

# Accuracy and Convergence of Free Energy Differences Calculated from Nonequilibrium Switching Processes

MAIK GOETTE, HELMUT GRUBMÜLLER

Theoretical and Computational Biophysics Department, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Received 17 April 2008; Revised 4 June 2008; Accepted 11 June 2008

DOI 10.1002/jcc.21073

Published online 1 August 2008 in Wiley InterScience (www.interscience.wiley.com).

**Abstract:** The molecular-dynamics-based calculation of accurate free energy differences for biomolecular systems is a challenging task. Accordingly, convergence and accuracy of established equilibrium methods has been subject of many studies, often focusing at small test systems. In contrast, the potential of more recently proposed nonequilibrium methods, derived from the Jarzynski and Crooks equalities, has not yet fully been explored. Here, we compare the performance of these methods by calculating free energy differences for test systems at different levels of complexity and varying extent of the involved perturbations. We consider the interconversion of ethane into methanol, the switching of a tryptophane-sidechain in a tripeptide, and the binding of two different ligands to the globular protein snurportin 1. On the basis of our results, we suggest and assess a new nonequilibrium free energy method, Crooks Gaussian Intersection (CGI), which combines the advantages of existing methods. CGI is highly parallelizable and, for the test systems considered here, is shown to outperform the other studied equilibrium and nonequilibrium methods.

© 2008 Wiley Periodicals, Inc. J Comput Chem 30: 447–456, 2009

**Key words:** molecular dynamics; free energy calculations; thermodynamic integration; benchmark; crooks gaussian intersection; gromaes; fep; cgi; simulation; nonequilibrium simulations

## Introduction

Free energy gradients are the driving forces of all biochemical processes; their accurate calculation is therefore essential for a quantitative understanding of biomolecular function. A particularly challenging task is to calculate free energies for the binding of putative drugs, which typically is only estimated by computationally much less expensive scoring functions. Limited accuracy in free energy calculations from atomic-level simulations often results from limited sampling and has been intensively reviewed.<sup>1–4</sup>

Over the years, several approaches to compute free energies have been developed, assessed, and applied, often focusing on smaller systems.<sup>5–14</sup> For larger systems such as proteins, the sampling problem becomes increasingly severe, mainly due to insufficient overlap of the involved thermodynamic phase space densities. This problem also complicates the reliable assessment of accuracy and convergence, simply due to the difficulty of obtaining a reliable reference value. Methods to calculate free energy differences between two states A and B fall into two classes, equilibrium methods and, developed more recently, nonequilibrium methods. The first equilibrium method considered here, free energy perturbation (FEP), was developed by Zwanzig<sup>15</sup> and has been widely used to calculate free energy differences for amino acid substitutions in proteins.<sup>16–18</sup> This method, however, suffers particularly from the above sampling problem. Because of the exponential growth of statistical uncertainty

with decreasing phase space density overlap, this method requires excessive sampling, especially if the size of the perturbation is large.<sup>19–21</sup>

A different technique is Thermodynamic Integration (TI),<sup>22</sup> which rests on the generalized force ( $\delta H/\delta\lambda$ ) with respect to the coupling parameter  $\lambda$  rather than on finite differences. Here  $\lambda$  is either varied continuously (slow-growth), or, similarly to FEP, sampled at discrete values. One consequence of the continuous shift of  $\lambda$  in slow-growth TI (SGTI) is that, strictly speaking, the system is never in equilibrium, as is assumed in the construction of the method. Therefore, accurate results are expected only for small systems or very long simulation times.<sup>23,24</sup>

Discrete TI (DTI) avoids this problem. Here, gradients of the free energy  $F$  are calculated at discrete  $\lambda$  values, and  $\Delta F$  is obtained from numerical integration (summation) of these gradients. In regions of  $\lambda$  where these gradients are large, insufficient sampling and numerical integration may cause problems and, hence, require increased computational effort. This effect is often particularly large for  $\lambda = 0$  and  $\lambda = 1$ , especially for differing atom numbers of the two involved states.<sup>12</sup> In this context, the Bennett Acceptance Ratio<sup>25</sup> typically provides an improved free energy estimate compared with numerical integration.<sup>14</sup>

**Correspondence to:** H. Grubmüller; e-mail: hgrubmu@gwdg.de

Here we will focus on an assessment of the second class of methods, which are based on nonequilibrium work simulations (NEW). Rather than spending computational effort on one long TI trajectory, Hendrix and Jarzynski suggested to carry out many short Fast Growth TI (FGTI) simulations from state A to B and to calculate the free energy difference between these two states via the Jarzynski equality.<sup>26,27</sup> Unfortunately, however, from the resulting set of trajectories, the ones which occur only rarely carry most of the statistical weight. Therefore, the severe sampling problem persists<sup>28,29</sup> and renders the treatment of larger biomolecules often still as computationally demanding, as for FEP.

It has been suggested to alleviate this problem by combining forward (A→B) and reverse (B→A) switching simulations as well as by using Gaussian approximations of the work distribution.<sup>30</sup> Following the Crooks Fluctuation Theorem (CFT),<sup>31</sup> Bennett's acceptance ratio—which was originally proposed and applied for equilibrium Monte-Carlo simulations—is also applicable to NEW calculations. Recently, this approach was combined with a maximum likelihood estimate,<sup>32</sup> and accurate free energy differences were obtained.<sup>33</sup> Here we refer to this method as BAR.

Despite the considerable amount of studies for the different methods described earlier, mostly on smaller and quite diverse systems, it is difficult to judge how these methods perform for larger and more complex systems such as biological macromolecules. We therefore have carried out a comprehensive evaluation of these methods for three test systems at different levels of complexity, namely the interconversion of ethane to methanol, of tryptophane to glycine in a tripeptide and of m<sub>3</sub>GpppG to m<sup>7</sup>GpppG in the globular protein Snurportin 1.

On the basis of these results, we suggest and evaluate an improved nonequilibrium free energy calculation method, which combines the advantages of the methods discussed earlier.

## Methods

Here we derive our new method to calculate free energies from nonequilibrium trajectories. To compare the accuracy of our method with existing methods from both, the equilibrium and the nonequilibrium class, four established methods are considered and used for extensive test simulations. Recently proposed nonequilibrium methods include Jarzynski's work averaging and Bennett's Acceptance Ratio; established equilibrium methods are Slow Growth Thermodynamic Integration (SGTI) and Discrete Thermodynamic Integration (DTI). As these methods have already been described in detail in the literature,<sup>22,25–27,34</sup> we therefore only sketch the respective approaches.

