THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

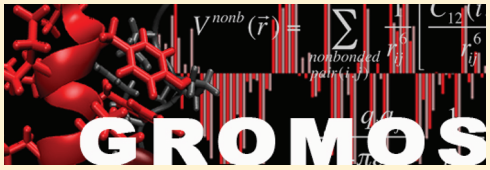
Calculation of Relative Free Energies for Ligand-Protein Binding, Solvation, and Conformational Transitions Using the GROMOS Software

Sereina Riniker,[†] Clara D. Christ,[†,§] Halvor S. Hansen,[†] Philippe H. Hünenberger,[†] Chris Oostenbrink,[‡] Denise Steiner,[†] and Wilfred F. van Gunsteren*[,†]

[†]Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, 8093 Zürich, Switzerland

[‡]Institute of Molecular Modeling and Simulation, University of Natural Resources and Life Sciences, Vienna, Austria

ABSTRACT: The calculation of the relative free energies of ligand-protein binding, of solvation for different compounds, and of different conformational states of a polypeptide is of considerable interest in the design or selection of potential enzyme inhibitors. Since such processes in aqueous solution generally comprise energetic and entropic contributions from many molecular configurations, adequate sampling of the relevant parts of configurational space is required and can be achieved through molecular dynamics simulations. Various techniques to obtain converged ensemble averages and their implementation in the GROMOS software for biomolecular simulation are discussed, and examples of their application to biomolecules in aqueous solution are given.



INTRODUCTION

The driving force for all (bio)molecular processes is the change of free energy. The estimation of free energy changes from computer simulations, however, remains a challenging task. Free energy calculations can be classified into two categories. The calculation of conformational free energies aims at evaluating the relative free energies of relevant conformational states of a given molecular system. The calculation of alchemical free energies on the other hand aims at evaluating the relative free energies of different molecules. The computational methods discussed in this paper will be classified accordingly. With the use of computer simulations, either the difference in free energy or the difference in free enthalpy of a process can be calculated. The relative (Helmholtz) free energy ΔF is obtained under isochoric-isothermic conditions (NVT), whereas the relative (Gibbs) free enthalpy ΔG is obtained under isobaric-isothermic conditions (NpT). The term free energy is also often used in a general sense, as will be done in the present article. Experimental data such as the thermodynamic binding constant K_i are usually measured at constant temperature and pressure as these are the conditions most typical of experimental situations. K_i is directly correlated to the free energy difference ΔG^{bind}

$$\Delta G^{\mathrm{bind}} = RT \ln(K_i) \tag{1}$$

which is the free energy difference between the bound state, i.e., the ligand−protein complex, and the unbound state, i.e., the free ligand and protein in solution

$$\Delta G^{\mathrm{bind}} = G^{\mathrm{complex}} - G^{\mathrm{free}} \tag{2}$$

ΔG^{bind} is a conformational free energy difference and its direct calculation from a molecular dynamics (MD) or Monte Carlo (MC) simulation requires many transitions of the ligand from the unbound to the bound state and back to occur in order to obtain an accurate ensemble average estimate of the population ratio of the two states. It is therefore often computationally more efficient to calculate the relative value $\Delta\Delta G_{\mathrm{BA}}^{\mathrm{bind}}$ for two ligands A and B using the alchemical perturbations from ligand A to ligand B in the bound and unbound state, $\Delta G_{\mathrm{BA}}^{\mathrm{complex}}$ and $\Delta G_{\mathrm{BA}}^{\mathrm{free}}$, instead of ΔG^{bind} for each ligand. The difference between $\Delta G_{\mathrm{BA}}^{\mathrm{complex}}$ and $\Delta G_{\mathrm{BA}}^{\mathrm{free}}$ is the relative binding free energy $\Delta\Delta G_{\mathrm{BA}}^{\mathrm{binding}}$ and can be directly compared to the experimental binding data using eq 1 and the relation

$$\begin{aligned}\Delta\Delta G_{\mathrm{BA}}^{\mathrm{binding}} &= \Delta G_{\mathrm{B}}^{\mathrm{bind}} - \Delta G_{\mathrm{A}}^{\mathrm{bind}} \\ &= \Delta G_{\mathrm{BA}}^{\mathrm{complex}} - \Delta G_{\mathrm{BA}}^{\mathrm{free}}\end{aligned} \tag{3}$$

For the estimation of alchemical free energy differences, the end states A and B generally have to be connected by a composite Hamiltonian. The composite Hamiltonian is a function of the end state Hamiltonians H_{A} and H_{B}, or in practice of the corresponding end state potential energies V_A and V_B, and can be formulated in two ways: dependent on a coupling parameter λ or without an explicit dependence on λ. The thermodynamic integration[1] (TI) method belongs to the first approach. The system is perturbed along a user-defined pathway between the end states by introducing a λ-dependence into the Hamiltonian, $H(\vec{r}^N,\vec{p}^N;\lambda)$, with $\vec{r}^N = (\vec{r}_1, \vec{r}_2, ..., \vec{r}_N)$ denoting the Cartesian coordinates of all N particles in the system and $\vec{p}^N = (\vec{p}_1, \vec{p}_2, ..., \vec{p}_N)$ denoting the corresponding conjugate momenta. The

Received: May 9, 2011
Revised: October 13, 2011
Published: November 01, 2011

sampling in TI can be enhanced using Hamiltonian replica exchange (H-RE),[2,3] where the different parallel simulations (replicas with different λ values) differ in their interactions; that is, they have different Hamiltonians. In one-step perturbation[4,5] (OSP), the composite Hamiltonian or reference Hamiltonian H_R is not dependent on a coupling parameter λ, but defined by the user such that its relevant configurations overlap as much as possible with those of all the end states of interest. From a single simulation of H_R, the free energy differences ΔF_{XR} between many end states X and the reference state R can be estimated. The reference Hamiltonian H_R in enveloping distribution sampling[6−9] (EDS) is, as in OSP, not dependent on a coupling parameter λ but, in contrast to H_R in OSP, it is defined through the end states. The sampled ensemble is expanded in such a way that all phase space regions important to the end states are also important to the reference state, thus H_R envelops the end state Hamiltonians.

