Lessons Learned from the Calculation of One-Dimensional Potentials of Mean Force [Article v1.0]

Daniel Markthaler[1], **Sven Jakobtorweihen**[2], **Niels Hansen**[1*]

[1]Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, D-70569 Stuttgart, Germany; [2]Institute of Thermal Separation Processes, Hamburg University of Technology, D-21073 Hamburg, Germany

Abstract

The origins of different computational artifacts that may occur in the calculation of one-dimensional potentials of mean force (PMF) via umbrella sampling molecular dynamics simulations and manifest as free energy offset between bulk solvent regions are investigated. By systematic studies, three distinct causes are elucidated: (i) an unfortunate choice of reference points for the umbrella distance restraint; (ii) a misfit in probability distributions between bound and unbound umbrella windows in case of multiple binding modes; (iii) artifacts introduced by the free energy estimator. Starting with a fully symmetric model system consisting of methane binding to a cylindrical host, complexity is increased through the introduction of dipolar interactions between the host and the solvent, the host and the guest molecule or between all involved species, respectively. The manifestation of artifacts is illustrated and their origin and prevention is discussed. Finally, the consequences for the calculation of standard binding free enthalpies is illustrated using the complexation of primary alcohols with α-cyclodextrin as an example.

*For correspondence:
hansen@itt.uni-stuttgart.de (NH)

1 Introduction

The field of *in silico* pharmaceutical drug design impressively demonstrates the potential of state-of-the-art free energy molecular dynamics (MD) simulations [1]. However, despite a sound theoretical basis [2, 3] and the emergence of best practices [4], reliable predictions of the standard binding free energy or rather free enthalpy [5] or Gibbs energy [6], for realistic host-guest systems from computer simulations are still far from routine. As revealed by different case studies,

the discrepancy between computed and experimental estimates of the standard binding free enthalpy is often beyond the threshold of 4.2 kJ mol^{-1}, commonly referred to as chemical accuracy [7]. Such deviations may arise from three main simulation-related sources: (i) the force-field problem [8, 9], (ii) the sampling problem [10] and (iii) the choice of the free-energy estimator [11]. In addition, experimental uncertainties also have to be considered [12] as well as incompatible thermodynamic state points [13], artifacts caused by the simulation method itself or an inappropriate use of it [14]. The

present article covers two of these issues - the sampling problem and methodological artifacts.

In general, two different strategies can be utilized to compute the binding free enthalpy, related either to alchemical double decoupling, or to physical pathway methods such as potential of mean force (PMF) computations [15]. The latter class of methods requires an integration of the PMF over a bound and unbound region, corresponding to the reversible work to transfer the ligand (or guest molecule) from the bulk to the binding pose inside the host. In principle, the PMF-derived estimate of the binding free enthalpy can be validated by results from double decoupling [16, 17] or, when possible, by direct counting estimation, based on long unbiased simulations [18, 19]. PMF calculations for a specific binding process, are based upon either equilibrium methods such as umbrella sampling [20, 21], local elevation [22] or metadynamics [23], adaptive biasing force [24], forward flux sampling [25] or on non-equilibrium methods such as steered MD [26]. In this article, we focus on one-dimensional PMFs obtained via umbrella sampling simulations for host molecules featuring a distinct hydrophobic cavity. Examples for these types of hosts range from rather low molecular-weight substances such as cyclodextrins [27] or cucurbiturils [28, 29], up to large moieties such as micelles [30] or protein channels inside a membrane [31–34]. The cavity, enabling a ligand to be bound with high affinity and specificity, makes such host molecules attractive for applications in (computer-aided) drug design. However, it also poses various challenges regarding the applied simulation protocol. Studies of cucurbituril complexes revealed that thermodynamic irreversibilities can occur when certain guest atoms, that are not directly controlled via the bias potential, become stuck inside the host and then suddenly jump outside [28, 29]. It was concluded that these dissipative conformational jumps might be a fundamental problem when applying steered MD but also umbrella sampling with fixed spring attachment points to flexible molecules. In typical applications of restrained MD simulations to such molecular systems, a one-dimensional PMF is evaluated by pulling or restraining the ligand along some (linear) path from the bulk at one side of the simulation box through the host to the bulk at the other side. Depending on the complexity of the system, it can be necessary to reduce the sampled space and thus to accelerate convergence by using auxiliary restraints in the simulation setup. The concrete choice of these auxiliary restraints is however non-trivial, since a rigorous way to estimate their effect on the calculated binding free enthalpy is required in order to remove it afterwards [3, 35]. For vanishing interactions at large distances between the binding partners, the PMF becomes flat and approaches a constant value. The fact that this constant has to be the same for every ligand position within the bulk region (due to the isotropy

of the bulk fluid in the absence of external potentials), can be used as a diagnostic test. In a couple of published examples [31–33, 36–39], an artificial offset is visible in the free energy profile between the two bulk regions of the solvent which violates the state function property of the free energy. In some of the cases, this offset was interpreted as indication of insufficient simulation time [38]. However, systematic studies about the origins of these artifacts are scarce [29]. Hub et al. [33] considered solute permeation across a protein channel and found that limited sampling inside the channel in the presence of locally different correlation times can lead to PMF offsets up to 15 kJ mol^{-1}. In Ref. 31, the sampling problem was interpreted as a very small average force across the channel due to the accumulation of noise, originating from all degrees of freedom other than the chosen order parameter. To remedy this problem, the authors followed similar routes: Ref. 31 proposed a symmetrization procedure by creating duplicate umbrella windows on opposite sides of the channel, while Ref. 33 implemented a modified version of the weighted histogram analysis method (WHAM), featuring an additional constraint to enforce periodicity. While such pragmatic solutions may suppress the occurrence of PMF offsets, they do not solve the underlying sampling problem itself. The latter can be solved however using sampling times in excess of microseconds combined with a systematic variation of initial conformations [40] or enhanced sampling methods [41].

The purpose of the present contribution is to demonstrate that computational artifacts may easily occur on much simpler systems for which advanced sampling techniques are not necessarily applied. We elucidate various causes for PMF offsets and relate them to properties of the host-guest system and the applied simulation protocol. The difficulty for setting up free energy molecular dynamics simulations decreased a lot over the last decades, allowing also less experienced users to obtain binding free enthalpy estimates for realistic biomolecular systems. The critical assessment of the results including inspection of convergence and artifacts will always require advanced experience and knowledge, however. With the systematic discussion of the reported artifacts, we aim to sensitize especially newcomers and non-experts in the field in order to prevent time-consuming pitfalls in the context of binding free enthalpy calculations.

2 Theory

Although the main goal is to discuss PMF artifacts, for better evaluation of the results, it is advisable to calculate binding free enthalpies ($\Delta G^{\circ}_{\text{bind}}$) from the PMFs. More important, when using additional restraints on the ligand, the PMF depends on the details of these restraints but $\Delta G^{\circ}_{\text{bind}}$ can be calculated by taking into account the specific restraints. In
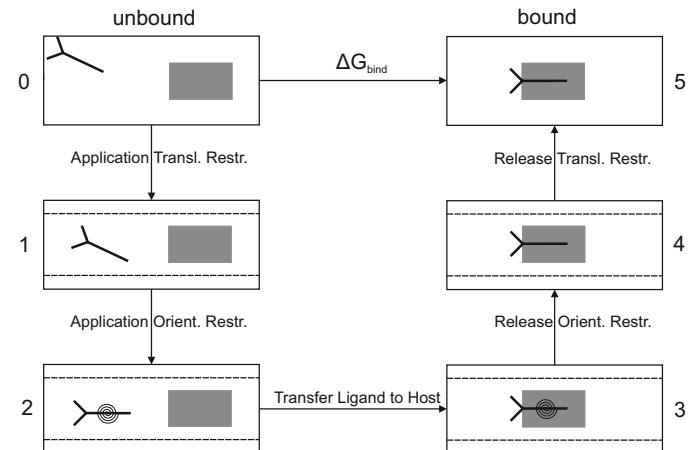
other words, $\Delta G^\circ_{\text{bind}}$ should be independent of the concrete choice of restraints. Furthermore, $\Delta G^\circ_{\text{bind}}$ is directly related to the binding equilibrium constant and as such enables validation against experimentally determined equilibrium constants [16]. A detailed derivation how to calculate $\Delta G^\circ_{\text{bind}}$ from the PMF is beyond the scope of the current work. Therefore, we will just outline the central ideas and present the final expressions. For rigorous derivations, we refer to Refs. 15 and 42. The link between an appropriately defined PMF and $\Delta G^\circ_{\text{bind}}$ can be formulated as the ratio of two configurational integrals over a bound (b) and unbound (u) region:

$$\Delta G^\circ_{\text{bind}} = -RT \ln \left(\frac{\int_b e^{-W(\mathbf{r}_{\text{HL}}, \boldsymbol{\omega}_{\text{HL}})/RT} |\mathbf{J}| \, d\mathbf{r}_{\text{HL}} d\boldsymbol{\omega}_{\text{HL}}}{\int_u e^{-W(\mathbf{r}_{\text{HL}}, \boldsymbol{\omega}_{\text{HL}})/RT} |\mathbf{J}| \, d\mathbf{r}_{\text{HL}} d\boldsymbol{\omega}_{\text{HL}}} \right)$$
$$-RT \ln \left(\frac{V_u}{V^\circ} \right) \quad (1)$$

where $V^\circ = 1.661$ nm^3 is the standard state volume and RT is the thermal energy. The PMF W appearing in the Boltzmann factor, originally depends on the relative separation vector $\mathbf{r}_{\text{HL}}$ and the relative orientation vector $\boldsymbol{\omega}_{\text{HL}}$ between host and ligand. In particular, the PMF does not depend on the external degrees of freedom of the complex corresponding to the absolute position and the overall orientation inside the simulation box. Depending upon the choice of coordinates, a Jacobian determinant $|\mathbf{J}| = |\mathbf{J}(\mathbf{r}_{\text{HL}}, \boldsymbol{\omega}_{\text{HL}})|$ may arise in the configurational integrals of Eq. (1). The second term accounts for the free energy contribution of the volume change from the standard state volume V° to the unbound volume V_u. It should be noted that V_u, which depends on the size of the simulation box cancels from the final expression for $\Delta G^\circ_{\text{bind}}$ [43]. At this point we want to emphasize the difference between a PMF and a free-energy curve (FEC). While these terms are often used synonymously in the literature, the FEC contains the Jacobian contribution, while the PMF does not. If, for example, umbrella sampling is applied to two non-interacting particles using the radial separation r as umbrella coordinate, the FEC decreases with $-2RT \ln r$, while the PMF becomes flat. For the one-dimensional setup as used in the present work (c.f. Sec. 3.2), the Jacobian contribution is equal to unity such that we mostly use the term PMF unless we refer to a three-dimensional calculation setup.

Due to the complexity of the systems, it is often necessary to use auxiliary restraints in the simulation setup. The effect of such additional restraints that limit the phase space to be sampled during the transfer of the ligand from the standard state volume to the binding pose of the host, can be incorporated by introduction of intermediate states into Eq. (1) [35]. The approach can be visualized in the form of a thermodynamic cycle as depicted in Fig. 1.