### Exponential Work Averaging Using a Gaussian Approximation (EXP)

Jarzynski has shown<sup>26,35</sup> that the Helmholtz free energy difference  $\Delta F$  can be derived from a series of nonequilibrium work measurements or computations,

$$e^{-\beta\Delta F} = \langle e^{-\beta W_\tau} \rangle, \quad (1)$$

where  $\langle \rangle$  denotes an average over an ensemble of  $N$  trajectories, which were started from an equilibrated canonical ensemble. Here,  $\beta$  is the reciprocal thermal energy, and  $W_\tau$  is the work

$$W_\tau = \int_0^1 \frac{\delta H_\lambda}{\delta \lambda} d\lambda \quad (2)$$

over a switching process of arbitrary length  $\tau$ , which may be very short. In eq. (2),  $\lambda$  is the so called coupling parameter, which switches the system during a simulation of length  $\tau$  from state A to state B (defined by Hamiltonians  $H_A$  and  $H_B$ , respectively), e.g. via  $H_\lambda = (1 - \lambda)H_A + \lambda H_B$ .

For very long switching times the system stays sufficiently close to equilibrium, such that the dissipated work is negligible and  $\Delta F = W$ ; in this case, eq. (2) describes thermodynamic integration.<sup>34,36</sup>

For method EXP, the distribution  $P(W)$  of values for the work obtained from the trajectory ensemble is approximated by a Gaussian function,

$$P_{f,r}(W) \approx \frac{1}{\sigma_{f,r}\sqrt{2\pi}} \exp\left[-\frac{(W - W_{f,r})^2}{2\sigma_{f,r}^2}\right], \quad (3)$$

where  $W_{f,r}$  and  $\sigma_{f,r}$  are the means and the standard deviations of the work distributions, respectively. The index  $f$  denotes the forward ensemble, where  $\lambda = 0 \rightarrow 1$ , and  $r$  the reverse ensemble, where  $\lambda = 1 \rightarrow 0$ , respectively.

As has been shown by Hummer,<sup>30</sup> for this approximation eq. (1) yields

$$\Delta F_f = W_f - \frac{1}{2}\beta\sigma_f^2 \quad (4)$$

and

$$\Delta F_r = -W_r + \frac{1}{2}\beta\sigma_r^2. \quad (5)$$

With the statistical accuracy of the two above expressions,

$$\bar{\sigma}_{f,r}^2 = \frac{\sigma_{f,r}^2}{N} + \frac{\beta^2\sigma_{f,r}^4}{2(N-1)}, \quad (6)$$

and using standard error propagation theory, we suggest to improve Hummer's estimate by using the weighted mean

$$\Delta F_{\text{EXP}} = \left( \frac{\Delta F_f}{\bar{\sigma}_f^2} + \frac{\Delta F_r}{\bar{\sigma}_r^2} \right) \bigg/ \left( \frac{1}{\bar{\sigma}_f^2} + \frac{1}{\bar{\sigma}_r^2} \right) \quad (7)$$

with statistical accuracy

$$\bar{\sigma}_{\text{EXP}}^2 = \left( \frac{1}{\bar{\sigma}_f^2} + \frac{1}{\bar{\sigma}_r^2} \right)^{-1}. \quad (8)$$

The latter two expressions have been used as method EXP in our simulations. We note, that also Hummer<sup>30</sup> has derived an expression for  $\sigma_f \neq \sigma_r$ ; however no error estimate has been derived for this case. Further, in our hands no improvement was observed for this method (data not shown).

#### Bennett Acceptance Ratio with a Maximum Likelihood Estimator (BAR)

Rather than relying on Jarzynski's equality, several methods are based on the more general Crooks Fluctuation Theorem (CFT).<sup>31</sup> Accordingly, the forward and reverse work distributions obey

$$\frac{P_f(W)}{P_r(-W)} = e^{\beta(W-\Delta F)}. \quad (9)$$

Using eq. (9), and assuming sufficiently smooth *a priori* distributions, a maximum likelihood approach<sup>32</sup> on Bennett's Acceptance Ratio<sup>25</sup> yields

$$\left\langle \frac{1}{1 + \exp[\beta(W - \Delta F_{\text{BAR}})]} \right\rangle_f = \left\langle \frac{1}{1 + \exp[-\beta(W - \Delta F_{\text{BAR}})]} \right\rangle_r, \quad (10)$$

assuming an equal number of forward (*f*) and reverse (*r*) simulations. From eq. (10)  $\Delta F_{\text{BAR}}$  is calculated numerically.

Following Shirts et al.<sup>32</sup> and Anderson,<sup>37</sup> we estimated the statistical error of  $\Delta F_{\text{BAR}}$  by

$$\frac{2}{\beta^2 N} \left[ \left\langle \frac{1}{1 + \cosh[\beta(W - \Delta F_{\text{BAR}})]} \right\rangle^{-1} - 2 \right], \quad (11)$$

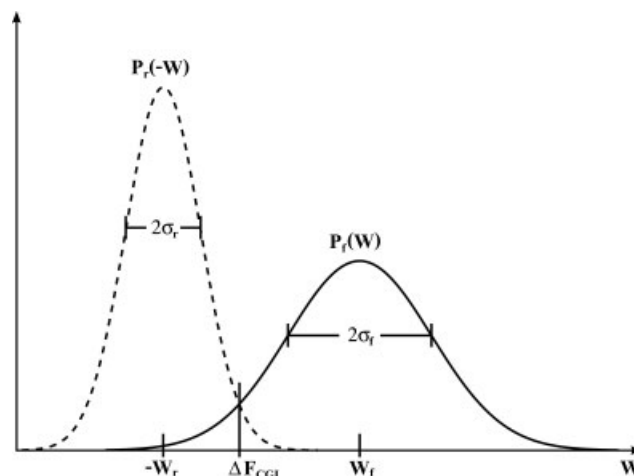
where  $\langle \rangle$  denotes the average over *all* (i.e., forward *and* reverse) work calculations. We will refer to this method as BAR in our work.

#### Crooks Gaussian Intersection (CGI)

An alternative set of methods employs CFT [eq. (9)] more directly. Again, for each trajectory the mechanical work for the switching process is calculated via eq. (2). As can be seen from eq. (9),  $\Delta F$  is the work  $W$  for which  $P_f(W) = P_r(-W)$ , i.e., the intersection point of the two work distributions (Fig. 1).<sup>29</sup>

Direct determination of the intersection point (e.g., from histograms) is prone to large statistical errors, however, because only those work values are used which fall into the bin containing the intersection point. Particularly if the forward and reverse distributions exhibit only a small overlap, this number can be very small or even zero.