Free energy differences between two conformational subspaces can be calculated using local elevation umbrella sampling[10] (LEUS), where the local elevation[11] (LE) method and the umbrella sampling[12,13] (US) method are combined for searching and sampling the conformational space. In US, a biasing potential energy term is added to the Hamiltonian with the aim to focus the sampling on the important regions in phase space. LE is thereby used to construct such a biasing potential energy term. By applying hidden restraints, TI can also be used for the calculation of free energy differences between conformational end states which are separated by high energy barriers.[14] Thereby, distance or dihedral-angle restraints are used in such a way that they only act in the intermediate states and are zero at the end states, leaving these unperturbed and therefore no corrections for the restraints have to be applied.

For yet other methods to compute alchemical or conformational free energy differences, we refer to ref 15. For all of these statistical-mechanically sound methods, the accuracy of the relative free energy estimates obtained in practice critically depends on the proper choice of some simulation parameter: the λ-dependence in TI, the biasing potential energy function in LEUS or the reference state H_R in OSP and EDS.

Here, we describe how these different computational methods that require a modified Hamiltonian to estimate free energy differences are implemented in the GROMOS biomolecular simulation software and give examples of their use. The implementation of methods to compute free energy differences from an MD trajectory generated using a physical, not alchemical Hamiltonian by postprocesing the configurational ensemble, such as Widom particle insertion, is discussed in ref 16. The architecture of the GROMOS software has been described in ref 17, whereas other new functionalities that were added since 2005 are described in ref 18 and 19 and the GROMOS++ set of analysis programs to obtain ensemble averages and time series of a variety of quantities from configurational trajectories is described in ref 16.

METHODS

Thermodynamic Integration. The Hamiltonian and the free energy of the system are made dependent on a coupling parameter λ, where the nature of the dependence on λ can be chosen freely as the free energy G is a state function; that is, it is path independent. The λ-dependence of the free energy, of the energies, of the Hamiltonian, and of its derivative $\partial H/\partial\lambda$ used in GROMOS are specified in refs 20−23. It may be specified independently for different interaction terms between given sets of atoms (energy groups). In addition, the level of softness for nonbonded interactions can be modified separately through individual λ values.[24] For the nonbonded potential energy term between two atoms i and j, assigned to energy groups I and J, it is possible to define the functions $\Lambda_{IJ}^{LJ}(\lambda)$ and $\Lambda_{IJ}^{CRF}(\lambda)$ to scale the Lennard-Jones and the Coulomb reaction-field interactions, respectively. Corresponding functions $\Lambda_{IJ}^{SL}(\lambda)$ and $\Lambda_{IJ}^{SC}(\lambda)$ are defined for the softness of these interactions. The individual λ values are user defined fourth-order polynomials of the overall value of λ. The nonbonded potential energy term becomes

$$\begin{aligned} V^{nonb}(\vec{r}_{ij};\lambda) = {} & (\Lambda_{IJ}^{LJ}(\lambda))^n V_B^{LJ}(\vec{r}_{ij};1-\Lambda_{IJ}^{SL}(\lambda)) \\ & + (1-\Lambda_{IJ}^{LJ}(\lambda))^n V_A^{LJ}(\vec{r}_{ij};\Lambda_{IJ}^{SL}(\lambda)) \\ & + (\Lambda_{IJ}^{CRF}(\lambda))^n V_B^{CRF}(\vec{r}_{ij};1-\Lambda_{IJ}^{SC}(\lambda)) \\ & + (1-\Lambda_{IJ}^{CRF}(\lambda))^n V_A^{CRF}(\vec{r}_{ij};\Lambda_{IJ}^{SC}(\lambda)) \end{aligned} \tag{4}$$

where n is a strictly positive integer and

$$V_X^{LJ}(\vec{r}_{ij};\lambda) = \frac{1}{\alpha_{LJ}(i,j)\lambda^2 C_{126}^X(i,j) + {r_{ij}}^6}\left[\frac{C_{12}^X(i,j)}{\alpha_{LJ}(i,j)\lambda^2 C_{126}^X(i,j) + {r_{ij}}^6} - C_6^X(i,j)\right] \tag{5}$$

$$V_X^{CRF}(\vec{r}_{ij};\lambda) = \frac{q_i^X q_j^X}{4\pi\varepsilon_0\varepsilon_{CS}}\left[\frac{1}{(\alpha_c(i,j)\lambda^2 + {r_{ij}}^2)^{1/2}} - \frac{\frac{1}{2}C_{RF}{r_{ij}}^2}{(\alpha_c(i,j)\lambda^2 + {R_{RF}}^2)^{3/2}} - \frac{1-\frac{1}{2}C_{RF}}{R_{RF}}\right] \tag{6}$$

where the interaction parameters that are different in states A and B are indicated by the superscript X (= A or B), q is the charge, ε_{CS} the dielectric permittivity inside the cutoff sphere, C_{RF} and R_{RF} the reaction-field parameters, and α_{LJ} and α_C the soft-core nonbonded interaction parameters which are used to avoid numerical instabilities when the nonbonded interactions between atoms are switched on or off during the perturbation. The soft-core parameters can be defined separately for each atom.[25] If the system is changed stepwise, under the control of λ, from state A ($\lambda_A = 0$) to state B ($\lambda_B = 1$), the free energy difference ΔG_{BA} is given by

$$\Delta G_{BA} = G(\lambda_B) - G(\lambda_A) = \int_{\lambda_A}^{\lambda_B}\left\langle\frac{\partial H(\lambda)}{\partial\lambda}\right\rangle_\lambda d\lambda \tag{7}$$

where the ensemble average for the system $H(\lambda)$ is denoted by $\langle\ldots\rangle_\lambda$. In practice, ΔG_{BA} is obtained by performing a finite number, N_λ, of simulations at discrete λ values ranging from 0 to 1 and subsequent numerical integration using an interpolation formula. Given a sufficiently long simulation time, the system is in equilibrium at every λ value and $\langle\partial H(\lambda)/\partial\lambda\rangle_\lambda$ is converged.

Using an optimal path along which the system is perturbed, the sampling problems are minimized. The choice of this pathway,

however, is not trivial in all cases and may strongly influence the convergence of $\langle \partial H(\lambda)/\partial\lambda \rangle_\lambda$ at particular λ values.