Application to the case of a PMF along a one-dimensional order parameter (ζ) including auxiliary translational and ori-

Figure 1. Thermodynamic cycle for the calculation of the binding free enthalpy ΔG_{bind}. The host and ligand molecule are represented by the grey rectangle and black structure, respectively. The free enthalpy difference between the unbound (point 0) and bound (point 5) state is given by ΔG_{bind}. When the volume in point 0 is given by the standard state volume V°, ΔG_{bind} corresponds to the standard binding free enthalpy $\Delta G^\circ_{\text{bind}}$. Due to the path-independence of ΔG_{bind}, it can be calculated not only from the direct path $0 \to 5$ as accessed experimentally but equivalently from an indirect path such as $0 \to 1 \to 2 \to 3 \to 4 \to 5$ as accessed from molecular simulations, including several intermediate states (see main text).

entational restraints on the ligand (c.f. Sec. 3.2), finally yields the following expression for $\Delta G^\circ_{\text{bind}}$ [16, 44]:

$$\Delta G^\circ_{\text{bind}} = \Delta G_V + \Delta G_\Omega + \Delta W_R + \Delta G_\theta + \Delta G_\rho \quad (2)$$

with terms ordered according to the cycle in Fig. 1:

$$\Delta G_V = -RT \ln \left(\frac{I_b A_{u,\rho}}{V^\circ} \right)$$

$$\Delta G_\Omega = -RT \ln \left(\frac{\Omega}{8\pi^2} \right)$$

$$\Delta G_\theta = RT \ln \left(\langle e^{-U_\theta(\theta)/RT} \rangle_{b, k_\theta = 0} \right)$$

$$\Delta G_\rho = RT \ln \left(\langle e^{-U_\rho(\rho)/RT} \rangle_{b, k_\rho = 0} \right)$$

ΔW_R denotes the thermally averaged depth of the one-dimensional PMF, i.e.

$$\Delta W_R = RT \ln \left(\frac{\int_u e^{-W_R(\zeta)/RT} d\zeta}{\int_u d\zeta} \right) \quad (3)$$

If the PMF is constant in the unbound region and the global PMF minimum is defined to be zero as it was done in this work, ΔW_R corresponds to the negative PMF value in the bulk ($W_{R,\infty}$): $\Delta W_R = -W_{R,\infty}$. The index "R" indicates that the PMF was evaluated in the presence of auxiliary restraints such as the orthogonal translational and orientational restraints (c.f. Sec. 3.2). That is, ΔW_R represents the step $2 \to 3$ in Fig. 1. The integration of the PMF over the bound region, is captured in the definition of the bound length I_b:

$$I_b = \int_b e^{-W_R(\zeta)/RT} d\zeta \quad (4)$$

Due to the Boltzmann-weighting in Eq. (4), the lowest values of $W_R(\zeta)$ contribute the most to the integral while larger values at increasing distances from the minimum have smaller weights. This makes the estimate of l_b and thus ΔG_{bind}° insensitive to the actual choice of the cut-off distance between bound and unbound region, in particular for rather tight binding situations as studied in the present work [2]. Here, the entire range of ζ-values between the flat parts of the PMF on both sides relative to the minimum were considered as the bound region. The free energy contribution due to the volume change from the standard state volume V° to $l_b A_{u,\rho}$ is described by the term ΔG_V in Eq. (2), corresponding to the step $0 \to 1$ in Fig. 1. Therein, $A_{u,\rho}$ denotes the cross-sectional area which is accessible to the unbound ligand in orthogonal directions in the presence of the applied translational restraint. Its value can be calculated analytically from the partition function of the restraining potential $U_\rho(\rho)$ used for restricting the lateral movement of the ligand in the bulk solvent [16, 44]:

$$A_{u,\rho} = \int_0^\infty e^{-U_\rho(\rho)/RT}\, 2\pi\rho\, d\rho \qquad (5)$$

where ρ is the orthogonal distance (c.f. Fig. 3). The term ΔG_ρ accounts for the free energy contribution of releasing the orthogonal translational restraint in the bound state ($4 \to 5$ in Fig. 1). As indicated by the notation $\langle ... \rangle_{b,k_\rho=0}$, this contribution may be evaluated numerically by free energy perturbation [45] using exponential averaging from an additional simulation with the bound ligand at vanishing restraining force constant $k_\rho = 0$ [44]. A more sophisticated way would be to perform the estimation within multiple simulations of decreasing values of k_ρ using thermodynamic integration [46] or the Multistate Bennett's Acceptance Ratio (MBAR) estimator [47]. The terms ΔG_Ω and ΔG_θ in Eq. (2) assess the free energy contributions from applying and releasing the angular restraint U_θ in the unbound and bound state, respectively ($1 \to 2$ and $3 \to 4$ in Fig. 1). Ω denotes the rotational volume available to the ligand in the bulk under the influence of the angular restraint. For a given functional form of the restraining potential U_θ, its value can be evaluated from a three-dimensional integral over the Euler angles [48, 49]:

$$\Omega = \int_0^{2\pi}\int_0^{2\pi}\int_0^\pi e^{-U_\theta(\theta)/RT}\, \sin\theta\, d\theta d\phi d\psi = 4\pi^2 \int_0^\pi e^{-U_\theta(\theta)/RT}\, \sin\theta\, d\theta$$
$$(6)$$

Depending on the functional form of U_θ, this integral may be solved analytically or numerically. If no angular restraint was applied, allowing the ligand to rotate freely in the bulk, Ω equals $8\pi^2$ and Eq. (2) becomes identical to Eq. (11) of Ref. 44. Therein, it is assumed that the change in rotational entropy

for the bound ligand is included in ΔW_R. In the following we will study situations in which this assumption does not hold. The free energy term ΔG_θ has to be evaluated numerically by free energy perturbation or thermodynamic integration for example.

Application of standard error propagation rules to all involved quantities associated with an uncertainty in Eq. (2) gives:

$$\sigma^2\{\Delta G_{bind}^\circ\} = \left(\frac{RT}{l_b}\right)^2 \sigma^2\{l_b\} + \sigma^2\{\Delta G_\Omega\} + \sigma^2\{\Delta W_R\}$$
$$+ \sigma^2\{\Delta G_\theta\} + \sigma^2\{\Delta G_\rho\} \qquad (7)$$

where $\sigma^2\{...\}$ denotes the variance. The uncertainty in ΔG_Ω arises from the numerical integration error associated with the applied quadrature scheme and is unnecessary if an analytical calculation is possible. Considering only the leading term in Eq. (7) which is given by $\sigma^2\{\Delta W_R\}$ and delivered by the applied estimator (c.f. Sec. 3.4) yields the following simplified expression for the uncertainty estimate of ΔG_{bind}°:

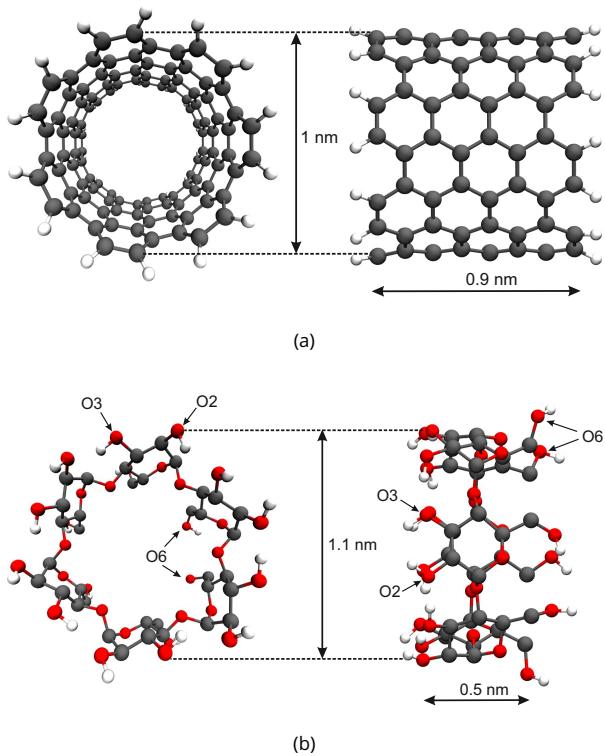$$\sigma^2\{\Delta G_{bind}^\circ\} \approx \sigma^2\{\Delta W_R\} \qquad (8)$$

3 Methods

3.1 Host-Guest Systems

The search for suitable host-guest benchmarks which are simple enough to approach accurately by MD simulations within reasonable time scales yet complex enough to feature properties of protein-ligand systems is an ongoing and non-trivial problem [50, 51]. The majority of simulations from the current work were based on a short carbon nanotube (CNT) host without partial charges (c.f. Fig. 2). This model system, featuring a hydrophobic, water-free cavity resembles the situation of an ideally symmetric and unpolar host molecule. On the other hand, it allows the effect of molecular "asymmetries" to be studied systematically. Here, such an asymmetry was introduced by distributing charge pairs on the terminal C-H atoms at one side of the CNT. The investigated ligands comprised united-atom models for methane, (elongated) ethane and hexane. The effect of dipolar ligands was modeled by placing a positive and neutralizing negative charge onto covalently bound neighboring carbon atoms in case of polyatomic ligands. To test the validity and transferability of the protocol in case of more realistic systems, it was applied to α-cyclodextrin (αCD, c.f. Fig. 2) as a host molecule of practical relevance, complexed with primary alcohols.

3.2 Simulation Protocol

The PMFs were constructed from the time series of a single order parameter sampled via umbrella sampling [20, 21], similar to the approach proposed by Doudou et al. [44]. As

(a)

(b)

Figure 2. Carbon nanotube (CNT) model host (a) and α-cyclodextrin (αCD) molecule (b). The CNT is a (7,7) tube in armchair structure. Nomenclature of αCD-oxygen types according to Ref. [52]. Host dimensions are depicted in front (left) and side view (right).

illustrated in Fig. 3, the order parameter (ζ) used primarily in this work is given by the projection of the instantaneous separation vector ($\boldsymbol{r}_{HL}$) between the centers of mass (COM) of the binding partners onto the host's instantaneous symmetry axis ($\boldsymbol{\omega}_H$): $\zeta \equiv \boldsymbol{r}_{HL} \cdot \boldsymbol{\omega}_H = r_{HL} \cos\varphi$. Here, the unity vector $\boldsymbol{\omega}_H$ was defined by the connecting line through the geometric centers at both sides of the CNT. φ denotes the angle between $\boldsymbol{\omega}_H$ and $\boldsymbol{r}_{HL}$. Instead of the centers of mass, two other characteristic reference points of the host and ligand could be used instead (c.f. Sec. 4.2). The usage of the COM-COM radial distance ($r_{HL} = |\boldsymbol{r}_{HL}|$) itself as order parameter would lead to artifacts around $r_{HL} = 0$ [53]. It would further require to remove the Jacobian contribution from the free-energy profile in order to obtain the PMF [54], as discussed above. Such a Jacobian term, which is of pure entropic nature and accounts for the increase in the accessible configurational area at increasing distances, does not arise when ζ is used instead [44]. Lateral movement of the ligand at every umbrella window was restricted with the aid of a flat-bottom potential acting on the orthogonal displacement ($\rho = r_{HL} \sin\varphi$, c.f. Fig. 3) of

the ligand's COM from the host's molecular axis:

$$U_\rho(\rho) = \begin{cases} k_\rho(\rho - \rho_{up})^n, & \text{if } \rho > \rho_{up} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