In analogy to method EXP, we here propose to use Gaussian approximations. This approach has the advantage that the assumption of a Gaussian work distribution has been shown to hold in the limit of large numbers of degrees of freedom.<sup>38</sup> Furthermore, as we will show below, this assumption can be tested through a Kolmogorov–Smirnov test.<sup>39</sup>



**Figure 1.** Schematic Gaussian work distributions for the switching from state A to B for a forward [ $P_f(W)$ , solid line] and a reverse [ $P_r(-W)$ , dashed line] process.  $W_f$ ,  $\sigma_f$ ,  $-W_r$ , and  $\sigma_r$  are the means and standard deviations of  $P_f(W)$  and  $P_r(-W)$ , respectively. These values are used to calculate the free energy difference  $\Delta F_{\text{CGI}}$ .

The intersection point is given by

$$\Delta F_{\text{CGI}} = \frac{\frac{W_f}{\sigma_f^2} - \frac{-W_r}{\sigma_r^2} \pm \sqrt{\frac{1}{\sigma_f^2 \sigma_r^2} (W_f + W_r)^2 + 2 \left( \frac{1}{\sigma_f^2} - \frac{1}{\sigma_r^2} \right) \ln \frac{\sigma_r}{\sigma_f}}}{\frac{1}{\sigma_f^2} - \frac{1}{\sigma_r^2}}, \quad (12)$$

where  $W_f$  and  $-W_r$  are the means, and  $\sigma_f$  and  $\sigma_r$  are the standard deviations of the respective Gaussian functions, as also defined in Figure 1.

We note that for  $\sigma_f \neq \sigma_r$ , two intersection points will generally be found, whereas  $\Delta F$  is of course uniquely defined. Only for  $\sigma_f^2 = \sigma_r^2$  a unique solution,  $\Delta F_{\text{CGI}} = (W_f - W_r)/2$ ,<sup>40</sup> is obtained. Typically, one intersection point is located between  $W_f$  and  $-W_r$ , and is close to  $(W_f - W_r)/2$ . The second intersection point is located far in the tail region and in this case is an artifact of our Gaussian approximation, which often fails in the tail region.<sup>38</sup> Accordingly, we chose the intersection value which is closest to  $(W_f - W_r)/2$  as our estimate for  $\Delta F_{\text{CGI}}$ .

If both Gaussians were too close to compute a proper intersection point, i.e., if  $W_f$  and  $-W_r$  were closer to each other than to one of the two intersection values, we empirically chose the mean of both as the best estimate for  $\Delta F_{\text{CGI}}$ . To estimate the statistical accuracy of  $\Delta F_{\text{CGI}}$ , we resorted to a Monte Carlo approach. Accordingly, 10,000 synthetic sets of work values were generated, each set containing  $N$  values that were randomly chosen from a Gaussian distribution with mean  $W_f$  and standard deviation  $\sigma_f$  for the forward simulation, and, similarly,  $N$  values for the reverse calculations.  $W_f$ ,  $-W_r$ , and  $\sigma_{f,r}$ , were taken from the above Gaussian fits to the work distribution obtained from the simulations. For each of the 10,000 synthetic forward/reverse work distributions, intersection points were calculated from the respective means and

standard deviations. The standard deviation of this ensemble of synthetic intersection points was used as an estimate for the expected statistical error of  $\Delta F_{\text{CGI}}$ .

### Slow Growth TI (SGTI) and Discrete TI (DTI)

To compare convergence behavior and accuracy of the above NEW methods with each other as well as with established equilibrium methods, we chose two widely-used approaches, Slow Growth TI (SGTI) and Discrete TI (DTI). Similar to FGTI, SGTI is also based on eq. (2). In contrast to the NEW methods, where many TI calculations are used to compute a statistical meaningful free energy difference, only one SGTI trajectory with a much longer  $\tau$  than in FGTI is computed and evaluated. Since in this case only one forward and one reverse trajectory is available, statistic convergence can be—and usually is—assessed only via the hysteresis of the forward and reverse free energy values. Here we provide this value as an indication of the expected statistical accuracy, which, however, is often likely to be larger.

The second equilibrium method considered here is DTI, which is currently becoming increasingly popular. In contrast to SGTI, the integral in eq. (2) is evaluated via discrete summation over ensemble averages of  $\delta H/\delta\lambda$ ,

$$\Delta F_{\text{DTI}} = \int_0^1 \left\langle \frac{\delta H}{\delta\lambda} \right\rangle_{\lambda} d\lambda \approx \sum_{i=0}^N \left\langle \frac{\delta H}{\delta\lambda} \right\rangle_{\lambda_i} \Delta\lambda, \quad (13)$$

where each ensemble average is obtained from one equilibration trajectory. The discrete summation here was carried out via Simpson's rule.<sup>41</sup> The statistical accuracy of  $\Delta F_{\text{DTI}}$  was computed via a Gaussian error propagation of the error contributions from each trajectory, calculated as usual by block averaging.<sup>42</sup>

### Test Systems

To evaluate the accuracy and convergence of the described CGI method as well as the conventional ones, we considered two small test systems and a large one (see Table 1). The first, "E2M", involves the interconversion of one solvated ethane into a methanol molecule, which affects essentially all eight atoms (Fig. 2, E2M). The second, "W2G", is a tripeptide, Gly-Trp-Gly, which is "mutated" into Gly-Gly-Gly. Here, 19 atoms are involved, where 17 were transformed into dummy particles which keep their mass and bond parameters, but do not interact via Lennard Jones or via Coulomb interactions (Fig. 2, W2G). For both systems the OPLSAA force field was used.<sup>43</sup>

As a third, large test system of biological relevance we chose the protein snurportin 1 (SPN) (PDB entry 1XK5<sup>44</sup>). The ligand from

the SPN crystal structure, m<sub>3</sub>GpppG, was transformed to a modified form, m<sup>7</sup>GpppG. Here, ten atoms are affected by the perturbation, and six become dummy particles (Fig. 2, SPN). For this system the AMBER99 force field<sup>45,46</sup> was used. For both SPN ligands, m<sub>3</sub>GpppG (resolved in the crystal structure) and m<sup>7</sup>GpppG (not resolved in the crystal structure), no force field parameters were available. We therefore used the standard AMBER99 values for guanosine and ribose. The m<sub>3</sub>G- and m<sup>7</sup>G-nucleoside parameters were taken from ref. 47, and the parameters for triphosphate, connecting the two nucleotides, from ref. 48. Additionally, we scaled the charges of the molecule at the connecting phosphates such that a net charge of -2 was obtained.