Thermodynamic Integration with Hamiltonian Replica Exchange. For cases in which the convergence at particular λ values poses a problem due to conformational changes on long time scales, sampling enhancement techniques such as Hamiltonian replica exchange (H-RE) can be applied.[2,3] In TI-H-RE, the replicas differ in their Hamiltonian $H(\vec{r}^N,\vec{p}^N;\lambda)$ and a MC exchange of configurations between two replicas i and j is attempted after fixed time intervals in a MD simulation, by using the switching probability p_{ij} derived from the requirement that the two states i and j are in detailed balance (Metropolis like criterion[26])

$$p_{ij} = \begin{cases} 1 & \text{for} \Delta \le 0 \\ \exp(-\Delta) & \text{for} \Delta > 0 \end{cases} \tag{8}$$

with

$$\Delta = \beta[V_i(\vec{r}_j^N) - V_i(\vec{r}_i^N)] - \beta[V_j(\vec{r}_j^N) - V_j(\vec{r}_i^N)] \tag{9}$$

$\beta = 1/k_B T$ and V_i being the potential energy function at λ_i.

Hidden Restraints. For systems where a high energy barrier separates the end states, leading to a large change of H as function of λ, or where the stabilization of a transition state is necessary, so-called hidden restraints can be applied to enhance the sampling around the barrier while keeping the end states A and B unperturbed.[14] A weight factor is introduced in the distance or dihedral-angle restraining energy function which is one at the transition state or on top of the separating barrier and smoothly decreases to zero at both end states

$$V^{\text{restr}}(\vec{r}^N) = 2^{n+m}\lambda^n(1-\lambda)^m V^{\text{AB}}_{\text{restr}}(\vec{r}^N;\lambda) \tag{10}$$

where $V^{\text{AB}}_{\text{restr}}$ is the λ-dependent restraining potential energy term that induces transitions between states A and B and $2^{n+m}\lambda^n(1-\lambda)^m$ is the weight factor (see ref 14 for further details). Thus, the end states remain unperturbed while transitions between them may be forced.

Implementation Details. When looping over the pairlist for nonbonded interactions, it is checked if an atom is perturbed or not and only the interactions with perturbed atoms are calculated twice: once for Hamiltonian A and once for Hamiltonian B.

The average $\langle \partial H(\lambda)/\partial\lambda \rangle_\lambda$ and a corresponding error estimate using block averaging can be obtained from the trajectory using the GROMOS++ program[16] *ene_ana*.

Local Elevation Umbrella Sampling. To overcome the barrier between two conformational subspaces of a system, e.g., cis and trans conformations of the torsional angle around a double bond, LEUS combines the advantages of the local elevation[11] (LE) and the umbrella sampling[12,13] (US) methods. In a LE phase of a simulation, a memory-based biasing potential energy term V_{LE} is built up which penalizes the resampling of previously visited configurations and allows the system to overcome barriers. As the corresponding Hamiltonian is time-dependent, LE can only be used for searching the conformational space but gives no information on the thermodynamics of the system. The biasing potential energy term obtained by LE is subsequently used in the US phase of a simulation. In US, an arbitrary, time-independent biasing potential energy term V_{US} is added to the Hamiltonian of the system

$$H_{\text{bias}}(\vec{r}^N, \vec{p}^N) = H(\vec{r}^N, \vec{p}^N) + V_{\text{US}}(\vec{R}) \tag{11}$$

where $V_{\text{US}} = V_{\text{LE}}$ and $\vec{R}$ is a vector associated with conformation $\vec{r}^N$ in the N_{US}-dimensional configurational subspace in which US is applied. The ensemble average of any observable Q of the real system is then obtained by reweighting

$$\langle Q \rangle = \frac{\langle Q(\vec{r}^N, \vec{p}^N) e^{\beta V_{\text{US}}(\vec{R})} \rangle_{\text{bias}}}{\langle e^{\beta V_{\text{US}}(\vec{R})} \rangle_{\text{bias}}} \tag{12}$$

The free energy of a conformational subspace S_L is then given by

$$G(\vec{L}) = -\frac{1}{\beta} \ln \frac{\langle \delta(\vec{R} - \vec{L}) e^{\beta V_{\text{US}}(\vec{R})} \rangle_{\text{bias}}}{\langle e^{\beta V_{\text{US}}(\vec{R})} \rangle_{\text{bias}}} \tag{13}$$

where δ is the N_{US}-dimensional Dirac delta function. If two conformational subspaces $\vec{A}$ and $\vec{B}$ are defined, the associated free energy difference ΔG_{BA} can be obtained from

$$\Delta G_{\text{BA}} = -\frac{1}{\beta} \ln \frac{\langle \delta(\vec{R} - \vec{B}) e^{\beta V_{\text{US}}(\vec{R})} \rangle_{\text{bias}}}{\langle \delta(\vec{R} - \vec{A}) e^{\beta V_{\text{US}}(\vec{R})} \rangle_{\text{bias}}} \tag{14}$$

Implementation Details. The quantity $V_{\text{US}}(\vec{R})$ is stored along with each successive configuration and the reweighting involved in eq 12 is done during postprocessing of the molecular trajectories. Equation 12 can be executed using the GROMOS++ program[16] *reweight*, eq 13 using the GROMOS++ program *dg_ener*.

One-Step Perturbation. If the relevant configurations of the Hamiltonian of a reference state R show overlap with those of the Hamiltonian of an end state A, the free energy difference between A and R can be estimated directly using[4,5]

$$\Delta G_{\text{AR}} = G_{\text{A}} - G_{\text{R}} = -\frac{1}{\beta} \ln \langle e^{-\beta(H_{\text{A}} - H_{\text{R}})} \rangle_{\text{R}} \tag{15}$$

If the relevant configurations of the Hamiltonian of a different end state B also show overlap with those of the reference state Hamiltonian, the free energy difference between the two end states is given by

$$\Delta G_{\text{BA}} = \Delta G_{\text{BR}} - \Delta G_{AR} \tag{16}$$

Thus, multiple free energy differences can be estimated from a single simulation of the reference state. However, because the ensembles of the end states are extrapolated from the ensemble of the reference state, the accuracy of the resulting free enthalpies is highly dependent on whether the configurations sampled by the reference state are relevant to the end states, and the choice of such a reference state may not be trivial. To enhance sampling when atoms are created or deleted, a soft-core interaction function as described in eqs 5 and 6 is to be used for such atoms in the reference state.

Implementation Details. To obtain the energies of an end state from the reference state simulation, the GROMOS++ program[16] *ener* can be applied which recalculates the interaction energies over the molecular trajectory files from the reference state simulation using the interaction parameters specified in the molecular topology file of the end state. The nonbonded potential energy term is calculated using eqs 5 and 6 and the bonded potential energy terms are described in refs 20−22. The free energy differences according to eq 15 can be calculated using GROMOS++ program[16] *dg_ener*.

