Kong XJ, Brooks CL: **Lambda-dynamics: a new approach to free energy calculations**. *J Chem Phys* 1996, **105**:2414-2423.

43. Guo ZY, Brooks CL: **Rapid screening of binding affinities: application of the lambda-dynamics method to a trypsin-inhibitor system**. *J Am Chem Soc* 1998, **120**:1920-1921.

44. Pomès R, Eisenmesser E, Post CB, Roux B: **Calculating excess chemical potentials using dynamic simulations in the fourth dimension**. *J Chem Phys* 1999, **111**:3387-3395.

45. Nymeyer H, Gnanakaran S, Garcia AE: **Atomic simulations of protein folding, using the replica exchange algorithm**. *Methods Enzymol* 2004, **383**:119-149.

46. Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, Garcia AE: **Peptide folding simulations**. *Curr Opin Struct Biol* 2003, **13**:168-174.

47. • Garcia AE, Onuchic JN: **Folding a protein in a computer: an atomic description of the folding/unfolding of protein A**. *Proc Natl Acad Sci USA* 2003, **100**:13898-13903.
REMD simulations provide a thorough description of the folding-unfolding mechanism of fragment B of protein A, a three-helix bundle of 46 amino acid residues. This is the largest protein studied to date with explicit solvent. Free energy, entropy and enthalpy surfaces computed over a wide temperature range characterize the thermal dependence of the folding-unfolding equilibrium and provide a quantitative picture of the enthalpy-entropy compensation modulating the equilibrium.

48. Paschek D, Garcia AE: **Reversible temperature and pressure denaturation of a protein fragment: a replica exchange molecular dynamics simulation study**. *Phys Rev Lett* 2004, **93**:238105.

49. Krivov SV, Karplus M: **Hidden complexity of free energy surfaces for peptide (protein) folding**. *Proc Natl Acad Sci USA* 2004, **101**:14766-14770.

50. •• Woods CJ, Essex JW, King MA: **The development of replica-exchange-based free-energy methods**. *J Phys Chem B* 2003, **107**:13703-13710.
This work combines the high sampling efficiency afforded by RE methods with finite-difference TI and FEP calculations. The method is tested and compared to more conventional approaches. It is argued that the method works better than adaptive methods because the latter are driven (i.e. non-equilibrium).

51. Woods CJ, Essex JW, King MA: **Enhanced configurational sampling in binding free-energy calculations**. *J Phys Chem B* 2003, **107**:13711-13718.

52. • Yang W, Bitetti-Putzer R, Karplus M: **Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence**. *J Chem Phys* 2004, **120**:2618-2628.
This paper presents a rigorous and straightforward procedure for removing non-equilibrium contributions from ensemble averages in equilibrium methods.

53. • Shirts MR, Pitera JW, Swope WC, Pande VS: **Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins**. *J Chem Phys* 2003, **119**:5740-5761.
The authors exploit the massive computing power provided by distributed computing to calculate the hydration free energy of amino acid sidechains in several commonly used biomolecular force fields.

54. Deng YQ, Roux B: **Hydration of amino acid side chains: nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules**. *J Phys Chem B* 2004, **108**:16567-16576.

55. MacCallum JL, Tieleman DP: **Calculation of the water-cyclohexane transfer free energies of neutral amino acid side chain analogs using the OPLS all-atom force field**. *J Comput Chem* 2003, **24**:1930-1935.

56. Nymeyer H, García AE: **Simulation of the folding equilibrium of a-helical peptides: a comparison of the generalized Born approximation with explicit solvent**. *Proc Natl Acad Sci USA* 2003, **100**:13934-13939.

57. Stultz CM: **An assessment of potential of mean force calculations with implicit solvent models**. *J Phys Chem B* 2004, **108**:16525-16532.

58. Feig M, Brooks CL: **Recent advances in the development and application of implicit solvent models in biomolecule simulations**. *Curr Opin Struct Biol* 2004, **14**:217-224.

59. Oostenbrink C, van Gunsteren WF: **Single-step perturbations to calculate free energy differences from unphysical reference states: limits on size, flexibility, and character**. *J Comput Chem* 2003, **24**:1730-1739.

60. • Oostenbrink C, van Gunsteren WF: **Free energies of binding of polychlorinated biphenyls to the estrogen receptor from a single simulation**. *Proteins* 2004, **54**:237-246.
This paper reports the calculation of relative binding affinities of a family of polychlorinated biphenyl compounds to the estrogen receptor from a single FEP simulation with multiple ligands. The judicious choice of an unphysical reference state representative of the 17 different ligands leads to relatively small errors for all but the largest ligands.