### Simulation Details

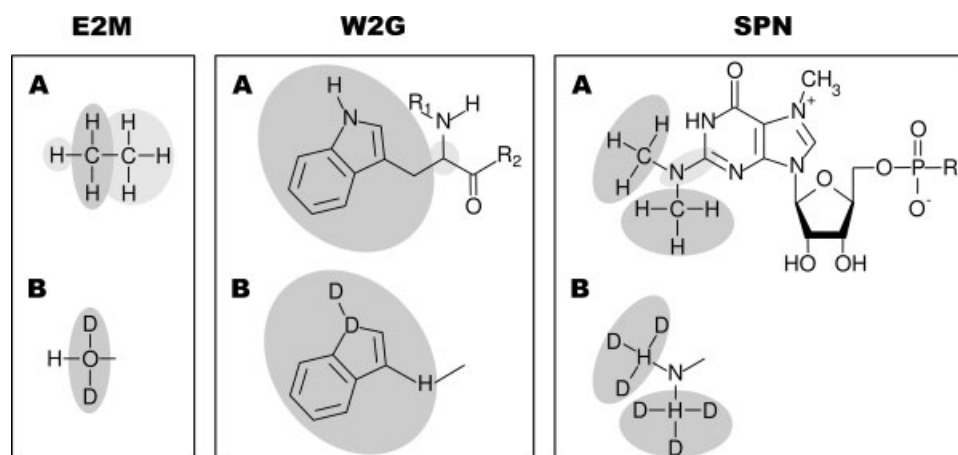
All simulations were carried out in explicit solvent, except one SGTI simulation of E2M, which was additionally carried out in vacuum to compute solvation free energies. All nonvacuum simulations were treated with a cutoff of 1.0 nm for short-range Coulomb as well as for Lennard-Jones interactions, and with Particle-Mesh-Ewald with a grid spacing of 0.12 nm and an interpolation order of 4 for long range electrostatics.<sup>49</sup> All simulations were carried out in an  $NpT$  ensemble using Berendsen pressure and temperature coupling<sup>50</sup> at 1 bar with a pressure coupling coefficient of  $\tau_p=1$  ps and at 300 K with a temperature coupling coefficient of  $\tau_T=0.1$  ps. The TIP4P water model<sup>51</sup> was used and, in the case of snurportin 1, a 150 mmol NaCl salt-concentration was used to mimic physiological solution. All simulations were carried out with the GROMACS software-package (version 3).<sup>52</sup> The E2M and W2G systems were equilibrated for one nanosecond; the SPN system was equilibrated for 50 ns. All test simulations were started from these equilibrated systems.

For the free energy calculations, we used a hardcore and, where necessary, a softcore potential. A softcore potential was used for perturbations involving changes of the number of atoms with van der Waals interactions. In addition, whenever required for proper comparison, a softcore potential was used also in those cases, where a hardcore potential would otherwise suffice according to the above criterion. For the softcore potential a 3-step scheme was used to avoid Coulomb interaction singularities for overlapping atoms. Accordingly, the switching process from state A to state B was split into three steps, involving two intermediate states. In the first step, the charges of all perturbed atoms of state A were switched to zero. In the second step, we switched the Lennard-Jones parameters, masses, as well as all bond, angle, and dihedral parameters of the perturbed atoms. Masses of dummy atoms and the parameters of bonds involving dummy atoms were not changed. A softcore potential<sup>53</sup> with  $\alpha = 0.25$  and  $\sigma = 0.3$  was used, with the two parameters defined as in ref. 52. In the last step, the charges of the perturbed non-dummy

**Table 1.** Summary of Simulation System Details of the Test-Systems Ethane (E2M), Tripeptide (W2G), and Snurportin 1 (SPN).

System	A-state	B-state	Force field	Water model	Ion conc. (mmol/L)	No. of atoms
E2M	C <sub>2</sub> H <sub>6</sub>	CH <sub>3</sub> OH	OPLSAA	TIP4P	–	1,432
W2G	GWG	GGG	OPLSAA	TIP4P	–	1,389
SPN	m <sub>3</sub> GpppG	m <sup>7</sup> GpppG	AMBER99	TIP4P	150	63,642

The tripeptide in the A- and B-state is given in the one-letter-code for amino acids.



**Figure 2.** Structures of the tested systems. (A) A-state of the free energy calculations for E2M, W2G, and SPN. (B) Respective B-state. The dark gray ellipsoids denote the atoms, which contribute most to the perturbation. The light gray ellipsoids indicate regions of minor perturbations.  $R_x$  denotes the rest of the respective molecule, D the dummy particles. In W2G B the complete imidazole ring is composed of dummy particles.

particles were switched from zero to their state B value. For the hardcore potential, the conventional 1-step scheme was used. The simulation length of all these steps was always chosen to be equal, and Gaussian error propagation was used to estimate the total error of  $\Delta F$ .

#### Comparison to Experiment

As a control, solvation free energies were calculated for ethane and methanol, respectively, and compared to experiments. To this aim, nine SGTI simulations were carried out for system E2M in solvent and one *in vacuo*. All SGTI simulations were performed in a 1-step process over 10 ns in the forward and reverse direction, yielding a total simulation time of 200 ns.

#### Test of the Gaussian Approximation

To test the assumption that the obtained distributions of nonequilibrium work values can be approximated by Gaussian functions, 1000 forward and 1000 reverse FGTI simulations of  $\tau = 50$  ps each were carried out for system E2M, using a 1-step switching process. After equilibrating the two states A and B for 11 ns, 1000 starting snapshots were extracted from the last 10 ns of each of the two trajectories. To the obtained set of work values, a Kolmogorov–Smirnov (KS-) test<sup>39</sup> was applied.

#### Convergence of Trajectories: Varying Number and Lengths

For the two small systems, E2M and W2G, extended test simulations with comparable total simulation times were carried out. FGTI was used to compute trajectories for the NEW methods, and SGTI for comparison. To assess the effect of the number of used trajectories, the number of trajectories for the nonequilibrium simulations was varied and a constant switching time  $\tau = 50$  ps was used. 12, 25, 50, and 150 trajectories were used, yielding a total simulation time of 3.6, 7.5, 15, and 45 ns, respectively, for both the forward and reverse simulations. The SGTI simulations switching times  $\tau$  were

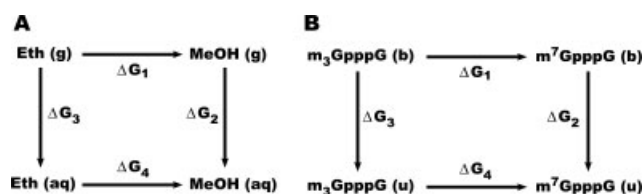
chosen to be 3.84, 7.5, 30, 37.5, 48, 60, and 75 ns, respectively, for the forward and reverse path. To allow proper comparison, the 3-step scheme was used for all simulations in this paragraph.