Enveloping Distribution Sampling. The reference state in EDS is related to the sum of the Boltzmann factors of the N end-state Hamiltonians. Two parameter sets, the smoothness

parameters s and the energy offsets E^R_i are introduced to provide more even sampling of all end states.[7] The optimized reference Hamiltonian $H_R = K_R + V_R$ has for two states the form[7]

$$V_R(\vec{r}^N; s, E^R) = -\frac{1}{\beta}\ln[(e^{-\beta s(V_A(\vec{r}^N) - E^R_A)} + e^{-\beta s(V_B(\vec{r}^N) - E^R_B)})^{1/s}] \quad (17)$$

where the kinetic terms K_R are omitted for simplicity, s is the smoothness parameter, and E^R_i is the energy offset parameter of end state i. The corresponding force on a particle k reads[7]

$$\vec{f}_k(t) = \left(-\frac{V_R(\partial \vec{r}^N; s, E^R)}{\partial \vec{r}_k}\right) = [1 + e^{+\beta s(\Delta V_{BA}(\partial \vec{r}^N) - \Delta E^R_{BA})}]^{-1}\left(-\frac{\partial V_B(\partial \vec{r}^N)}{\partial \vec{r}_k}\right) + [1 + e^{-\beta s(\Delta V_{BA}(\partial \vec{r}^N) - \Delta E^R_{BA})}]^{-1}\left(-\frac{\partial V_A(\partial \vec{r}^N)}{\partial \vec{r}_k}\right) \quad (18)$$

with $\Delta V_{BA}(\vec{r}^N) = V_B(\vec{r}^N) - V_A(\vec{r}^N)$ and $\Delta E^R_{BA} = E^R_B - E^R_A$. The reference Hamiltonian can be generalized to multiple states by determining either an s parameter for each pair or by using a maximum spanning tree where only those states closest in phase space, i.e., with the largest s ($0 < s \leq 1$), are directly connected.[9]

The free energy difference ΔG_{BA} between two end states A and B can then be obtained from

$$\Delta G_{BA} = G_B - G_A = -\frac{1}{\beta}\ln\frac{\langle e^{-\beta(V_B - V_R)}\rangle_R}{\langle e^{-\beta(V_A - V_R)}\rangle_R} \quad (19)$$

where the ensemble average over the system H_R is denoted as $\langle ... \rangle_R$.

Parameter Update Scheme. If the energy difference $\langle V_{ji}\rangle$ between two states i and j is large, only one of the states would be sampled in a simulation. Therefore, the energy offset parameters E^R_i and E^R_j were introduced into H_R to remove the bias toward the state with the lower energy. If the important parts of phase space of two states i and j are well separated, a small s ($0 < s \leq 1$) is required in H_R for transitions between these regions to occur. A first update scheme to determine the optimal reference state parameters was proposed by Christ and van Gunsteren.[8,9] It was found, however, that the s parameter was systematically underestimated by the approach, which led to the development of an improved update scheme.[27] The new approach makes use of the condition that, using an optimal reference state, all end states should be sampled equally. To find these parameters, the non-reweighted end-state energies in the reference state simulation, e.g., for two states V_A and V_B, are evaluated. Alternatively, the end-state energy difference ΔV_{BA} can be investigated. The occupancy numbers, N_A and N_B, are determined by assigning each configuration of a simulation to one of the end states depending on its energy. As the ratio of the occupancy numbers N_A/N_B should be 1 for optimal parameters, it can be used to evaluate the direction in which the s parameter has to be evolved. The initial parameters can either be chosen in a top-down (starting with $s = 1$) or in a bottom-up (starting with a small $s > 0$) approach. The modification of the energy offset parameters E^R_i is based on counting the number of visits of a state

$$E^R_i(\text{new}) = -\frac{1}{\beta}\ln\left\langle\left[1 + \sum_{\substack{j=1\\ j\neq i}}^{N} e^{-\beta(\Delta V_{ji} - \Delta E^R_{ji})}\right]^{-1}\right\rangle_{s(\text{new}), E^R(\text{old})} + E^R_i(\text{old}) \quad (20)$$

for all states i, where the ensemble average over the reference state is defined using the new s parameters s(new) and the old energy offsets E^R(old) to which state the energy trajectories are reweighted. The ensemble average $\langle Q\rangle_{s(\text{new}),E^R(\text{old})}$ of a quantity $Q(\vec{r}^N)$ obtained from a simulation using $V_R(\vec{r}^N;s(\text{new}),E^R(\text{old}))$ can be expressed by reweighting in terms of two ensemble averages obtained from a simulation using $V_R(\vec{r}^N;s(\text{old}),E^R(\text{old}))$,

$$\langle Q\rangle_{s(\text{new}),E^R(\text{old})} = \frac{\langle Q e^{-\beta[V_R(s(\text{new}),E^R(\text{old})) - V_R(s(\text{old}),E^R(\text{old}))]}\rangle_{s(\text{old}),E^R(\text{old})}}{\langle e^{-\beta[V_R(s(\text{new}),E^R(\text{old})) - V_R(s(\text{old}),E^R(\text{old}))]}\rangle_{s(\text{old}),E^R(\text{old})}} \quad (21)$$

Optimal s and energy offset parameters are found when the reference state configurations switch between relevant end state configurations. The new update scheme is described in more detail in ref 27.

Implementation Details. In EDS, either a dual or a single topology approach can be used.[28−31] In the case of different ligands, a common core of atoms is left unperturbed in a single topology and the substituents are introduced as dummy atoms which are perturbed. In a dual topology approach on the other hand, each ligand is simulated as a separate end state and an EDS state comprises thus one active end state ligand and $N - 1$ inactive (dummy) end state ligands. To avoid spatial drifting of ligands in their inactive state, all ligands are held together by weak distance restraints.

Reweighting of the end state energies can be performed using the GROMOS++ program[16] *reweight*. The free energy difference ΔG_{BA} can be calculated from eq 19 using the GROMOS++ program *dfmult*.

EXAMPLE APPLICATIONS

The literature contains many examples of the application of techniques to obtain free energy differences from ensemble averages, e.g., refs 15 and 32. Below we present one example for each of the methods, i.e., thermodynamic integration in combination with Hamiltonian replica exchange, local elevation umbrella sampling, one-step perturbation, and enveloping distribution sampling. Since the applicability and efficiency of the different techniques to compute relative free energies depend on the particular shape of the energy hypersurface and its regions of interest, the examples are chosen such as to highlight the type of efficient application for each method.