The flat-bottom potential (harmonic ($n = 2$) or quartic ($n = 4$), force constant k_ρ) is activated only when the actual displacement exceeds a certain threshold ρ_{up}. In this case, calculation of $A_{u,\rho}$ according to Eq. (5) yields [16]:

$$A_{u,\rho} = \pi\rho_{up}^2 + \begin{cases} \frac{2\pi}{k_\rho^*} + \pi\rho_{up}\frac{(2\pi)^{1/2}}{k_\rho^{*1/2}}, & \text{if } n = 2 \\ \frac{\pi^{3/2}}{2\,k_\rho^{*1/2}} + \pi\rho_{up}\frac{\Gamma(1/4)}{2\,k_\rho^{*1/4}}, & \text{if } n = 4 \end{cases} \quad (10)$$

with the reduced restraining force constant $k_\rho^* \equiv k_\rho/RT$ and the Gamma function Γ. If the threshold value ρ_{up} for the flat-bottom potential is chosen to be large enough compared to the size of the host's cavity such that the ligand's dynamic is not affected in the bound state, the term ΔG_ρ in Eq. (2) makes no contribution. It should be stressed that while the PMF itself and the terms $\Delta W_R, \Delta G_V, \Delta G_\rho$ in Eq. (2) are influenced by the restraining parameters n, k_ρ and ρ_{up}, the estimate for ΔG_{bind}° should be independent when all contributions are evaluated adequately (c.f. Sec. 4.1). Major modifications compared to the original approach of Doudou et al. [44] can be summarized as follows: (i) the order parameters used for both the actual PMF calculation and for measuring the ligand's orthogonal movement are defined in a relative manner between ligand and host. Instead of using a particular Cartesian component such as the z-component of the COM-COM separation vector $\boldsymbol{r}_{HL}$ with respect to an arbitrary external laboratory coordinate system, we look at projections of $\boldsymbol{r}_{HL}$ onto axes of a body-fixed coordinate system which is centered inside the host. The usage of relative order parameters relaxes the requirement of a translationally and/or rotationally restrained host and allows the same approach to be used in case of a fully mobile host molecule without further modifications (c.f. Sec. 4.1); (ii) for the majority of ligands, an additional angular or orientational restraint in the form of a harmonic potential

$$U_\theta(\theta) = \frac{k_\theta}{2}(\theta - \theta_0)^2 \quad (11)$$

was applied, acting on the angle (θ) between the molecular axes of host ($\boldsymbol{\omega}_H$) and ligand ($\boldsymbol{\omega}_L$) in order to suppress flipping of the ligand relative to the host. The molecular axis of the ligand, expressed as unity vector $\boldsymbol{\omega}_L$, was defined by the connecting line through two peripheric atoms of the ligand. The value of k_θ should be chosen high enough to prevent transitions between different ligand orientations. As in case of the translational restraint, the estimate for ΔG_{bind}° should be independent of the concrete choice of k_θ. In case of a translationally and rotationally restrained host aligned along the z-axis without orientational restraint on the ligand, the approach corresponds to the original setup described in Ref. [44].
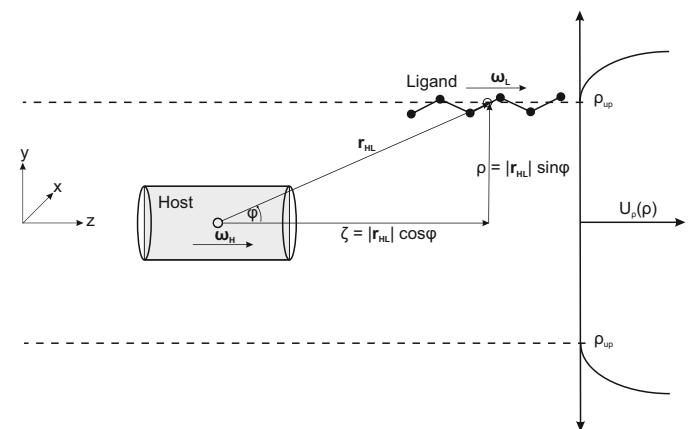
In this case, the order parameter ζ corresponds to the Cartesian z-component of $\mathbf{r}_{HL}$ and ρ becomes $\rho = \sqrt{\Delta x^2 + \Delta y^2}$. Here, Δx and Δy denote the orthogonal displacements of the ligand's COM from the central z-axis.

The free energy contributions corresponding to the release of the translational and orientational restraint in the bound state (ΔG_ρ and ΔG_θ in Eq. (2)) were each calculated from a sequence of 20 simulations with the bound ligand located at the PMF minimum. The individual simulations were conducted at different scaled force constants $k_\rho(\lambda) = \lambda \cdot k_\rho$ with the scaling parameter λ equally distributed between 0 and 1 (analogously for k_θ). The endpoints correspond to the unrestrained case at $\lambda = 0$ and the actual force constant value as used for umbrella sampling at $\lambda = 1$, respectively. Using the configurations sampled from a particular state λ_i, all possible pairwise potential energy differences $\Delta_{ij}U = U_\rho(\lambda_j) - U_\rho(\lambda_i)$ towards the reference state potential $U_\rho(\lambda_i)$ were evaluated (analogously for U_θ). From these potential energy differences, the free energy calculation was performed via the MBAR estimator. For enhanced sampling, Hamiltonian Replica Exchange between neighboring λ-points was applied with attempted exchanges every 1000 steps.

Initial configurations for the production simulations of each umbrella window were generated within a prior equilibration phase (500 ps per window) in the following manner: starting in the bulk at one side of the CNT, the ligand was sequentially displaced in 0.1 nm increments along a linear path through the cavity, until the unbound ligand was located in the bulk again, but relative to the other side of the CNT. For production, all considered systems were simulated at least for 20 ns per window until converged PMF estimates were obtained. Specifications regarding the applied restraints in the protocol are summarized in Tab. 1.

3.3 Simulation Code and Parameters

The GROMOS biomolecular force field was applied throughout this work using the 54A7 [55] and 53A6$_{GLYC}$ [56] parameter sets for studies based on the CNT and αCD, respectively. The standard atom types 12 and 20 were used to represent the CNT carbon and hydrogen atoms, respectively. All systems were solvated in water based on the three-site simple point charge (SPC) water model [57]. Simulations were conducted under periodic boundary conditions using the leap-frog algorithm [58] for integrating Newton's equations of motion with a time step of 2 fs. The majority of simulations were performed with the GROMACS 2016.4 program package [59–61]. In the light of recent publications reporting on the sensitivity of simulation results on the choice of the pairlist algorithm, the electrostatics treatment, the cut-off scheme or other technical details [62–65], complementary simulations were conducted with the GROMOS11 program

Figure 3. Schematic representation of the host-guest system and the relevant collective variables. The orientation of the ligand ($\boldsymbol{\omega}_L$) may be aligned towards the orientation of the host ($\boldsymbol{\omega}_H$) by the usage of an orientational restraint acting on the angle θ between $\boldsymbol{\omega}_L$ and $\boldsymbol{\omega}_H$ (see main text). φ denotes the angle between $\boldsymbol{\omega}_H$ and the separation vector between the centers of mass of host and ligand ($\mathbf{r}_{HL}$). The chosen order parameter (ζ) is the projection of $\mathbf{r}_{HL}$ onto the host's molecular axis $\boldsymbol{\omega}_H$. When the host itself is aligned along the z-axis of the laboratory coordinate system, as depicted here, the order parameter corresponds to the Cartesian z-component of $\mathbf{r}_{HL}$. The ligand's movement orthogonal to the order parameter outside the host is restricted via a flat-bottom potential ($U_\rho(\rho)$) acting on the orthogonal distance (ρ) between the center of mass of the ligand and the molecular axis of the host. The flat-bottom potential is activated when the actual distance ρ exceeds a certain threshold distance (ρ_{up}), as depicted by the dashed lines.

package (release version 1.5.0) [66–68] which has different recommended settings. In particular, GROMOS is usually used with a reaction field scheme for treating long-range electrostatic interactions. Since this approach is also used by other codes in the context of free energy simulations [69, 70], it is interesting to study the effect on a PMF calculation. In the following, separated computational details are given for the two simulation codes.

GROMACS Simulations
Simulations using the particle-mesh Ewald (PME) method [71, 72] for treating electrostatic interactions were conducted with the GROMACS 2016.4 program [59–61] patched to the free-energy library PLUMED 2.4.2 [73] for restraints definition and biasing selected collective variables. The center of mass translation of the computational box was removed every 1000 steps. All bond lengths were constrained using the LINCS algorithm [74, 75] with a LINCS-order of 4. The number of iterations to correct for rotational lengthening in LINCS was set to 2. SPC water was constrained using the SETTLE algorithm [76]. Equilibration of solvated energy-minimized systems was performed within a prior 100 ps constant-volume simulation at reference tempera-

Table 1. Default values for restraints specifying the umbrella sampling protocol as used for the majority of studies in the current work. In case of differing settings, the parameter choice is explicitly given. For all simulations involving a polar CNT, a value of k_ζ = 3000 kJ mol^{-1} nm^{-2} was used for the distance restraint force constant. Lateral translational movement of the ligand (as measured by the orthogonal displacement ρ) was restrained using a flat-bottom potential (c.f. Eq. (9)). To restrain the ligand's orientation towards a specific bound state orientation, an orientational restraint acting on the angle θ between the molecular axes of host and ligand was applied (c.f. Eq. (11)).

Distance Restraint			
k_ζ [kJ mol^{-1} nm^{-2}]	ζ_{min} [nm]	ζ_{max} [nm]	$\Delta\zeta$ [nm]
500	-2.5	2.5	0.1
Translational Restraint			
k_ρ [kJ mol^{-1} nm^{-n}]	ρ_{up} [nm]	n [–]	
500	0.4	2	
Orientational Restraint			
k_θ [kJ mol^{-1} rad^{-2}]	θ_0 [rad]		
500	0.0		

ture of 300 K, followed by a 1 ns constant-pressure simulation at 300 K and 1 bar for pressure equilibration. Initial velocities were sampled from a Maxwell-Boltzmann distribution at 300 K. During the equilibration phase, both temperature and pressure were controlled by application of the weak coupling scheme [77] with corresponding relaxation times of τ_T = 0.1 ps and τ_P = 0.5 ps and an (isotropic) isothermal compressiblity of 4.5 × 10^{-5} bar^{-1} [78]. For production simulations, the Nosé-Hoover thermostat [79–81] and Parrinello-Rahman barostat [82, 83] were applied with corresponding coupling constants of τ_T = 1.0 ps and τ_P = 2.0 ps. The solute (comprising the host and ligand molecule) and solvent were coupled to separate heat baths. A Verlet-buffered neighbor list [84] which was updated every 25 steps, was applied for the treatment of short-range electrostatic and van der Waals interactions with potentials shifted to zero at 1.4 nm. The latter were modeled by the Lennard-Jones potential. Analytic dispersion corrections were applied for energy and pressure calculation. Long-range electrostatic interactions were treated with the smooth particle-mesh Ewald (PME) method [71, 72] using a real-space cut-off of 1.4 nm with a cubic splines interpolation scheme and a grid spacing of 0.12 nm. In most simulations reported here, the host's orientation was aligned along the z-axis of the simulation box (box dimensions: 3.4 x 3.4 x 12 nm) alongside with a translational restraint (500 kJ mol^{-1} nm^{-2}) to keep its COM close to the box center. The bias on the orientation was realized

by an orientational restraint (500 kJ mol^{-1} rad^{-2}) acting on the angle between the host's symmetry axis and the external z-axis. Biased collective variables were written to file every 100 steps.