The effect of trajectory length (for a fixed number of trajectories) was assessed for all three systems, E2M, W2G, and SPN. The 3-step scheme was used for systems E2M and W2G; the 1-step scheme was used for system SPN. Switching times  $\tau$  of 1, 5, 10, 25, 50, 80, 100, 128, 160, and 200 ps were used for all FGTI calculations. For each value of the switching time, 50 trajectories were calculated. The total simulation times for E2M and W2G were thus 0.3, 1.5, 3, 7.5, 15, 24, 30, 38.4, 48, and 60 ns, respectively.

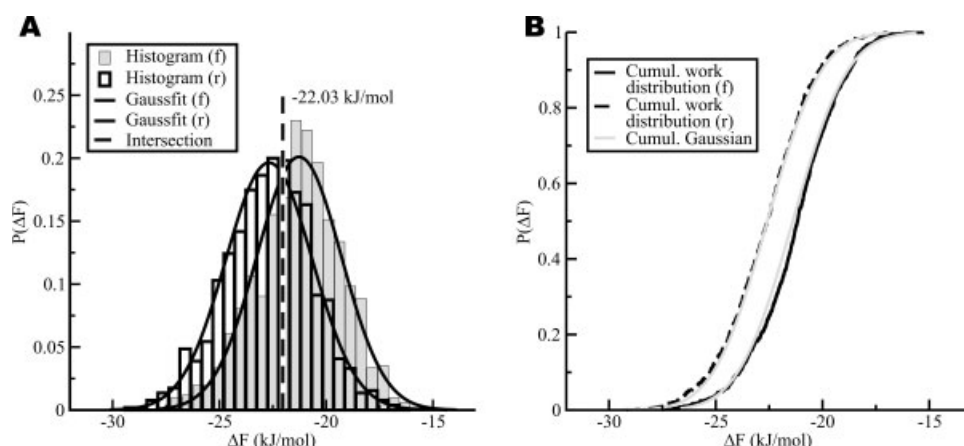
For SPN, we additionally computed the interconversion of the two ligands  $m_3$ GpppG and  $m^7$ GpppG, bound to snurportin 1 and in solution to calculate their binding free energy difference. The total simulation times for these simulations were 0.6, 1, 2, 5, 10, 16, 20, 25.6, 32, and 40 ns.

#### Accuracy for Given Computational Effort

To evaluate the accuracy of all five methods SGTI, DTI, EXP, BAR, and CGI for given computational effort, the test systems E2M and W2G were used. A total simulation time of roughly 20 ns was spent for each method. The 3-step scheme was used for these simulations.



**Figure 3.** Used thermodynamic cycles. (A) switching from ethane (Eth) to methanol (MeOH) in vacuum (g) and aqueous solution (aq), (B) switching from  $m_3$ GpppG to  $m^7$ GpppG bound to SPN (b) and unbound (u) in solution.



**Figure 4.** (A) Distribution of 1000 work measurements for the switching process of ethane to methanol in aqueous solution, forward (f) and reverse (r). A Gaussian function was fitted to the forward and reverse data, with their intersection yielding the free energy. (B) Distribution functions for the same data, which are assessed by the Kolmogorov–Smirnov-test. For comparison, the respective cumulative distribution function (error function) of the respective Gaussian functions is shown in gray.

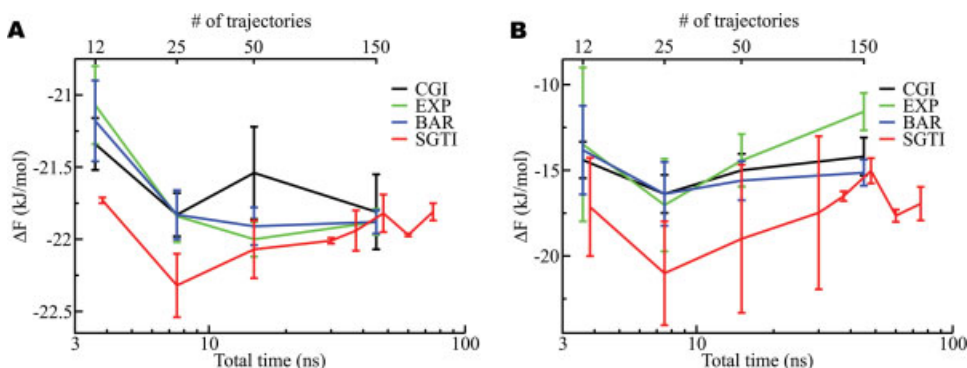
The SGTI simulation times for both systems were 9.6 ns each for the forward and reverse process, yielding a total simulation time of 19.2 ns. For the DTI simulations, the switching process was split into 11 equally spaced discrete  $\lambda$  values. For each of these values, the system was sampled for 600 ps, yielding a total simulation time of 18.6 ns. The first 200 ps of each simulation were discarded as additional equilibration time and were not used for the calculation of the intermediate free energies  $F_\lambda$  and their statistical accuracies.

For EXP, BAR, and CGI, identical trajectories were used as input. For the 1-step process (system SPN), the A- and B-states were equilibrated for 1 ns. For the 3-step process (systems E2M and W2G), the intermediate states were also equilibrated for 1 ns. From the last 100 ps of each equilibration trajectory 50 snapshots were extracted as starting structures for the subsequent 50 ps FGTI simulations. Statistical independence was assessed via an autocorrelation analysis of  $dH/d\lambda$  which yielded an autocorrelation time well below 100 ps/50 = 2 ps. The total simulation time including the necessary equilibration runs was 21 ns.

## Results

### Comparison to Experiment

As a check, we compared computed free energies with experimental solvation free energies for E2M. The appropriate thermodynamic cycle (Fig. 3A) required interconversions of ethane into methanol both in solvent and in vacuum. From nine SGTI simulations in solvent,  $\Delta G_4 = -21.91 \pm 0.09$  kJ/mol was obtained. Here, the error of the mean was estimated from the standard deviation ( $\sigma = 0.27$  kJ/mol) of the obtained nine free energy values. The observed value for the hysteresis was 0.09 kJ/mol on average, which agrees with this estimate. The vacuum free energy difference was  $\Delta G_1 = 6.72 \pm 0.03$  kJ/mol, with the error estimated from the hysteresis. The resulting solvation free energy difference  $\Delta \Delta G = \Delta G_1 - \Delta G_4 = 28.63 \pm 0.10$  kJ/mol agrees well with the experimental value of  $\Delta \Delta G = 29.01$  kJ/mol<sup>54</sup> as well as with earlier simulations,  $\Delta \Delta G = 28.25 \pm 1.13$  kJ/mol.<sup>5</sup> We note that we consider Gibbs' free energies here, as the experiments refer to constant pressure conditions.



**Figure 5.** Comparison of 3-step free energy calculations. (A) System E2M, (B) System W2G. The total simulation time is shown on a logarithmic time scale. The upper scale shows the corresponding number of FGTI trajectories used (unidirectional). For each SGTI free energy calculation, one trajectory was used.