Thermodynamic Integration with Hamiltonian Replica Exchange. The free energy differences of a series of inhibitors binding to the protein plasmepsin (PM) II were studied using TI in conjunction with Hamiltonian replica exchange (TI-H-RE).[33] Simulations of the ligands in solution and bound to the protein were performed at 310 K and constant volume using the GROMOS force field[34] 53A6 and the GROMOS software package.[17,19,22] The initial structure for the protein−ligand complex was taken from the Protein Data Bank, PDB code

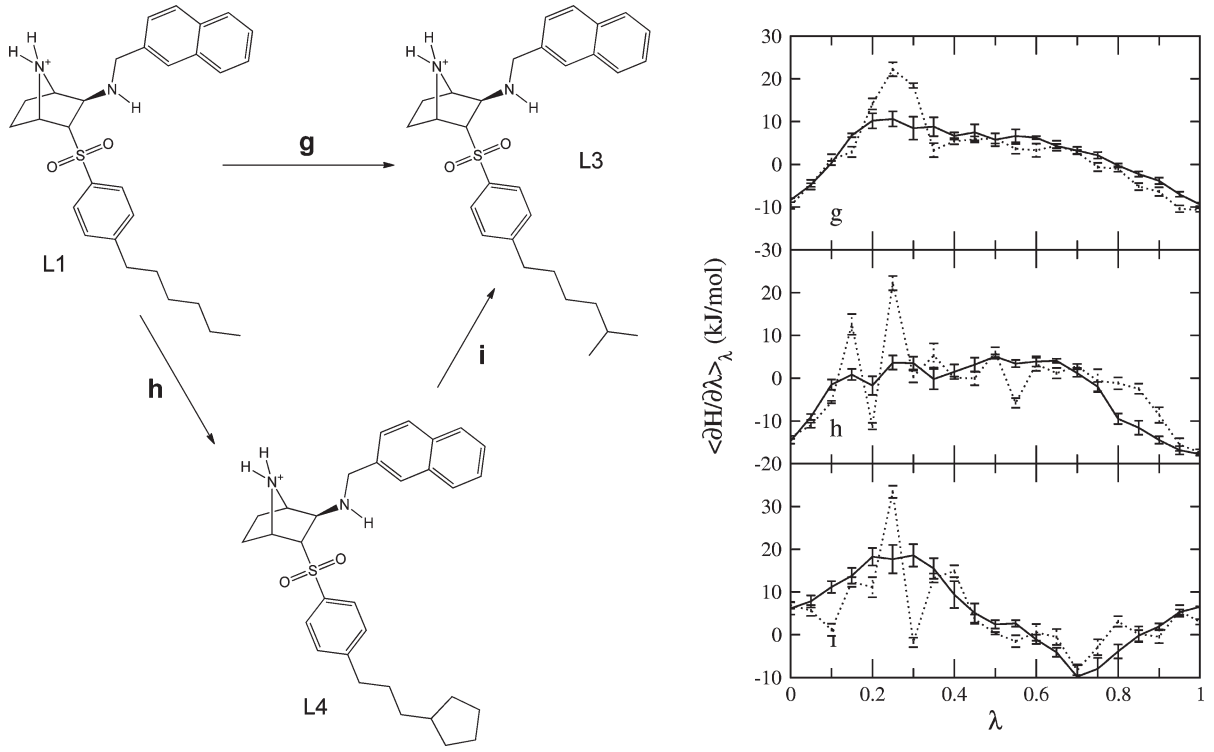


Figure 1. Example of the use of thermodynamic integration in combination with Hamiltonian replica exchange.[33] Left: Perturbation between the ligands L1 and L3 (**g**), L1 and L4 (**h**), and L3 and L4 (**i**) of the protein plasmepsin II in water. Right: $\langle \partial H/\partial \lambda \rangle_\lambda$ as function of λ for the transitions **g**, **h**, and **i** for the ligand−protein complex. The results from standard TI are shown as dotted lines and from TI in combination with H-RE as solid lines. Error bars indicate the statistical error estimate on the ensemble averages.

2BJU,[35] and subsequently energy minimized and equilibrated. Simulations of 300 ps length were carried out at 21 equidistant λ values and were extended up to 1200 ps if needed. Using conventional TI simulations, the protein−ligand complex was found to be too large and too flexible to give converged averages $\langle \partial H/\partial \lambda \rangle_\lambda$ at each λ value. Poor convergence was found both in terms of the statistical error estimates per λ value as well as in terms of the smoothness of the λ curves.

By applying TI-H-RE to the system, more extended sampling at each λ value occurred, smoother λ curves were obtained and the total error over the whole transition was decreased, although the simulation time per λ value was only 200 ps compared to 300−1200 ps in case of conventional TI. In Figure 1, the λ curves from standard TI and TI in combination with H-RE are shown for three perturbations of the ligand as example. Although the application of H-RE markedly enhanced the sampling and the convergence of the TI-results, reproducing the trends in the experimental IC_{50} values remained challenging for other reasons.[33]

Local Elevation Umbrella Sampling. LEUS was applied to estimate the free energy differences between sugar ring conformations of β-D-glucopyranose in water, which represents a particulary difficult conformational transition problem.[10] Pyranose systems can adopt at least eight relevant conformers, i.e., two chair conformations and six boat forms. One of the chair conformations, i.e., $^4C_1$, is the only ring conformer significantly populated in aqueous solution. The stability difference between the chair conformers is estimated to lie between 25 and 45 kJ mol^{-1} and the interconversion time is on the μs time scale. Processes on such time scales cannot be investigated using standard MD simulations, thus requiring sampling enhancement techniques such as LEUS. To this end, a three-dimensional conformational subspace was defined by the three out-of-plane dihedral angles α_1, α_2, and α_3, and the various ring conformers, i.e., conformational states, were described in terms of these three angles (top panel in Figure 2). Simulations were performed at 300 K and 1 atm using the GROMOS force field[36] 45A4 and a modified version of the GROMOS96 software package.[20,21]

The results of a 100 ns standard MD simulation were compared with the results from a 100 ns sampling period of a LEUS simulation (lower panels in Figure 2). Using standard MD, the chair conformation $^4C_1$ was dominantly sampled and to much lesser extent the boat conformation $^{3,0}B$. Using LEUS on the other hand, the sampling was dramatically enhanced and, after reweighting of the ensemble to the original Hamiltonian, a third high probability region appeared which corresponds to the second chair conformation $^1C_4$. This second chair conformation has also a low free energy but due to a high interconversion barrier, it is not sampled with standard MD. A fourth (shallow) minimum was assigned to the skew-boat conformation $^1S_3$. The relative free energies of these four conformations were estimated to be 0.0 kJ mol^{-1} for $^4C_1$, 16.2 kJ mol^{-1} for $^1C_4$, 18.7 kJ mol^{-1} for $^{3,0}B$, and 32.8 kJ mol^{-1} for $^1S_3$. The corresponding time scales for the transition from $^4C_1$ to $^1C_4$, $^{3,0}B$, and $^1S_3$ were found to be 0.8 μs, 15 ns, and 0.5 μs, respectively, which explains why only $^{3,0}B$ is sampled in a standard MD simulation of 100−200 ns length. The thermodynamic and kinetic results obtained using LEUS[10] were qualitatively comparable with the limited amount of experimental and quantum-mechanical data.