GROMOS Simulations

Simulations using the Barker-Watts reaction field (RF) scheme [85] for treating electrostatic interactions were conducted with the GROMOS11 program package (release version 1.5.0) [66–68]. The center of mass translation of the computational box was removed every 1000 steps. All bond lengths including the water hydrogen-hydrogen distances were constrained using the SHAKE algorithm [86] with a relative geometric tolerance of 10^{-4}. Equilibration of solvated energy-minimized systems was performed within a prior 100 ps constant-volume simulation followed by a 1 ns constant-pressure simulation at 300 K and 1 bar for pressure equilibration. During the constant-volume equilibration, temperature was raised by increments of 60 K to the final value of 300 K with initial velocities assigned according to a Maxwell-Boltzmann distribution centered around 60 K. Temperature was maintained close to its reference value by weak coupling [77] to individual external baths for solute and solvent with relaxation times of 0.1 ps. Pressure was held constant at 1 bar by the weak coupling method with a relaxation time of 0.5 ps and an isothermal compressibility of 4.5 × 10^{-5} bar^{-1} [78]. A Barker-Watts RF contribution [85] was applied to account for the long-range electrostatic effect beyond the (long-range) cut-off. The relative dielectric permittivity of the dielectric continuum outside the cut-off sphere was set to ϵ_{RF} = 61, as appropriate for SPC water [87]. In case of van der Waals interactions, no long-range correction was incorporated. Non-bonded interactions were either calculated using a single-range (SR) or a twin-range (TR) cut-off scheme [88]. In case of the TR scheme, interactions within the short-range cut-off radius of 0.8 nm were calculated every time step from a pairlist updated every five steps, while interactions between 0.8 and the long-range cut-off of 1.4 nm were reevaluated for each pairlist update and kept constant in between. In case of the SR scheme using a cut-off radius of 1.4 nm, the pairlist update was performed every time step. On top of the two cut-off schemes, the influence of different construction schemes for the non-bonded pairlist was further investigated, specifying whether the interactions are calculated based on distances between individual atoms (AT) or neutral charge groups (CG). In total, this results in four different non-bonded interaction setups that were tested in conjunction with the RF approach: (i) RF using a twin-range cut-off scheme based on charge groups (RF, TR-CG), (ii) RF using a twin-range and atomistic cut-off scheme (RF, TR-AT), (iii) RF using a single-range cut-off

scheme based on charge groups (RF, SR-CG), (iv) RF using a single and atomistic cut-off scheme (RF, SR-AT). An overview of the systems which were treated with the different RF setups is given in Tab. 2.

Due to different implementations, the restraints handling was different in the GROMOS package compared to analogue simulations conducted with GROMACS/PLUMED (see above) and can be summarized as follows: (i) alignment of the host along the z-axis of the simulation box was realized by four individual position restraints (1000 kJ mol^{-1} nm^{-2}) imposed for two pairs of peripheric C-atoms located at opposing sides of the CNT; (ii) the coordinates ζ and ρ for measuring progression and lateral movement of the ligand, respectively, were defined by Cartesian components of the separation vector between the COM of the ligand and a fixed anchor point on the z-axis instead of using the separation vector between the COM of ligand and host (c.f. Fig. 3). It was verified that the difference in restraining the host's degrees of freedom does not affect the PMF (c.f. Sec. 4.1), while the usage of a translated reference point along the z-axis only shifts the whole PMF by the same offset along the range of ζ-values without affecting its shape or the barrier heights. Biased collective variables were written to file every 100 steps.

3.4 Free Energy Estimation

PMFs were evaluated employing three commonly used free-energy estimators or analysis methods: (i) the Weighted Histogram Analysis Method (WHAM) [89–91], (ii) Umbrella Integration (UI) [92–94] and (iii) the Multistate Bennett's Acceptance Ratio (MBAR) estimator [47]. For WHAM, the GROMACS implementation `g_wham` [33] was used, while in case of UI and MBAR, open source python packages [95, 96] were employed. While each estimator aims to recover a statistically optimal estimate for the unbiased distribution function of the order parameter, differences become apparent from the underlying working equations and the uncertainty estimates. Detailed information regarding these aspects can be found in the specialized literature cited above. Both WHAM and MBAR result in a coupled set of non-linear equations for the free energy estimates which have to be solved iteratively in a self-consistent manner. This is avoided in the UI approach which was the primarily used estimator throughout this work. In UI, the biased distributions are approximated as normal distributions (fully characterized by the mean and variance) and the restraint forces from each window are combined instead of the unbiased distributions itself. As illustrated in Sec. 4.3, the assumption of normal distributions might not be fulfilled for certain conditions depending on the molecular system and simulation protocol. Analytic expressions for PMF uncertainties corresponding to the UI method involve a segment-based analysis (similar to block averaging)

for mean and variance of the sampled biased distributions and follow from repeated application of error propagation as described in detail in Ref. 94. The resulting uncertainty over the interval $[\zeta_a, \zeta_b]$ refers to the 95% confidence interval such that the presented PMFs are reported in the form $\Delta W_R(\zeta_b; \zeta_a) \pm 1.96 \sqrt{\sigma^2 \{\Delta W_R(\zeta_b; \zeta_a)\}}$ [94]. ζ_a denotes the minimal value of the order parameter (left border) and ζ_b some running upper value (right border). In that sense, the error bar represents a cumulative estimate with respect to a chosen reference point (ζ_a), resulting in larger error bars for increasing values of the order parameter $\zeta > \zeta_a$ [94].

4 Results

The results as presented in the following were obtained from systematic series of studies with the objective to analyze the influences of (i) restraining the host's degrees of freedom, (ii) restraining the ligand's degrees of freedom via translational and orientational restraints, (iii) the choice of reference points as used in the restraining setup, (iv) the treatment of electrostatic interactions (PME vs. RF) and (v) the free energy estimator. Issue (iv) also includes influences of the used cut-off scheme (SR vs. TR) as well as the underlying pairlist-generation scheme (AT vs. CG) in case of simulations based on the RF approach (c.f. Sec. 3.3). Except for the paragraphs considering different approaches for long-range electrostatics, all reported PMFs refer to simulations based on the PME approach. To separate the various influences, we started with united-atom methane binding to the completely symmetric and unpolar CNT host before studying polyatomic unpolar ligands. To investigate issues associated with intrinsically asymmetric systems, complexity was further increased by considering the binding of unpolar as well as dipolar ligands to a CNT with a polar pore mouth at one side. The consequences with respect to the calculation of the standard binding free enthalpy according to Eq. (2) are elucidated. For several cases, the PMF-derived estimates for $\Delta G_{\text{bind}}^{\circ}$ were compared with results from alchemical double decoupling. Details about the double decoupling approach can be found in the appendix. Finally, the application to α-cyclodextrin (αCD) complexed with primary alcohols is presented. Special focus is given to the occurrence of computational artifacts which manifest as flawed PMFs featuring a significant offset between the two flat bulk regions. Specific parameters as used in the umbrella sampling setup are summarized in Tab. 1.

4.1 Unpolar CNT / Methane

This section reports PMFs between united-atom methane and the unpolar CNT. Since it was found that all three estimators (WHAM, UI, MBAR) yield indistinguishable PMFs within error bars, only the UI results will be reported in the following.

Table 2. Overview of simulated systems based on the reaction field treatment for long-range electrostatics using different cut-off schemes (SR, TR) and pairlist generation schemes (CG, AT) as specified in the main text (c.f. Sec. 3.3). All simulations were conducted with the GROMOS MD package. The CNT was aligned along the z-axis of the computational box such that the order parameter ζ corresponds to the z-component of the COM-COM separation vector between the binding partners. For ethane and hexane, no restraint was imposed on the orientation. Labels S1 to S3 refer to different box sizes - S1: 3.4 x 3.4 x 8.0 nm, S2: 4.0 x 4.0 x 8.0 nm, S3: 5.0 x 5.0 x 8.0 nm.
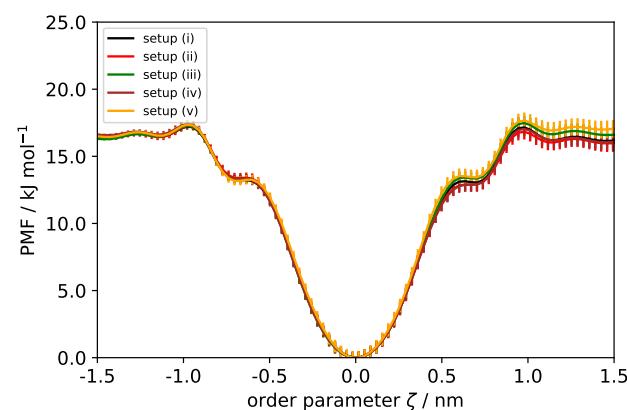
System	SR-AT			TR-AT			SR-CG			TR-CG		
	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
unp. CNT / CH_4	√	-	-	√	√	-	√	-	-	√	√	-
unp. CNT / unp. C_2H_6	√	-	-	√	√	-	√	-	-	√	√	-
unp. CNT / unp. C_6H_{14}	-	-	-	√	-	-	-	-	-	√	√	-
pol. CNT / CH_4	√	√	-	√	√	√	√	√	-	√	√	√
pol. CNT / unp. C_2H_6	√	√	-	√	√	√	√	√	-	√	√	√