Because only double differences are considered, the  $p\Delta V$  term is not expected to affect our results.

#### Test of the Gaussian Approximation

Methods EXP and CGI rely on the assumption that the distribution of work values from nonequilibrium switching processes can be approximated by Gaussian functions. To test this assumption, 1000 independent simulations for the system E2M were performed, from which 1000 work values were obtained. Figure 4A shows the obtained distributions as well as the Gaussian fits.

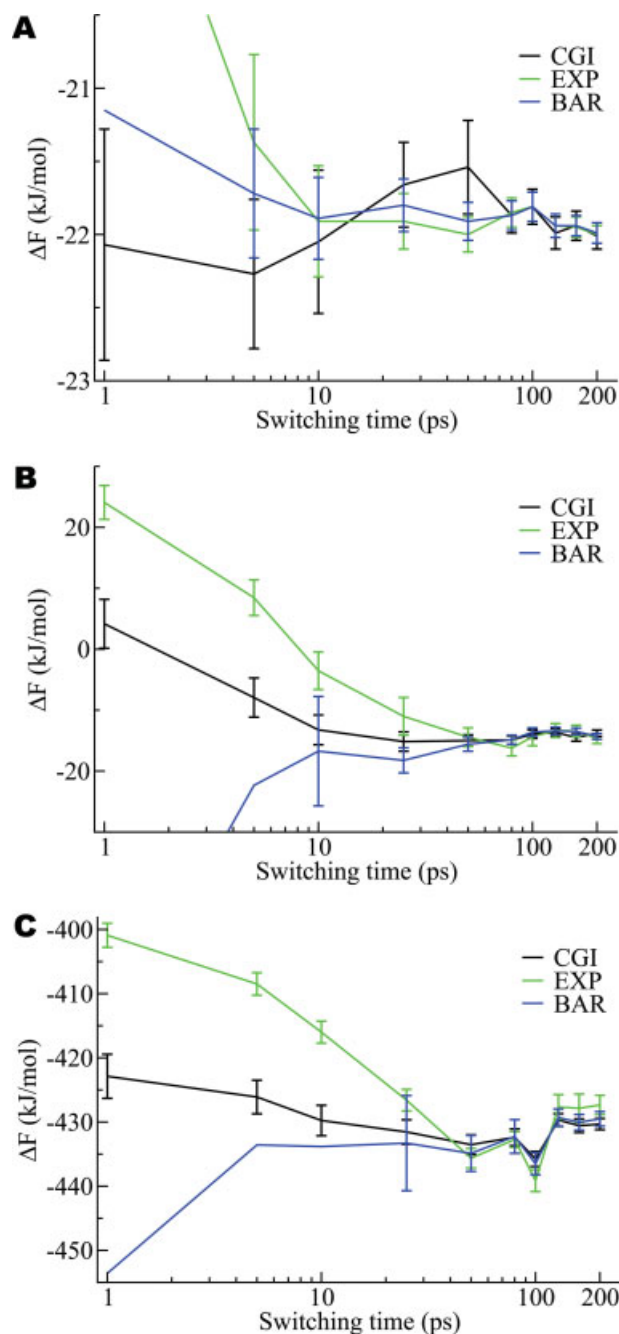
To test the hypothesis that these 1000 values are actually distributed according to a Gaussian function, the Kolmogorov–Smirnov-test was applied (Fig. 4B). The obtained significance levels of  $\alpha_f=0.10$  and  $\alpha_r=0.50$  for the forward and, respectively, reverse work distributions imply that the hypothesis cannot be rejected at the usual 5% (or lower) significance level. We thus assume that the Gaussian approximation holds for the case at hand as well as for the other test systems considered here. We note that the limited set of work values lacks data points in the far tail regions of the distributions, which therefore cannot be reliably assessed by the KS-test. Because the intersection of the forward and reverse distribution falls typically within the “main body” of the Gaussian function and not in the tails, however, this uncertainty will not affect our free energy estimates.

#### Convergence for Varying Numbers of Trajectories

To assess the accuracy and convergence behavior of the three NEW methods, test simulations were carried out for the two test systems E2M and W2G. SGTI was used as a reference. Figure 5 shows the convergence of  $\Delta F$  for the four methods as a function of the total computational effort spent. For SGTI, convergence is reached for both systems after about 40 ns. The mean and statistical accuracy of the last four energy values are  $\Delta F = -21.89 \pm 0.20$  kJ/mol for E2M and  $\Delta F = -16.54 \pm 1.31$  kJ/mol for W2G, respectively. We used these two mean values as our best reference estimates, against which convergence of the NEW methods is assessed.

As can be seen, beyond 7.5 ns, all NEW methods, except EXP for system W2G, yield free energy estimates, which agree with the reference within their respective statistical accuracy. Accordingly, for the systems at hand we consider 25 trajectories for the forward process and 25 for the reverse, respectively, sufficient to obtain reasonably converged results for the NEW methods.

Next, we studied the effect of trajectory length for a given number of 50 trajectories for systems E2M, W2G, and SPN. As can be seen in Figure 6, all NEW methods converge for all three systems at about 80 ps trajectory length. However, for the large system SPN, as well as for the large perturbation in system W2G, CGI turns out to converge markedly faster, such that accurate results are obtained already from quite short trajectories. Closer analysis reveals that CGI provides the same accuracy as EXP or BAR at about half of the computational effort. Furthermore, CGI provides reliable error estimates, whereas EXP drastically underestimates the statistical error.



**Figure 6.** Convergence of NEW free energies for differing FGTI trajectory lengths on a logarithmic scale. (A) system E2M, (B) system W2G, (C) system SPN. Trajectory lengths vary from 1 to 200 ps. The errors of the BAR method for the very short trajectories (1–25 ps) are often very large and are therefore not shown in the plot.

#### Accuracy for Given Computational Effort

We compared the accuracy of both equilibrium (SGTI, DTI) and nonequilibrium methods (EXP, BAR, CGI) for given computational effort. For systems E2M and W2G, a total simulation time of roughly

**Table 2.** Free Energy Contributions of a 3-Step Switching Process of System E2M.

Method	$\Delta G_{\text{QQoff}}$	$\sigma$	$\Delta G_{\text{VdW}}$	$\sigma$	$\Delta G_{\text{QQon}}$	$\sigma$	$\Delta G_{\text{Tot}}$	$\sigma$
SGTI	-8.44	0.01	-0.77	0.09	-12.86	0.17	-22.07	0.19
DTI	-8.43	0.05	-0.70	0.49	-12.68	1.51	-21.81	1.59
EXP	-8.43	0.01	-0.77	0.05	-12.80	0.12	-22.00	0.13
BAR	-8.43	0.001	-0.78	0.002	-12.70	0.01	-21.91	0.01
CGI	-8.43	0.01	-0.78	0.24	-12.33	0.20	-21.54	0.31

“QQ<sub>off</sub>” denotes switching the charges of the perturbed atoms in the state A to zero, “VdW” the interconversion of the Lennard Jones parameters from state A to B with softcore potentials, “QQ<sub>on</sub>” the switching of the charges to their values in state B, and “Tot” the sum of these three contributions, yielding the total free energy difference. All free energies are given in kJ/mol. The estimated error for all values is  $\sigma$ . For slow growth TI,  $\sigma$  provides a lower bound error estimate.