61. Hermans J, Shankar S: **The free-energy of xenon binding to myoglobin from molecular-dynamics simulation**. *Isr J Chem* 1986, **27**:225-227.

62. Hermans J, Wang L: **Inclusion of the loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme**. *J Am Chem Soc* 1997, **119**:2707-2714.

63. Gilson MK, Given JA, Bush BL, McCammon JA: **The statistical-thermodynamic basis for computation of binding affinities: a critical review**. *Biophys J* 1997, **72**:1047-1069.

64. McCammon JA: **Theory of biomolecular recognition**. *Curr Opin Struct Biol* 1998, **8**:245-249.

65. Roux B, Nina M, Pomès R, Smith JC: **Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study**. *Biophys J* 1996, **71**:670-681.

66. • Boresch S, Tettinger F, Leitgeb M, Karplus M: **Absolute binding free energies: a quantitative approach for their calculation**. *J Phys Chem B* 2003, **107**:9535-9551.
This work proposes a general approach for imposing biases that restrict the translational and rotational motion of bound ligands in the calculation of absolute binding free energies from alchemical molecular dynamics simulations. A simple method to account for these restraints is shown to address the standard-state dependence of binding equilibria. Test cases include the binding of benzene to the L99A mutant of T4 lysozyme.

67. • Swanson JM, Henchman RH, McCammon JA: **Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculations of the association free energy**. *Biophys J* 2004, **86**:67-74.
This paper examines the conformational free energy of both ligand and receptor molecules upon binding from a long molecular dynamics simulation. This quantity is hard to evaluate because it requires knowledge of changes in the configurational entropy of the system.

68. Li Z, Lazaridis T: **Thermodynamic contributions of the ordered water molecule in HIV-1 protease**. *J Am Chem Soc* 2003, **125**:6636-6637.

69. Petrone PM, García AE: **MHC-peptide binding is assisted by bound water molecules**. *J Mol Biol* 2004, **338**:419-435.

70. Olano LR, Rick SW: **Hydration free energies and entropies for water in protein interiors**. *J Am Chem Soc* 2004, **126**:7991-8000.

71. • Woo HJ, Dinner AR, Roux B: **Grand canonical Monte Carlo simulations of water in protein environments**. *J Chem Phys* 2004, **121**:6392-6400.
The chemical potential of water in protein cavities in equilibrium with a bulk water reservoir is calculated in the grand canonical ensemble and illustrated for several biomolecular systems.

72. • Hamelberg D, McCammon JA: **Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method**. *J Am Chem Soc* 2004, **126**:7683-7689.
The authors present a rigorous method for computing the standard free energies of water molecules in the binding sites of enzyme–ligand complexes. Implications for drug design are discussed.

73. Schafer H, Smith LJ, Mark AE, van Gunsteren WF: **Entropy calculations on the molten globule state of a protein: side-chain entropies of alpha-lactalbumin**. *Proteins* 2002, **46**:215-224.

74. Gohlke H, Case DA: **Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf**. *J Comput Chem* 2004, **25**:238-250.

75. • Chen W, Chang CE, Gilson MK: **Calculation of cyclodextrin binding affinities: Energy, entropy, and implications for drug design**. *Biophys J* 2004, **87**:3035-3049.
This paper shows that changes in host–guest conformational entropy oppose binding by as much as 24 kcal/mol in calculations of cyclodextrin binding affinities for a range of ligands. Interestingly, these changes result from a narrowing of energy wells upon binding rather than from a decrease in the number of conformers.

76. • Lu N, Kofke DA, Woolf TB: **Staging is more important than perturbation method for computation of enthalpy and entropy changes in complex systems**. *J Phys Chem B* 2003, **107**:5598-5611.
The authors examine various protocols to compute entropy and enthalpy changes from single-step and multistage FEP simulations of simple model systems. They analyze the source of errors and argue that perturbations should be performed towards the state with lower entropy.

77. • Peter C, Oostenbrink C, van Dorp A, van Gunsteren WF: **Estimating entropies from molecular dynamics simulations**. *J Chem Phys* 2004, **120**:2652-2661.
The authors examine various approaches for entropy calculations from molecular dynamics simulations. They conclude that a promising approach to obtaining entropy in alchemical transformations is to determine the temperature dependence of the free energy difference by TI.