Restraining the Host

In the integrals of the PMF expression in Eq. (1), six external degrees of freedom, corresponding to overall rotation and translation of the host-guest complex were integrated out. In practical applications it is often desirable to restrain the position and orientation of the host molecule in order to limit the size of the computational box. Such a position restraint may influence the potential of mean force if conformational fluctuations of the host molecule are suppressed. For the CNT host studied here, we confirmed that the restraints acting on the external degrees of freedom of the host molecule do not influence the PMF. Therefore, five different setups were compared: (i) no external restraints applied for the host, (ii) a three-dimensional position restraint acting on the host's COM to keep it close to the box center, (iii) application of an axial restraint to keep the host aligned along the z-axis, (iv) a three-dimensional position restraint on the host's COM combined with an axial restraint (combination of (ii) and (iii)) and (v) three-dimensional position restraints acting on every host atom. In setup (i), the host-guest complex as a whole can translate and rotate in three dimensions. In setup (ii), the host (and thus the complex as a whole) can not translate, but it can rotate without hindrance. In setup (iii) in contrast, the axial restraint on the host restricts the rotation of the host-guest complex, but it can still translate in three dimensions. The setups (iv) and (v) hamper both, the translational and the rotational movement of the host molecule. Setup (v) even restricts a rotation of the host around its axis which is possible for setup (iv). From the perspective of a moving observer located in the host's COM, all setups are identical as long as the host's internal dynamic is not affected by the external restraints, which is only the case in setup (v). While the setups become more restrictive from (i) to (v), the system size (and thus the computational effort) is increased considerably for setup (i) and (ii), since a uniform simulation box is required in contrast to (iii), (iv) and (v). It should be stressed, that due to the relative formulation of the order parameter (guest relative to the host) and auxiliary quantities such as the angle φ and the orthogonal distance ρ (c.f. Fig. 3), identical restraints specifications between host and guest can be used for all setups without modifications.
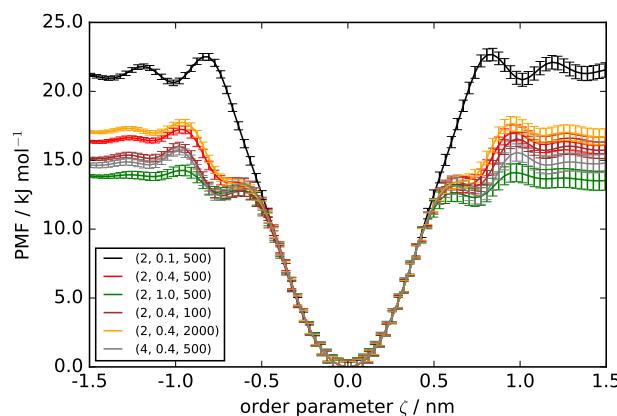
It was found that all five setups yield indistinguishable PMFs within uncertainties (c.f. Fig. 4). The fact that even the very restrictive setup (v) has no effect on the PMF can be probably attributed to the rather rigid structure of the CNT cavity. For other more flexible host molecules, the effect of restraining so many degrees of freedom might be more pronounced and should be avoided as outlined above. We conclude, that the way we restrained the host's external degrees of freedom (overall rotation and translation) does not affect the calculated PMFs, as expected from theory such that the PMF artifacts reported below have a different cause. Unless explicitly stated otherwise, all results presented in the remainder of the article refer to setup (iv).

Figure 4. Effect of different restraining setups (i)-(v) as used to restrain the external degrees of freedom of the CNT host (c.f. Sec. 4.1). PMFs refer to the system methane / unpolar CNT.

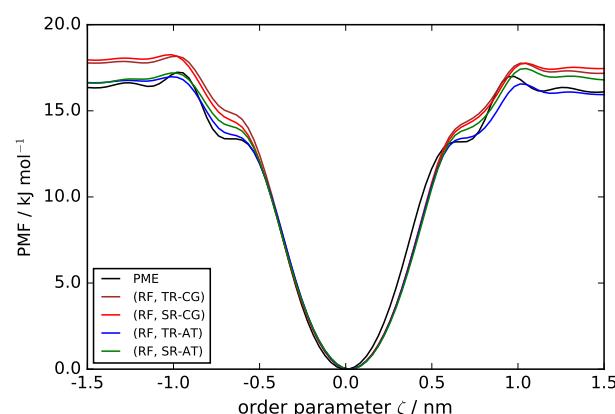### Restraining the Ligand's Lateral Movement

While the flat-bottom potential U_ρ used for limiting the ligand's lateral movement influences the PMF, the final estimate of the standard binding free enthalpy ΔG°_{bind} should be independent. Fig. 5 shows the PMFs for methane / CNT obtained for different restraining parameters in terms of the exponent n, the threshold ρ_{up} and force constant k_ρ (c.f. Eq. (9)). All PMFs show perfect symmetry as expected for such a system with a global minimum at $\zeta = 0.0$ nm, corresponding to configurations where methane is located at the cavity center. Different parameter combinations basically scale the PMFs while the overall shape remains very similar. Here, the usage of smaller threshold parameters (at constant k_ρ) as well as higher force constants (at constant ρ_{up}) leads to higher absolute numbers of ΔW_R. Corresponding estimates of ΔG°_{bind} for every PMF according to Eq. (2) are summarized in Tab. 3. Since no orientational restraint was applied in this case, the terms ΔG_Ω and ΔG_θ in Eq. (2) make no contribution. While the estimates for ΔW_R and $A_{u,\rho}$ are strongly influenced by the parameters of U_ρ and as such also ΔG_V, the bound length l_b is virtually independent. The free energy contribution associated with the orthogonal translational restraint in the bound state (ΔG_ρ) is close to zero due to the naturally restricted conformational space accessible to the bound ligand inside the host's cavity. In accordance with theoretical expectation, all PMFs yield very similar estimates for ΔG°_{bind} independent of the choice of orthogonal restraining parameters (c.f. last column in Tab. 3). In addition, the PMF-based estimates are in reasonable agreement with the value of $\Delta G^\circ_{bind} = -13.0$ kJ mol^{-1} as obtained from alchemical double decoupling (c.f. Tab. A1).

Figure 5. Effect of the orthogonal translational restraint on the PMF for methane / unpolar CNT. Restraining parameters (n, ρ_{up} [nm], k_ρ [kJ mol^{-1} nm^{-n}]) as used for the flat-bottom potential U_ρ according to Eq. (9), are given by number triplets in the legend.

Treatment of Electrostatic Interactions

Fig. 6 shows the influence of different treatments for long-range electrostatics (PME vs. RF) alongside with different cut-off schemes (SR vs. TR) and pairlist generation schemes (CG vs. AT) on the PMF. As can be seen, all setups yield very similar PMFs. In Ref. 97, a system size dependence of the PMF for ion association was observed in case of simulations based on the RF treatment. Therefore, additional simulations using box sizes of different x- and y-dimensions were conducted for the two TR-setups (c.f. Tab. 2) as well as for the PME treatment. In all cases, no system size dependence could be observed (data not shown).

Figure 6. Effect of the treatment of electrostatic interactions (PME vs. RF), the cut-off scheme (SR vs. TR) and the pairlist generation scheme (CG vs. AT) on the PMF for unpolar methane / unpolar CNT (c.f. Sec. 3.3). Bounds for statistical uncertainties are below 1.0 kJ mol^{-1} and have been omitted in the interest of clarity.

Lesson Learned

For the CNT / methane system, consistent PMFs were obtained leading to binding free enthalpies within maximal statistical bounds of ± 1.5 kJ mol^{-1}, regardless of how the host and the ligand's lateral movement was restrained (c.f. Tab. 3). This conservative estimate for the maximal error encompasses the PMF uncertainty as delivered by the UI estimator as well as the spread of ΔG°_{bind} values obtained from the different setups. This distribution of ΔG°_{bind} values also emphasizes however, that even for such a simple system, no perfect agreement can be expected. The treatment of electrostatic interactions and the pairlist generation scheme have an effect on ΔW_R on the order of ± 1.5 kJ mol^{-1}. These results are an important basis to judge the artifact reported in Sec. 4.3.

4.2 Unpolar CNT / Multiatomic Ligand

This section reports the PMFs for the unpolar CNT host complexed with different multiatomic unpolar ligands. Here, rigid
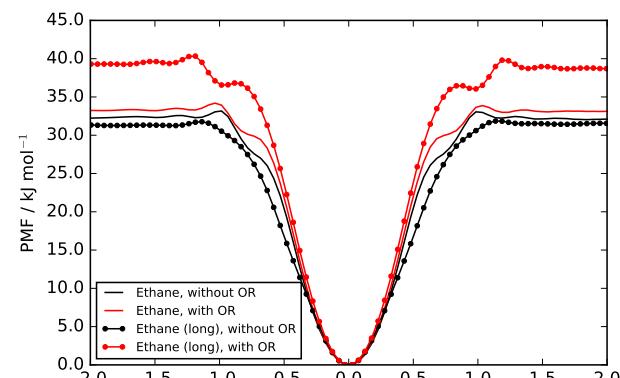
Table 3. Influence of the translational restraint settings on the calculated standard binding free enthalpy ΔG°_{bind} for united-atom methane / unpolar CNT (c.f. Sec 4.1). Corresponding PMFs are depicted in Fig. 5. First three columns specify the parameters used for the flat-bottom potential U_ρ (c.f. Eq. (9)). Calculations of l_b, $A_{u,\rho}$ and ΔG°_{bind} were performed according to Eq. (4), Eq. (10) and Eq. (2), respectively. The contribution of the translational restraint in the bound state ΔG_ρ was calculated using the MBAR estimator from a sequence of simulations in the bound state with force constants k_ρ varying from zero to the final value as given in the table. Error estimates refer to the UI result.

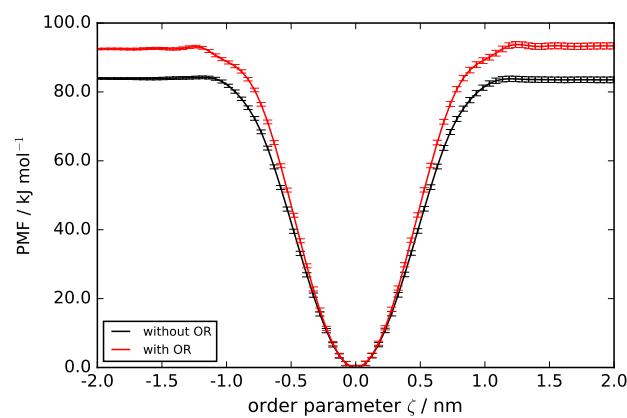| Setup | | | | | | | | |
n	ρ_{up}	k_ρ	ΔW_R	l_b	$A_{u,\rho}$	ΔG_V	ΔG_ρ	ΔG°_{bind}
[–]	[nm]	[kJ mol^{-1} nm^{-n}]	[kJ mol^{-1}]	[nm]	[nm^2]	[kJ mol^{-1}]	[kJ mol^{-1}]	[kJ mol^{-1}]
2	0.1	500	-21.37 ± 0.54	0.3827	0.1184	8.98	-0.06	-12.44 ± 0.54
2	0.4	500	-16.27 ± 0.62	0.3832	0.7565	4.35	-0.13	-12.05 ± 0.62
2	1.0	500	-13.80 ± 0.72	0.3869	3.7291	0.35	-0.14	-13.59 ± 0.72
2	0.4	100	-15.51 ± 0.61	0.3874	1.1569	3.27	-0.50	-12.74 ± 0.61
2	0.4	2000	-16.93 ± 0.60	0.3835	0.6217	4.84	-1.03	-13.12 ± 0.60
4	0.4	500	-14.84 ± 0.66	0.3856	1.3047	2.98	-0.64	-12.50 ± 0.66

diatomic ligands in the form of ethane and a modified model with increased bond length were studied as well as hexane. In contrast to (ordinary) ethane, the elongated variant (in the remainder denoted as "elongated ethane") is unable to rotate inside the CNT cavity once it is bound. This ligand selection enables to study the impact of the ligand's flexibility and rotational degrees of freedom inside the binding pose. Since it was found that all three estimators (WHAM, UI, MBAR) yield indistinguishable PMFs within errors bars, only the UI results will be reported in the following.