20 ns was spent as described in the methods section. Table 2 shows the free energy contributions of the 3 steps as well as the total free energy difference obtained from the 3-step switching processes for system E2M. Table 3 shows the same results for system W2G. As can be seen from Table 2, for system E2M all applied methods yield free energy differences which converge to  $\Delta G_{\text{Tot}} \approx -21.9$  kJ/mol. For a larger perturbation (system W2G), however, convergence to  $\Delta G_{\text{Tot}} \approx -15.2$  kJ/mol with reasonable statistical accuracy was only reached by the NEW methods (Table 3). In addition to the DTI result using 11 discrete  $\lambda$  values listed in Table 3, DTI was also used with 21 discrete  $\lambda$  values, sampling each for 300 ps. However, no significant differences in the resulting free energies were seen (data not shown).

We finally applied the NEW techniques to SPN using the thermodynamic cycle shown in Figure 3B. As can be seen from Table 4, the free energy differences obtained for all three NEW methods are sufficiently accurate to allow a reliable calculation of the required double differences  $\Delta\Delta G$ . Despite the fact that quite small differences of large values are involved, CGI provides reliable values already for much shorter trajectory lengths than EXP and BAR. Comparison with the measured value of  $5.61 - 11.22$  kJ/mol<sup>44,55</sup> suggests that, given the relatively large experimental uncertainty, good agreement is obtained for both, CGI and BAR.

## Discussion and Conclusions

We compared established equilibrium (SGTI, DTI) and nonequilibrium (EXP, BAR) free energy calculation methods. On the basis of our results, an improved nonequilibrium method, CGI, has been derived and tested, which combines the advantages of existing methods. Interconversions between ethane and methanol (E2M),

between tryptophane and glycine within a tripeptide (W2G), and between  $m_3\text{GpppG}$  into  $m^7\text{GpppG}$  bound to the globular protein snurportin 1 (SPN) were used as test cases. The main aim here was to assess recently proposed nonequilibrium methods, particularly for large perturbations, which typically occur in the context of complex biomolecular systems.

For system E2M, the calculated free energy difference of  $28.63 \pm 0.1$  kJ/mol agrees well with the measured value of 29.01 kJ/mol, which suggests that the used force field and simulation times provide sufficiently accurate results for our further assessments.

To assess their convergence behavior, the three NEW methods as well as SGTI were compared by calculating the free energy of the two systems E2M and W2G for different simulation times. Extended SGTI simulations were found to converge after about 40 ns and were therefore taken as reference. All NEW methods, except EXP for system W2G, yielded converged and, compared to SGTI, accurate free energies (within the error bars) already above 7.5 ns, i.e., for five times shorter simulation times.

The effect of trajectory length on the convergence behavior of the NEW methods was then studied in more detail by varying the trajectory length of the FGTI simulations performed for the systems E2M, W2G, and SPN. Using ensembles of 50 trajectories, all NEW methods converged for all studied systems for trajectory lengths above 80 ps. Remarkably, our new method CGI converged significantly faster, particularly for the more complex perturbations.

Both equilibrium (SGTI, DTI) and nonequilibrium methods (EXP, BAR, CGI) were then compared for given computational effort. To compute the free energy differences for the systems E2M and W2G, a total simulation time of roughly 20 ns was spent. For the very small system E2M, all five methods turned out to perform reasonably well. For such a small system, convergence after 20 ns is indeed expected. Importantly, for the larger perturbation of

**Table 3.** Free Energy Contributions of a 3-Step Switching Process of System W2G.

Method	$\Delta G_{\text{QQoff}}$	$\sigma$	$\Delta G_{\text{VdW}}$	$\sigma$	$\Delta G_{\text{QQon}}$	$\sigma$	$\Delta G_{\text{Tot}}$	$\sigma$
SGTI	26.61	0.3	-52.00	4.17	6.39	1.13	-19.00	4.33
DTI	26.79	2.94	-48.42	9.46	7.68	1.66	-13.95	10.04
EXP	25.55	0.21	-46.84	1.52	6.86	0.05	-14.43	1.54
BAR	25.54	0.03	-48.38	1.24	7.24	0.02	-15.60	1.24
CGI	25.53	0.20	-47.63	0.92	7.24	0.16	-14.98	0.96

Symbols are defined as in Table 2.



**Table 4.** Free Energy Contributions of a 3-Step Switching Process of SPN.

Method	$\Delta G_{\text{Cpx}}$	$\sigma$	$\Delta G_{\text{Sol}}$	$\sigma$	$\Delta \Delta G$	$\sigma$
EXP	-427.35	1.51	-441.18	0.67	13.83	1.65
BAR	-429.52	1.29	-440.91	0.17	11.39	1.30
CGI	-430.34	0.86	-441.16	0.51	10.82	1.00

“Cpx” denotes the free energy difference calculated for the ligands bound to the protein, and “Sol” for the ligands in solvent, respectively. All free energies are given in kJ/mol. The estimated error for all values is  $\sigma$ .

system W2G, the NEW methods provide significantly more accurate results than the equilibrium methods. Among the equilibrium methods, DTI was found to provide the most accurate free energy values; however, the very large statistical uncertainty involved is clearly a disadvantage.

As our new method CGI (as well as EXP) relies on the validity of the Gaussian approximation of work distributions, this assumption deserved particular attention. To test this assumption, Kolmogorov–Smirnov-tests were applied to the work distributions calculated for the systems mentioned above. 1000 forward and 1000 reverse work values were obtained from a corresponding number of FGTI trajectories, and their distributions were subjected to a Kolmogorov–Smirnov-test. Indeed, the hypothesis of an underlying Gaussian distribution could not be rejected, such that the Gaussian approximation for work distributions seems to be sufficiently accurate, which is in line with the accurate values obtained by CGI.

We attribute the improved accuracy of CGI over EXP to the fact that CGI is derived directly from the more general Crooks’ theorem rather than from Jarzynski’s equality, in particular, CGI exploits the relation between forward and reverse distributions, which is not implied in Jarzynski’s equality.