One-Step Perturbation. The effect of force-field changes on the folding equilibrium of a hepta-β-peptide in methanol was

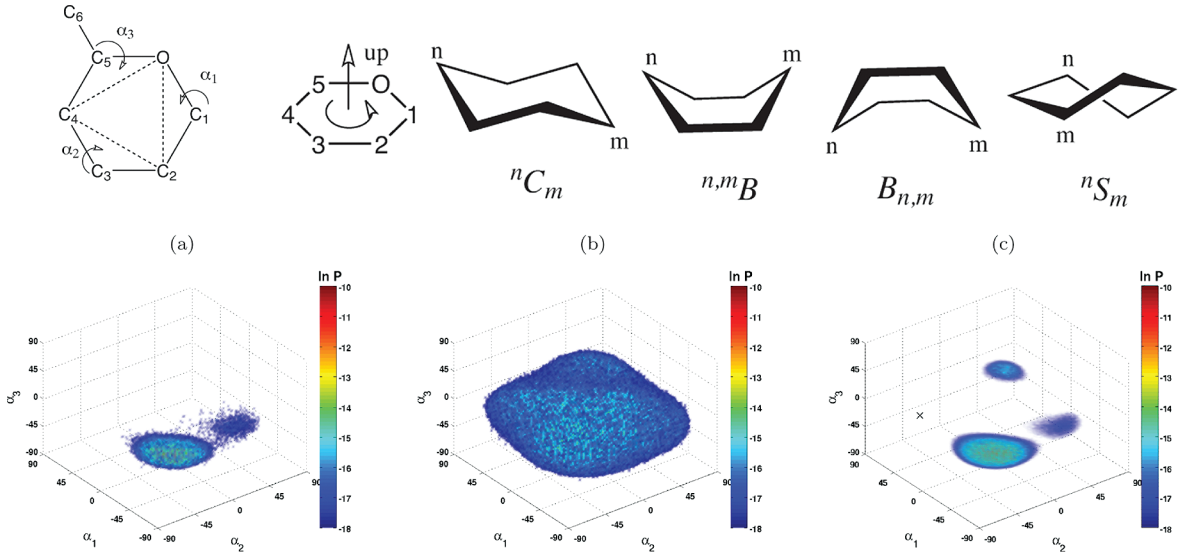


Figure 2. Example of the use of local-elevation umbrella sampling.[10] Top: Definition of the out-of-plane dihedral angles α_1, α_2, and α_3, and the idealized ring conformations with their nomenclature for a pyranose in water. Bottom: Conformational probability distributions in the α space of (a) the 100 ns unbiased simulation and (b) the 100 ns sampling phase of LEUS; (c) represents the distributions of (b) after unbiasing using eq 12 (the cross indicates the location of the $^1S_3$ idealized conformation).

studied by Lin et al.[37] using one-step perturbation. The method permits the estimation of the change in the free energy of folding for a variety of force-field parameter sets from a single simulation, the prediction of which would be computationally very expensive using TI or other multistep methods. Simulations were performed at 340 K and 1 atm using the GROMOS software package.[17,19,22] A simulation of 101 ns length was carried out for the reference state which corresponded to the GROMOS parameter set 45A3.[38] The configurations were separated in two subensembles, folded (C_F) and unfolded (C_U), based on their atom-positional root-mean-square differences (rmsd) from the most stable, helical folded conformation. The free energy of folding in the reference state, $\Delta G_R^{F,U}$, was calculated by direct counting of the number of conformations in the two ensembles. The free energy change of a subensemble C due to changing the Hamiltonian from H_R to H_{pert}, involving a change of force-field parameters, was estimated using 15, and subsequently the free energy of folding in the perturbed state with the changed force-field parameters was predicted using the thermodynamic cycle $\Delta G_{pert}^{F,U} = \Delta G_R^{F,U} + \Delta G_{pert,R}^{F} - \Delta G_{pert,R}^{U}$.

For selected perturbations, the results were verified by performing simulations using the modified force-field parameters. The change in the free energy of folding was estimated correctly by one-step perturbation for changes in the backbone partial charges, the backbone dihedral angle energy force constants, the cutoff distance, and the dielectric permittivity of the reaction-field force.[37] The relative perturbation free enthalpies $\Delta\Delta G_{pert,R}^{F,U} = \Delta G_{pert,R}^{F} - \Delta G_{pert,R}^{U}$ for changing the backbone partial charges and the backbone dihedral angle energy force constants are shown in Figure 3. Increasing the solute partial charges or backbone torsional-angle force constants leads to a stabilization of the helical fold. A perturbation involving the change of partial charges of the solvent or the entire solvent model, however, was too large to yield precise results through one-step perturbation.[37]

Enveloping Distribution Sampling. The performance of EDS was tested on a set of 10 tetrahydroisoquinoline derivatives which are inhibitors of human phenylethanolamine N-methyltransferase (PNMT).[27] The free energy differences between various pairs of ligands in the bound and unbound state obtained using EDS were compared to results from TI simulations. The simulations were performed at 298 K and 1 atm using the GROMOS force field[34] 53A6 and the GROMOS software package.[17,19,22] Both a single and a dual topology approach were investigated and the reference state parameters s and E_i^R were determined with the original and the new update schemes. No difference between the two topology approaches were found. For the original update scheme proposed by Christ et al.[8] a systematic underestimation of the s parameter in case of large perturbations was observed due to an artificial dependence of s on the difference in the energy offsets ΔE_{BA}^R.