Restraining the Ligand's Orientation

In practice it can often be essential to restrain not only the translational movement of the ligand but also its orientation towards the host molecule (c.f. Sec. 4.4). Fig. 7 shows the PMFs for (a) ethane, elongated ethane and (b) hexane, obtained from the setup with (red curves) and without (black curves) orientational restraint. As can be seen, the restraint on the ligand's rotation leads to higher absolute numbers of ΔW_R. Comparison of the diatomic ligands shows that this increase is more pronounced for increasing bond lengths. Tab. 4 contains the calculated estimates for ΔG°_{bind}. For each ligand, two estimates are provided, corresponding to the setup with and without orientational restraint. As revealed by the data, the application of an orientational restraint in case of ethane has a marginal effect on ΔW_R but the free energy contribution from releasing this restraint in the bound state is the highest for all ligands. The fact that this contribution is almost identical for the bound and unbound state shows that the confinement inside the host's cavity has no significant effect on the populated ligand orientations in this case, as expected. For elongated ethane and hexane, which are not able to rotate in the bound state, the free energy gain of releasing the restraint is much smaller. The good agreement of the corresponding values for ΔG°_{bind} from simulations with and without orientational restraint confirms consistency be-

(a)

(b)

Figure 7. Effect of a restrained ligand orientation on the PMFs for (a) unpolar (elongated) ethane / unpolar CNT and (b) unpolar hexane / unpolar CNT. Red and black curves correspond to the setup with and without orientational restraint (OR), respectively. Error bars in graph (a) have been omitted in the interest of clarity.
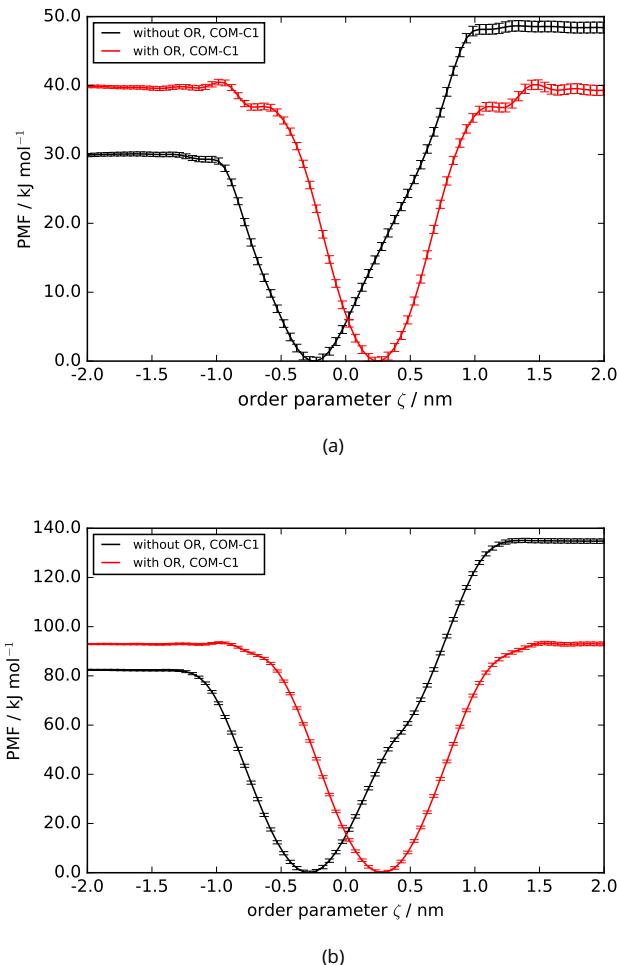
tween the setups. Results from double decoupling which was performed for ethane and elongated ethane, was also found to be in good accordance with the PMF-based estimates (c.f. Tab. A1). We conclude that the effect of an orientational restraint included in the simulation protocol with respect to the calculation of ΔG_{bind}° is captured adequately by the terms ΔG_{Ω} and ΔG_{θ} in Eq. (2). Therefore, a variation of the restraining force constant k_{θ} was not performed in this work.

Choice of Restraining Reference Points

The decision which (pseudo) atoms to choose in the host and ligand molecule to serve as reference or anchor points for the applied distance restraint in the umbrella sampling simulations is often not clear a priori. Though the centers of mass might be an intuitive choice (and were selected for the majority of studies of the current work), other choices might appear more suitable in practical application [44]. Fig. 8 (a) and (b) show the PMFs for elongated ethane / CNT and hexane / CNT, respectively, as obtained when a peripheric carbon atom was picked as reference point in the ligand. The COM of the CNT was chosen as reference point within the host as has been the case hitherto. Every graph contains two free energy profiles, corresponding to the PMF evaluated with (red curves) and without (black curves) orientational restraint. In the setup lacking an orientational restraint, a substantial free energy offset between 20 (elongated ethane) and 55 kJ mol^{-1} (hexane) is present in the PMF. The reason for this offset lies in the differences of the sampled configurational space at the two pore mouths. Depending on which part of the ligand is buried (the part with or without the anchor atom for the distance restraint), the configurational space accessible to the partly bound ligand is quite different. The estimation of ΔG_{bind}° from such a PMF would lead to very different results depending on which branch of the PMF would have been taken as a basis for the analysis. In contrast, if the COM of the ligand is chosen as anchor atom, the rotational behavior of the ligand is symmetric at both CNT ends and no PMF offset is present, even when the ligand's orientation is not restrained (c.f. Fig. 7). However, as demonstrated in Fig. 8, even in case of such an "unfortunate" choice of anchor points, the PMF offset can be eliminated through the usage of an orientational restraint. Comparison of the corresponding profiles of Fig. 8 and Fig. 7 obtained from the setup including an orientational restraint (but different reference points in the ligand), shows that the PMFs are identical except for a marginal shift along the order parameter axis which will not affect the estimate for ΔG_{bind}°.

Treatment of Electrostatic Interactions

Fig. 9 shows the influence of the treatment for long-range electrostatics (PME vs. RF) alongside with the pairlist gener-

(a)

(b)

Figure 8. Effect of the choice of reference points for the distance restraint on the PMFs for (a) unpolar elongated ethane / unpolar CNT and (b) unpolar hexane / unpolar CNT. Here, the COM of the CNT and the C1 carbon atom of the ligand was chosen as reference points. Red and black curves correspond to PMFs obtained from the setup with and without orientational restraint (OR), respectively.

ation scheme (CG vs. AT) on the PMFs for two systems: (a) ethane / CNT and (b) hexane / CNT. In addition, the effect of different cut-off schemes (SR vs. TR) was tested in case of ethane / CNT. None of the cases included an orientational restraint. As can be seen, all setups yield almost indistinguishable PMFs for ethane / CNT, whereas for hexane / CNT, the two RF setups yield a slightly narrower PMF well compared to the PME result. Referring to the effect on the binding free enthalpy, such a different shape only affects the calculation of the bound length l_b (c.f. Eq. (4)) leading to a marginal discrepancy in the order of ± 0.5 kJ mol^{-1} compared to the PME-based estimate. As in case of methane / CNT, no system size dependence for the PMFs could be observed (data not shown).

Table 4. Calculated standard binding free enthalpies $\Delta G^\circ_{\text{bind}}$ for the binding of unpolar ethane, elongated ethane and hexane to the unpolar CNT (c.f. Sec 4.2). Two rows of data are associated with every ligand, corresponding to the setup with and without orientational restraint (OR). Corresponding PMFs are depicted in Fig. 7. Calculations of l_{b}, ΔG_V and ΔG_Ω were performed as described in Sec. 2. The joint contribution of the translational and orientational restraint in the bound state ($\Delta G_\rho + \Delta G_\theta$) was calculated using the MBAR estimator from a sequence of simulations in the bound state with force constants k_ρ and k_θ varying from zero to the final values as specified in Tab. 1. The estimate of $\Delta G^\circ_{\text{bind,Conf.1}}$ as obtained from the setup including an orientational restraint corresponds to one distinct binding configuration and was corrected by a symmetry term of $-RT \ln 2$ [48, 98] to obtain $\Delta G^\circ_{\text{bind}}$ in case of elongated ethane and hexane which are unable to rotate inside the CNT cavity in the absence of an orientational restraint (c.f. Sec. 4.5). Error estimates refer to the UI result.

System	Setup	ΔW_{R} [kJ mol^{-1}]	l_{b} [nm]	ΔG_V [kJ mol^{-1}]	ΔG_Ω [kJ mol^{-1}]	$\Delta G_\rho + \Delta G_\theta$ [kJ mol^{-1}]	$\Delta G^\circ_{\text{bind,Conf.1}}$ [kJ mol^{-1}]	$\Delta G^\circ_{\text{bind}}$ [kJ mol^{-1}]
Ethane	No OR	-32.16 ± 0.87	0.2989	4.97	0.00	0.00	-	-27.19 ± 0.87
	OR	-33.23 ± 0.74	0.2925	5.03	14.95	-14.41	-27.66 ± 0.74	-27.66 ± 0.74
Long Ethane	No OR	-31.38 ± 0.85	0.2864	5.08	0.00	0.00	-	-26.30 ± 0.85
	OR	-38.97 ± 0.81	0.2707	5.22	14.95	-5.65	-24.44 ± 0.81	-26.17 ± 0.81
Hexane	No OR	-83.65 ± 0.89	0.1995	5.98	0.00	0.00	-	-77.67 ± 0.89
	OR	-92.99 ± 0.92	0.1954	6.03	14.95	-4.72	-76.72 ± 0.92	-78.45 ± 0.92

Lesson Learned

For the binding of symmetric unpolar multiatomic ligands to the unpolar CNT, the change in rotational entropy upon binding is included in the PMF if no orientational restraint is used. The two setups (with and without orientational restraint) lead to standard binding free enthalpies which are indistinguishable within statistical uncertainties. An orientational restraint is required however, if the anchor points for the umbrella distance restraint in the ligand and in the CNT are chosen in such a way, that the configurational space accessible to the partly bound ligand at the both cavity entrances are different. The treatment of electrostatic interactions and the pairlist generation scheme have a marginal effect on ΔW_{R} on the order of ± 1 kJ mol^{-1}.

4.3 Polar CNT / Unpolar Ligand

This section reports PMFs for the association of a polar CNT with different unpolar ligands. The polar CNT was modeled by distributing balancing charges to terminal pairs of C-H-atoms at one side of the CNT (C-atoms: -0.5 e, H-atoms: +0.5 e where "e" denotes the elementary charge). Every balancing pair of C-H atoms was assigned to one neutral charge group in case of simulations based on the RF approach for long-range electrostatics in combination with the CG pairlist scheme. In contrast to the unpolar systems treated so far, care has to be taken in order to avoid a bias due to the applied PMF analysis method. This issue is discussed explicitly in a separate subsection.