As a check against experiment involving a large system, we finally calculated binding free energies for two ligands,  $m_3$ GpppG and  $m^7$ GpppG, bound to the globular protein snurportin 1 (SPN). The obtained values of  $11.39 \pm 1.3$  kJ/mol (BAR) and  $10.82 \pm 1.0$  kJ/mol (CGI), respectively, agree with the estimate from experiments of 6 – 11 kJ/mol, albeit such large experimental error does not allow to attribute a pronounced significance to this result. Nevertheless, the value of  $13.83 \pm 1.65$  kJ/mol obtained from method EXP, is clearly too large.

Overall, for the three test systems considered, the nonequilibrium methods were shown to outperform the traditional equilibrium methods. The best results were obtained for our newly proposed CGI method. We attribute the substantially faster convergence of the nonequilibrium methods to the fact that these do not rely on the assumption that the system is sufficiently close to equilibrium at all times. Rather, only the ensembles from which the required sets of trajectories are started need to be sufficiently equilibrated.

A further advantage of the nonequilibrium methods is that they are inherently parallel, and therefore scale optimally on parallel machines, in contrast to calculations of single long trajectories.

## Acknowledgments

The authors thank Guillem Portella and Gerrit Groenhof for helpful discussions. Help by David Mobley is also gratefully acknowledged.

## References

- van Gunsteren, W. *Protein Eng Des Select* 1988, 2, 5.
- Kollman, P. *Chem Rev* 1993, 93, 2395.
- Rodinger, T.; Pomes, R. *Curr Opin Struct Biol*, 2005, 15, 164.
- Shirts, M.; Mobley, D.; Chodera, J. *Ann Rep Comp Chem* 2007, 3, 41.
- Jorgensen, W.; Ravimohan, C. *J Chem Phys* 1985, 83, 3050.
- van Gunsteren, W.; Daura, X.; Mark, A. *Helvetica Chim Acta* 2002, 85, 3113.
- Hu, H.; Yun, R.; Hermans, J. *Mol Simul* 2002, 28, 67.
- Lu, N.; Kofke, D.; Woolf, T. *J Comput Chem* 2004, 25, 28.
- Woo, H.; Roux, B. *Proc Natl Acad Sci* 2005, 102, 6825.
- Oostenbrink, C.; van Gunsteren, W. *Chemistry(Weinheim)* 2005, 11, 4340.
- Oostenbrink, C.; van Gunsteren, W. *Chem Phys* 2006, 323, 102.
- Shirts, M.; Pande, V. *J Chem Phys* 2005, 122, 144107.
- Kofke, D. *Mol Phys* 2006, 104, 3701.
- Mobley, D.; Chodera, J.; Dill, K. *J Chem Phys* 2006, 125, 084902.
- Zwanzig, R. *J Chem Phys* 1954, 22, 1420.
- Warshel, A.; Sussman, F. *Proc Natl Acad Sci USA* 1986, 83, 3806.
- Hwang, J.; Warshel, A. *Biochemistry* 1987, 26, 2669.
- Rao, S.; Singh, U.; Bash, P.; Kollman, P. *Nature* 1987, 328, 551.
- Lu, N.; Kofke, D. *J Chem Phys* 2001, 114, 7303.
- Lu, N.; Kofke, D. *J Chem Phys* 2001, 115, 6866.
- Lu, N.; Singh, J.; Kofke, D. *J Chem Phys* 2003, 118, 2977.
- Kirkwood, J. *J Chem Phys* 1935, 3, 300.
- Pearlman, D.; Kollman, P. *J Chem Phys* 1989, 91, 7831.
- Straatsma, T.; McCammon, J. *J Chem Phys* 1991, 95, 1175.
- Bennett, C. *J Comput Phys* 1976, 22, 245.
- Jarzynski, C. *Phys Rev Lett* 1997, 78, 2690.
- Hendrix, D.; Jarzynski, C. *J Chem Phys* 2001, 114, 5974.
- Lua, R.; Grosberg, A. *J Phys Chem B* 2005, 109, 6805.
- Jarzynski, C. *Phys Rev E* 2006, 73, 46105.
- Hummer, G. *J Chem Phys* 2001, 114, 7330.
- Crooks, G. *J Stat Phys* 1998, 90, 1481.
- Shirts, M.; Bair, E.; Hooker, G.; Pande, V. *Phys Rev Lett* 2003, 91, 140601.
- Shirts, M.; Pande, V. *J Chem Phys* 2005, 122, 134508.
- Beveridge, D.; DiCapua, F. *Annu Rev Biophys Biophys Chem* 1989, 18, 431.
- Jarzynski, C. *Phys Rev E* 1997, 56, 5018.
- Straatsma, T.; McCammon, J. *Ann Rev Phys Chem* 1992, 43, 407.
- Anderson, J. *Biometrika* 1972, 59, 19.
- Crooks, G.; Jarzynski, C. *Phys Rev E* 2007, 75, 21116.
- Massey, F. Jr, *J Am Stat Assoc* 1951, 46, 68.
- Kosztin, I.; Barz, B.; Janosi, L. *J Chem Phys* 2006, 124, 064106.
- Atkinson, K. *An Introduction to Numerical Analysis*; Wiley: New York, 1978.
- Hess, B. *J Chem Phys* 2002, 116, 209.
- Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. *Am Chem Soc* 1996, 118, 11225.

44. Strasser, A.; Dickmanns, A.; Luhrmann, R.; Ficner, R. *EMBO J* 2005, 24, 2235.
45. Wang, J.; Cieplak, P.; Kollman, P. *J Comput Chem* 2000, 21, 1049.
46. Sorin, E.; Pande, V. *Biophys J* 2005, 88, 2472.
47. Aduri, R.; Psciuk, B.; Saro, P.; Taniga, H.; Schlegel, H.; SantaLucia, J. Jr *J Chem Theory Comput* 2007, 3, 1464.
48. Meagher, K.; Redman, L.; Carlson, H. *J Comput Chem* 2003, 24, 1016.
49. Darden, T.; York, D.; Pedersen, L. *J Chem Phys* 1993, 98, 10089.
50. Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. *J Chem Phys* 1984, 81, 3684.
51. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J Chem Phys* 1983, 79, 926.
52. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. *J Comput Chem* 2005, 26, 1701.
53. Beutler, T.; Mark, A.; van Schaik, R.; Gerber, P.; van Gunsteren, W. *Chem Phys Lett* 1994, 222, 529.
54. Ben-Naim, A.; Marcus, Y. *J Chem Phys* 1984, 81, 2016.
55. Huber, J.; Cronshagen, U.; Kadokura, M.; Marshallsay, C.; Wada, T.; Sekine, M.; Luhrmann, R. *EMBO J* 1998, 17, 4114.