On the other hand, the reference state parameters obtained with the new update scheme produced converged results which agree well with the results from TI for the ligands in water and reasonably well for the ligands bound to the protein. The nonbonded potential energy distributions of the end states A and B from simulations using reference state parameters obtained from the original and the new update scheme are compared in the left panel of Figure 4 for two perturbations, 1−5 and 1−7, in the bound and unbound state. The end states were sampled well with both parameter sets in case of the small perturbation 1−5, but for the large perturbation 1−7 only the new parameter set yielded converged distributions.[27]

The reason for the deviations between EDS and TI results in case of the ligand-protein complexes lies in the flexibility of some ligands inside the binding pocket. Although each end state was sampled well, as was shown by converged end state energy distributions, the distributions do not overlap for the same end state in different perturbations which means that not the same range of potential energies was sampled and thus not a similar range of configurations. Therefore the thermodynamic cycles did not close well for the ligand-protein complexes. The relative binding free enthalpies obtained using EDS did agree well with experimental data.[27]

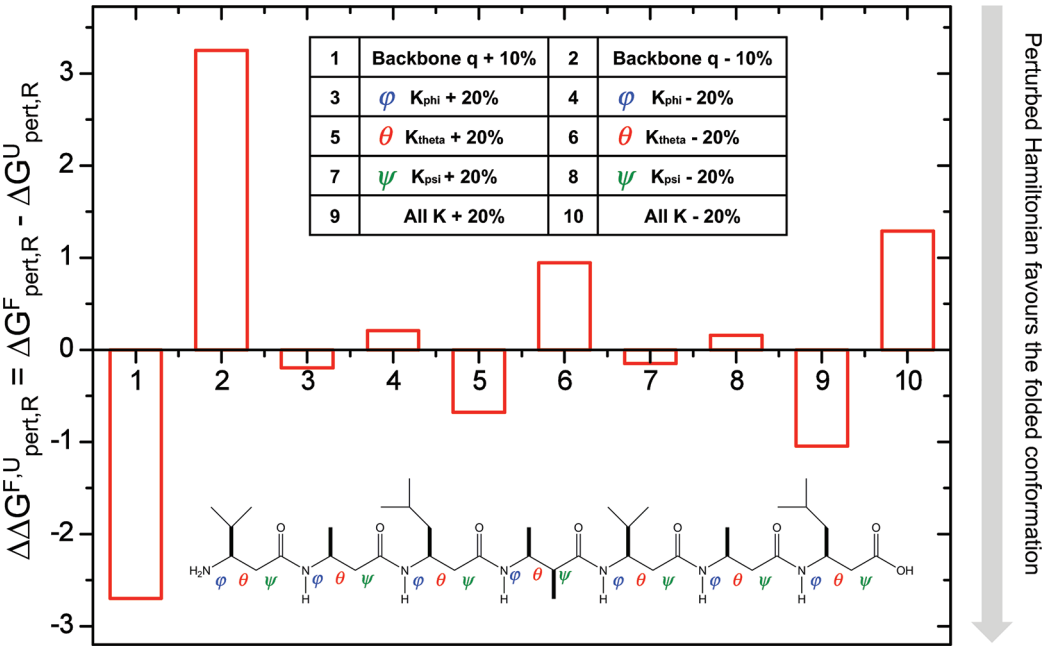
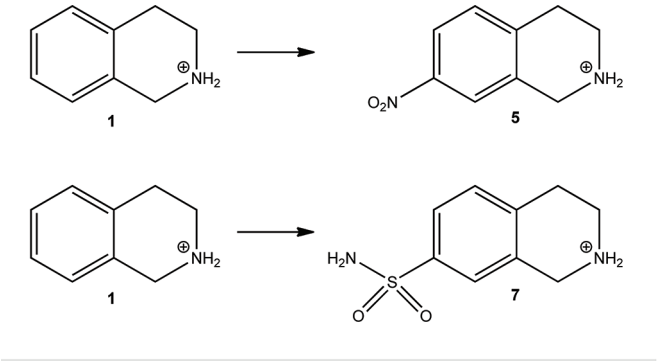


Figure 3. Example of the use of one-step perturbation.[37] Relative perturbation free enthalpies of folding $\Delta\Delta G^{F,U}_{pert,R} = \Delta G^{F}_{pert,R} - \Delta G^{U}_{pert,R}$ in kJ mol^{-1} associated with perturbing the backbone charges q or the force constants K of the torsional potential energy terms for the backbone dihedral angles ϕ, θ, and ψ of a hepta-β-peptide in methanol estimated by one-step perturbation. A negative value indicates that the perturbed Hamiltonian favors the helical folded conformation more than the nonperturbed Hamiltonian.



Pert.	$\Delta G^{complex}$	ΔG^{free}	$\Delta\Delta G^{binding}$(calc)	$\Delta\Delta G^{binding}$(exp)
1-5	-4.0 ± 0.9	6.2 ± 0.1	-10.2 ± 1.0	-10.3
1-7	-352.0 ± 2.0	-343.6 ± 0.2	-8.4 ± 2.2	-9.6

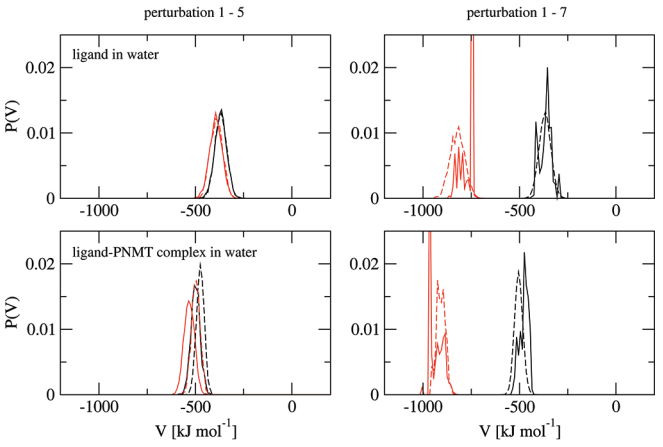


Figure 4. Example of the use of enveloping distribution sampling.[27] Left: Perturbation between inhibitors 1 and 5, and 1 and 7 of the protein PNMT for the bound and unbound ligands with the corresponding free energy differences and the resulting relative binding free enthalpies using the single topology approach and reference state parameters obtained from the new update scheme. Experimental relative binding free enthalpies were calculated from binding constants[40] K_i using eq 1 and eq 3. Right: Nonbonded potential energy distributions of the end states A (black) and B (red) from the perturbations 1−5 and 1−7 using reference state parameters obtained from the original update scheme (solid lines) and from the new update scheme (dashed lines).