Impact of the Host's Polarity

Fig. 10 shows PMFs for a set of unpolar ligands (methane, (elongated) ethane, hexane) binding to the polar CNT. No orientational restraint was imposed on the ligand. In contrast to the previous examples corresponding to the binding to an unpolar 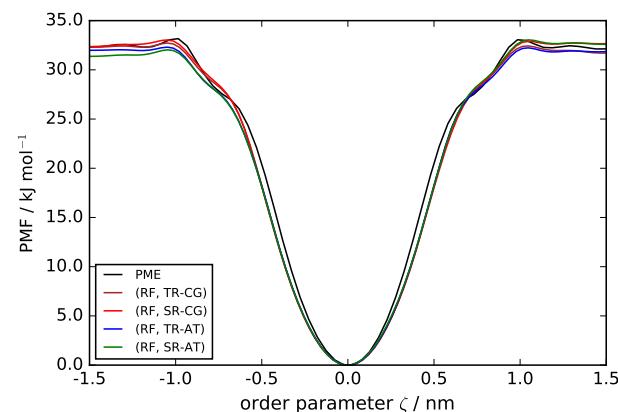CNT, the resulting PMFs are highly asymmetric featuring a considerable barrier to be overcome by the ligand at the polar entrance of the CNT. This barrier which is caused by the modified water structure in proximity to the polar mouth, makes the binding path through that particular side energetically unfavorable.
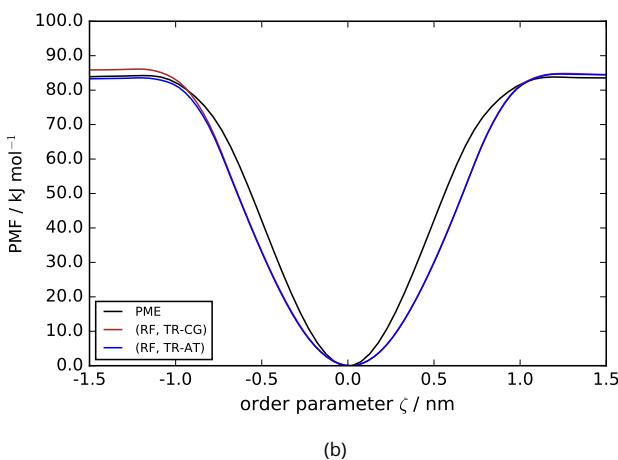
Treatment of Electrostatic Interactions

Fig. 11 shows the influence of the treatment for long-range electrostatics (PME vs. RF) alongside with the pairlist generation scheme (CG vs. AT) and cut-off scheme (SR vs. TR) on the PMFs for two systems: (a) methane / polar CNT and (b) unpolar ethane / polar CNT. In case of ethane, no orientational restraint was applied. As can be seen, PMFs based on an atomistic interaction scheme (TR-AT, SR-AT) show a smaller well (measured by ΔW_{R}) and barrier at the polar entrance compared to analogue simulations based on charge groups (TR-CG, SR-CG). The combination of the RF approach with an atomistic interaction scheme also shows higher resemblance with the PME solution as judged by the value of ΔW_{R}. We found that typically much longer simulations times (more than 40 ns per window) were required compared to PME-based simulations (typically 20 ns per window were sufficient) in order to achieve converged estimates. No significant impact of the underlying cut-off schemes (SR vs. TR) could be observed. As in the previous sections, no systematic system size dependence was found.

Impact of the Free Energy Estimator

For the considered systems unpolar ligand / polar CNT it was found that artifacts in the form of a PMF offset as detected previously in another context (c.f. Sec. 4.2), can be introduced by the analysis method. Fig. 12 shows a comparison between PMFs as obtained from different estimators (UI, WHAM, MBAR) for the example of unpolar ethane / polar CNT
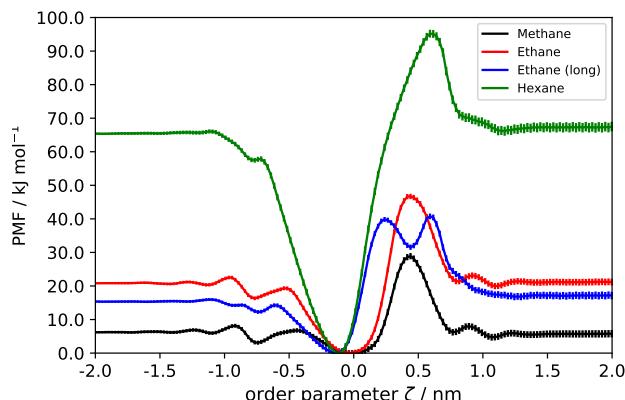
(a)

(b)

Figure 9. Effect of the treatment of electrostatic interactions (PME vs. RF), the cut-off scheme (SR vs. TR) and the pairlist generation scheme (CG vs. AT) on the PMF for (a) unpolar ethane / unpolar CNT and (b) unpolar hexane / unpolar CNT (c.f. Sec. 3.3). No orientational restraint was applied to the ligands. Bounds for statistical uncertainties are below 1.0 kJ mol^{-1} and have been omitted in the interest of clarity.

based on the (RF, TR-CG) setup. As can be seen, an offset between the flat bulk water regions of around 7 kJ mol^{-1} is present in the profile obtained from the UI method, which significantly exceeds bounds due to the statistical uncertainty. Its origin can be explained by means of the sampled biased distribution functions of the order parameter ζ (c.f. Fig. 13). As noted previously in Sec. 3.4, a central assumption in the UI approach is that the biased distributions can be approximated as Gaussian distributions. In Fig. 13 (a) it can be seen that the distribution sampled from window close to ζ = 0.5 nm at the polar CNT entrance differs from the rest and is non-Gaussian in shape. The set of distributions from corresponding PME-based simulation in contrast, does not contain such a window (c.f. Fig. 13 (b)). Such an offset was exclusively observed for simulations based on the reaction field treatment

Figure 10. PMFs for the association of a polar CNT with different unpolar ligands. No orientational restraint was imposed on the ligand. Long-range electrostatic interactions were treated with the PME method. PMFs were estimated via the UI method.
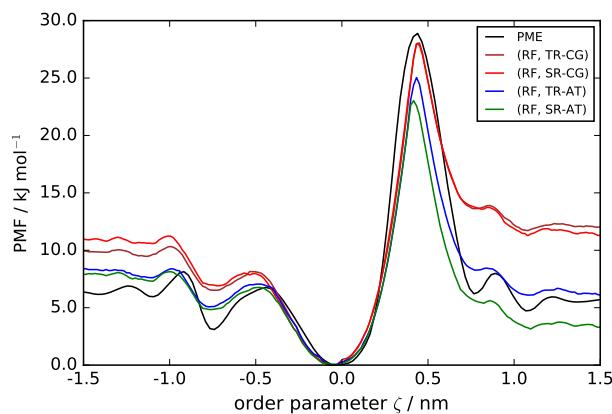
(including different ligands and combinations of cut-off and pairlist generation schemes) which is more susceptible for cut-off artifacts in structural solvation properties [99] but not for PME-based simulations. Nonetheless, we stress that this artifact is only indirectly caused by the electrostatics treatment but actually results from application of an estimator to a situation for which it was not designed for. The usage of a higher force constant for the umbrella distance restraint might probably remedy such a bias.
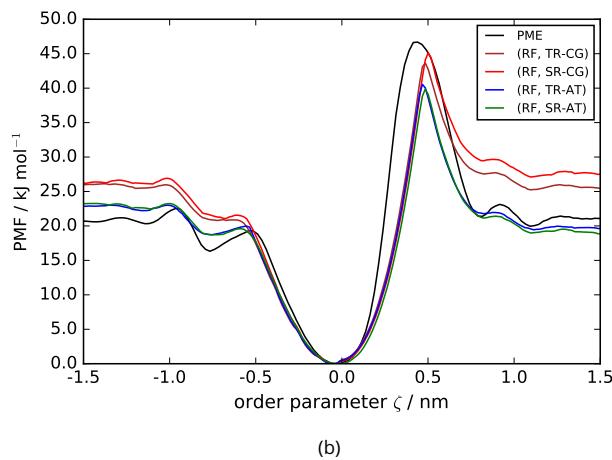
Lesson Learned
The examples show that also in the presence of considerable polar interactions between host and solvent, neither the differences in the treatment of electrostatic interactions nor in schemes for the cut-off or pairlist generation affect the estimated PMFs systematically. The artifact caused by the UI estimator demonstrates the benefit to compare different analysis methods on the same data set. Furthermore, if the UI estimator is used, the shape of the sampled distributions should be checked.

4.4 Polar CNT / Dipolar Ligand
This section reports PMFs for the association of a polar CNT with different dipolar ligands based on (elongated) ethane and hexane. The modeling of the polar CNT was described in the previous section. Dipolar ligands were modeled in a similar way by distributing a pair of balancing partial charges to the peripheric pair of covalently bound carbon atoms (C1-atom: +0.5 e, C2-atom: -0.5 e where "e" denotes the elementary charge). For all simulations considered in this paragraph, the PME treatment for long-range electrostatics was utilized. Since it was found that the different estimators (WHAM, UI,
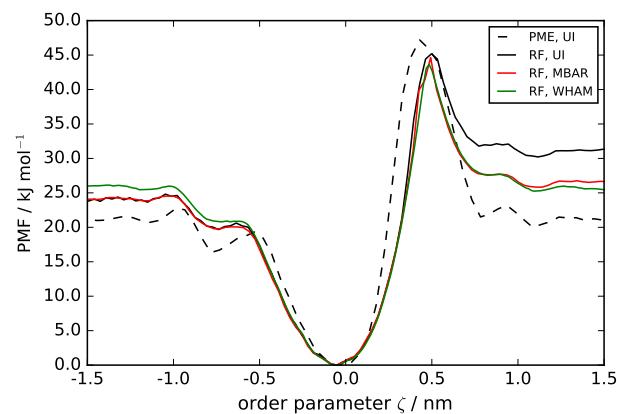
(a)

(b)

Figure 11. Effect of the treatment of electrostatic interactions (PME vs. RF), the cut-off scheme (SR vs. TR) and the pairlist generation scheme (CG vs. AT) on the PMF for (a) methane / polar CNT and (b) unpolar ethane / polar CNT (c.f. Sec. 3.3). No orientational restraint was applied for ethane. PMFs were estimated via the WHAM method. Error bars have been omitted in the interest of clarity.

MBAR) yield indistinguishable PMFs within errors bars, all profiles reported in the following refer to the UI result.

Sampling of Ligand Orientations

In contrast to the systems treated so far, two distinct binding configurations with different binding affinities can be distinguished. The bound configuration for which the positively charged ligand head (C1-atom) is facing (away from) the negatively charged C-atoms of the CNT is denoted as Conf. 1 (Conf. 2). PMFs for dipolar ethane, elongated ethane and hexane binding to the polar CNT are depicted in Fig. 14. Profiles in (a) and (b) were obtained without and with imposed orientational restraint on the ligand, respectively. For all simulations, the dipolar ligand was initially prepared in Conf. 1. Significant differences become apparent from comparison