DISCUSSION

The functionality of the GROMOS software for biomolecular simulation in regard to the calculation of free energy differences based on MD simulations with modified Hamiltonians was described. The simulations can be performed under canonical conditions, NVT, to obtain free energy differences ΔF, or under isothermic-isobaric conditions, NpT, to obtain free enthalpy differences ΔG. Four methods are implemented into GROMOS for this purpose, i.e., thermodynamic integration (TI), local elevation umbrella sampling (LEUS), one-step perturbation (OSP), and enveloping distribution sampling (EDS). The sampling in TI or OSP simulations can be enhanced by combining it with Hamiltonian replica exchange (TI-H-RE). The methods TI, TI-H-RE, EDS, and OSP are best suited to estimate the relative free energy of an alchemical perturbation, whereas LEUS and TI in combination with hidden restraints are applicable for estimating the free energy or energy difference between conformational subspaces of a system. In the case of alchemical perturbations, the computational demand varies greatly between the three methods. TI requires the specification of a pathway, along which the system is driven from one end state to the other.

A sufficient number of λ values has to be chosen along the pathway and for each λ value a separate simulation has to be performed. OSP is computationally the cheapest method as only a single simulation needs to be carried out, but it is also most limited, because the phase space regions of the reference state and the end state need sufficient overlap. EDS requires in principle a single simulation similar to OSP. Although in OSP the user has to choose the reference state parameters such that this overlap is guaranteed, in EDS the reference state parameters are optimized in a first simulation, which is then followed by a sampling simulation with fixed EDS parameters. The advantage of EDS compared to TI is that the reference state is defined through the end states and no λ dependence has to be specified by the user. The GROMOS software allows the use of the thermodynamically calibrated GROMOS force fields in free energy calculations,[34,38,39] and offers the user a variety of statistical-mechanics based methods to obtain free energy differences from configurational ensemble averages for (bio)molecular systems of practical interest.

AUTHOR INFORMATION

Corresponding Author

*E-mail: wfvgn@igc.phys.chem.ethz.ch.

Present Addresses

[§]Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG, 88397 Biberach, Germany.

ACKNOWLEDGMENT

Contributions from Markus Christen, Jozef Hritz, Zhixiong Lin, and Niels Hansen are gratefully acknowledged. This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology, by Grant No. 200020-121913 of the Swiss National Science Foundation, and by Grant No. 228076 of the European Research Council, which is gratefully acknowledged.

REFERENCES

(1) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.

(2) Sugita, Y.; Kitao, A; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.

(3) Woods, C. J.; Essex, J. W. *J. Phys. Chem. B* **2003**, *107*, 13703–13710.

(4) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(5) Liu, H.; Mark, A. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1996**, *100*, 9485–9494.

(6) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Phys.* **2007**, *126*, 184110.

(7) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Phys.* **2008**, *128*, 174112.

(8) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2009**, *5*, 276–286.

(9) Christ, C. D.; van Gunsteren, W. F. *J. Comput. Chem.* **2009**, *30*, 1664–1679.

(10) Hansen, H. S.; Hünenberger, P. H. *J. Comput. Chem.* **2010**, *31*, 1–23.

(11) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput. Aided Mol. Des.* **1994**, *8*, 695–708.

(12) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.

(13) Valleau, J. P.; Torrie, G. M. In *Modern Theoretical Chemistry*; Berne, B. J., Ed.; Plenum Press: New York, 1977; Vol. 5, pp 169−194.

(14) Christen, M.; Kunz, A.-P. E.; van Gunsteren, W. F. *J. Phys. Chem. B* **2006**, *110*, 8488–8498.

(15) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2010**, *31*, 1569–1582.

(16) Eichenberger, A. P.; Allison, J. R.; Dolenc, J.; Geerke, D. P.; Horta, B. A. C.; Meier, K.; Oostenbrink, C.; Schmid, N.; Steiner, D.; Wang, D.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2011**, DOI: 10.1021/ct2003622.

(17) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2011**submitted.

(18) Schmid, N.; Allison, J. R.; Dolenc, J.; Eichenberger, A. P.; Kunz, A.-P. E.; van Gunsteren, W. F. *J. Biomol. NMR* **2011**, DOI: 10.1007/s10858-011-9534-0.

(19) Kunz, A.-P. E.; Allison, J. R.; Geerke, D. P.; Horta, B. A. C.; Hünenberger, P. H.; Riniker, S.; Schmid, N.; van Gunsteren, W. F. *J. Comput. Chem.* **2011**in press.

(20) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf Hochschulverlag AG an der ETH Zürich: Zürich, 1996.

(21) 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.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.

(22) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719–1751.

(23) Riniker, S.; Kunz, A.-P. E.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2011**, *7*, 1469–1475.

(24) Hritz, J.; Oostenbrink, C. *J. Chem. Phys.* **2008**, *128*, 14421.

(25) Christen, M.; Christ, C. D.; van Gunsteren, W. F. *ChemPhysChem* **2007**, *8*, 1557–1564.

(26) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(27) Riniker, S.; Christ, C. D.; Hansen, N.; Mark, A. E.; Nair, P. C.; van Gunsteren, W. F. *J. Chem. Phys.* **2011**, *135*, 024105.

(28) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science* **1989**, *244*, 1069–1072.

(29) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487–1493.

(30) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 103–118.

(31) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 119–136.

(32) Hritz, J.; Läppchen, T.; Oostenbrink, C. *Eur. Biophys. J.* **2010**, *39*, 1573–1580.

(33) Steiner, D.; Oostenbrink, C.; Diederich, F.; Zürcher, M.; van Gunsteren, W. F. *J. Comput. Chem.* **2011**, *32*, 1801–1812.

(34) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(35) Prade, L.; Jones, A. F.; Boss, C.; Richard-Bildstein, S.; Meyer, S.; Binkert, C.; Bur, D. *Biol. Chem.* **2005**, *280*, 23837–23843.

(36) Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400–1412.

(37) Lin, Z.; Liu, H.; van Gunsteren, W. F. *J. Comput. Chem.* **2010**, *31*, 2419–2427.

(38) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–1218.

(39) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. *Eur. Biophys. J.* **2011**, *40*, 843–856.

(40) Lee, C. L.; Drinkwater, N.; Tyndall, J. D.; Grunewald, G. L.; Wu, Q.; McLeish, M. J.; Martin, J. L. *J. Med. Chem.* **2007**, *50*, 4845–4853.