Figure 12. Influence of the free energy estimator (UI vs. MBAR vs. WHAM) on the PMF for unpolar ethane / polar CNT. Electrostatics treatment refers to the (RF, TR-CG) setup as described in the main text. The profile obtained from simulations using PME and evaluated via UI is shown for comparison (black dashed line). No orientational restraint was imposed on the ligand.
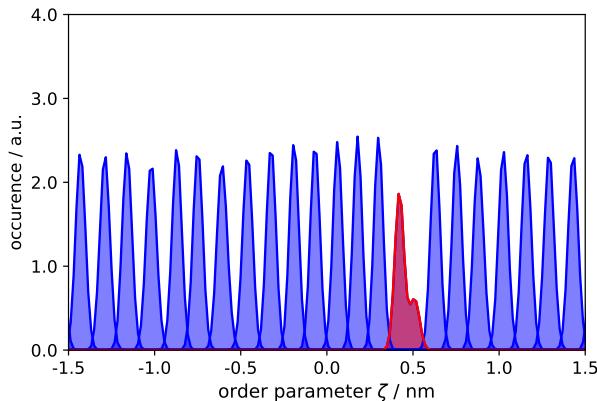
with corresponding profiles in Fig. 10 where the ligands "feel" the influence of the polar CNT only in an indirect manner mediated by the solvent. Fig. 14 (a) reveals substantial PMF offsets of 10 and 60 kJ mol^{-1} for hexane and elongated ethane, respectively, both of which are unable to rotate inside CNT. For ethane in contrast, which can rotate inside the CNT due to its small size, no offset is present. As demonstrated by Fig. 14 (b), such an offset can be removed for all considered ligands through inclusion of an orientational restraint in the simulation protocol.
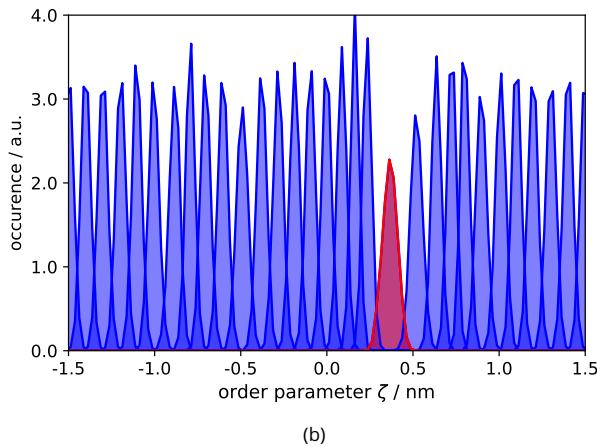
Lesson Learned

The examples illustrate that in case of asymmetric ligands binding to an asymmetric host (which is probably the most common case in practice), the biased distributions sampled in umbrella windows outside the binding site in which the ligand is free to rotate does not fit to the biased distributions sampled in windows for the bound state in which the binding pose is prescribed. This misfit can be illustrated by excluding configurations exhibiting the "wrong" orientation from the analysis which reduces the offset considerably (data not shown). We point out that excluding states from the analysis was done just to support our findings, and is not meant to be a suitable method to avoid offsets. The use of an orientational restraint is therefore mandatory in such cases unless the umbrella sampling is combined with Hamiltonian Replica Exchange as discussed in the following section.

4.5 Cyclodextrin / Alcohols

This section reports PMFs for the association of α-cyclodextrin (αCD) with two different primary alcohols (1-butanol, 1-
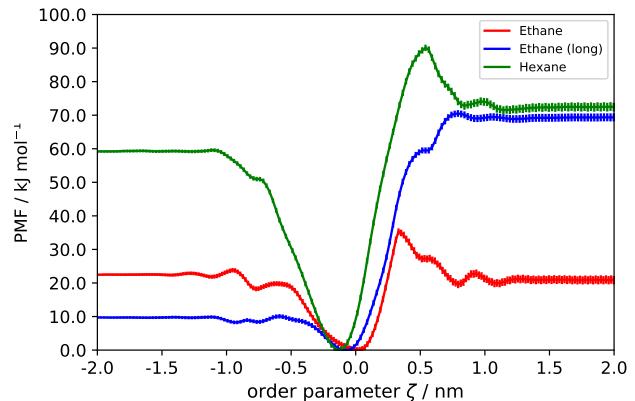
(a)

(b)

Figure 13. Influence of electrostatics treatment on sampled biased distributions of the order parameter ζ along the considered path for unpolar ethane / polar CNT. (a): RF-based simulations using the TR-CG setup, (b): PME-based simulations. The distribution close to ζ = 0.5 nm is highlighted in red to support the discussion in the main text. The abbreviation a.u. refers to arbitrary units.

(a)

(b)

Figure 14. PMFs for the association of a polar CNT with different dipolar ligands. All ligands were initially prepared in the same binding configuration (Conf. 1). Profiles in (a) and (b) refer to the setup without and with imposed orientational restraint on the ligand, respectively.
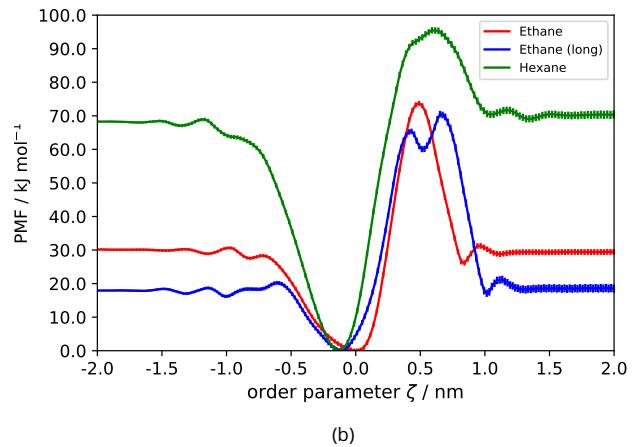
dodecanol). Results for further alcohols were reported in our previous work [16]. Since it was found that the different estimators (WHAM, UI, MBAR) yield indistinguishable PMFs within errors bars, all profiles reported in the following refer to the UI result.

Sampling of Ligand Orientations

As in the case of dipolar ligands binding to the polar CNT, two different binding configurations can be distinguished which will be denoted as Conf. 1 and Conf. 2 according to Refs. 16 and 100. Fig. 15 (a) and (b) show the PMFs for 1-butanol and 1-dodecanol binding to αCD, respectively. Each graph contains two PMFs, corresponding to the setup with (red curve) and without (black curve) orientational restraint with the ligand bound to αCD (Conf. 1). The third profile (green curve) in both

graphs corresponds to the PMFs as obtained from umbrella sampling combined with Replica Exchange (RE-US) [101]. In RE-US, the Hamiltonians of neighboring windows defined by the individual values for the bias centers are allowed to swap after predefined time instances, based on the Metropolis-Hastings criterion. Here, an exchange was attempted every 1000 steps. For RE-US simulations, no orientational restraint was applied. A significant offset is visible for the PMFs lacking an orientational restraint. As observed previously in case of the dipolar ligand / polar CNT system (c.f. Sec. 4.4), this offset can be remediated by restricting the ligand orientation. The fact that also the RE-US approach (without orientational restraint) yields an offset-free PMF, further demonstrates that this artifact is caused by a bias introduced when prescribing the binding pose in standard umbrella sampling.

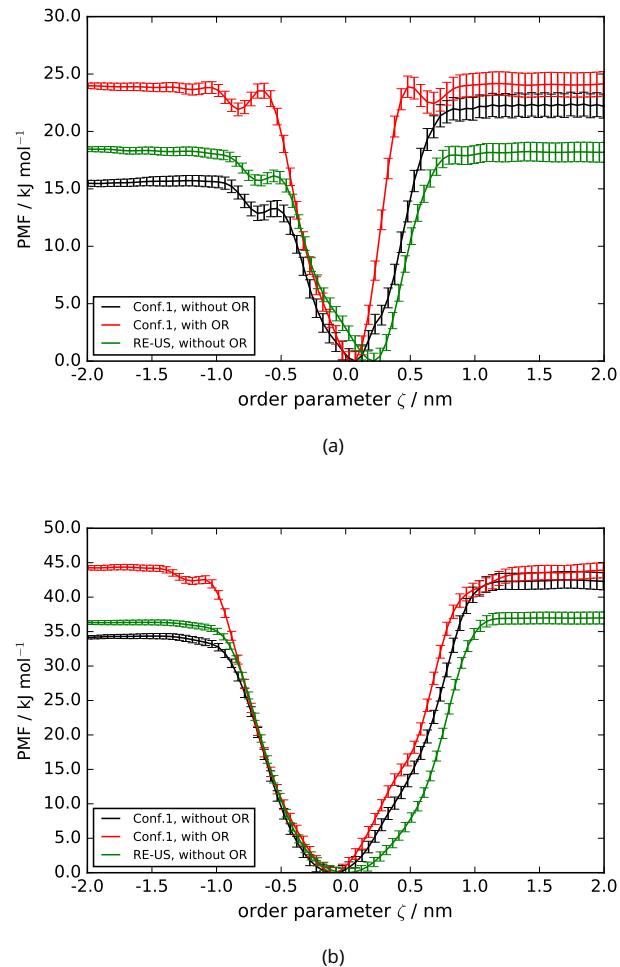The estimate of the binding free enthalpy as obtained

from the protocol including an orientational restraint, corresponds to one particular binding configuration (in this case Conf. 1) and should be therefore denoted as $\Delta G^{\circ}_{\text{bind,Conf.1}}$. For comparison with experiments which measure a configurational average, the binding free enthalpy for the second binding configuration (obtained from additional simulations and denoted as $\Delta G^{\circ}_{\text{bind,Conf.2}}$) can be combined with $\Delta G^{\circ}_{\text{bind,Conf.1}}$ via exponential averaging [102]:

$$\Delta G^{\circ}_{\text{bind}} = -RT \ln \left(e^{-\Delta G^{\circ}_{\text{bind,Conf.1}}/RT} + e^{-\Delta G^{\circ}_{\text{bind,Conf.2}}/RT} \right) \quad (12)$$

In RE-US simulations (without orientational restraint) in contrast, both binding configurations are sampled and the corresponding PMF can not be attributed to either Conf. 1 or Conf. 2 but already represents a configurational average. Thus, the estimate for $\Delta G^{\circ}_{\text{bind}}$ inferred from such a PMF can be directly compared with the corresponding experimental value without the need of additional simulations. However, this gain in efficiency might be offset by an overhead in terms of hardware resources and (depending on the system) computation time for reaching convergence. Here, it was found that in case of butanol 20 ns per window were sufficient to obtain converged PMFs while 140 ns per window were required for dodecanol. For standard umbrella sampling including an orientational restraint in contrast, 20 ns per window were found to be sufficient for all systems, at least in case of simulations based on the PME treatment for long-range electrostatics as mentioned previously. The good agreement between the $\Delta G^{\circ}_{\text{bind}}$ estimates obtained from the setup including an orientational restraint and the RE-US simulations (c.f. Tab. 5) indicates that the RE scheme not only removes the PMF artifact but especially samples both binding configurations with the correct weighting. Moreover, the results were found to be in good agreement with corresponding estimates from double decoupling (c.f. Fig. 5 in Ref. [16]).

Lesson Learned
The examples considering the binding of primary alcohols to the αCD-host show that in case of multiple binding configurations which are separated by significant energy barriers, artifacts in the form of a PMF offset might occur if the umbrella sampling protocol only includes a distance restraint. This artifact was already observed for the artificial model system dipolar ligand / polar CNT (c.f. Sec. 4.4) and is caused by insufficient sampling of ligand orientations in the binding site. The simulation protocol can be modified in two ways for such situations: (i) restraining the ligand's orientation to a specific binding configuration. For each binding configuration, one can calculate a binding free enthalpy and combine the distinct estimates during post processing (c.f Eq. (12)). (ii) combination of umbrella sampling with Replica Exchange to allow sampling of multiple ligand orientations in the binding

(a)

(b)

Figure 15. PMFs for (a) 1-butanol and (b) 1-dodecanol binding to αCD (Conf.1) [16, 100]. Red and black profiles refer to the setup with and without imposed orientational restraint (OR), respectively. The PMF as obtained from the RE-US approach (green curve) represents a configurational average of Conf. 1 and Conf. 2.

site. In this case no orientational restraint is required and the estimate for $\Delta G^{\circ}_{\text{bind}}$ represents a configurational average.
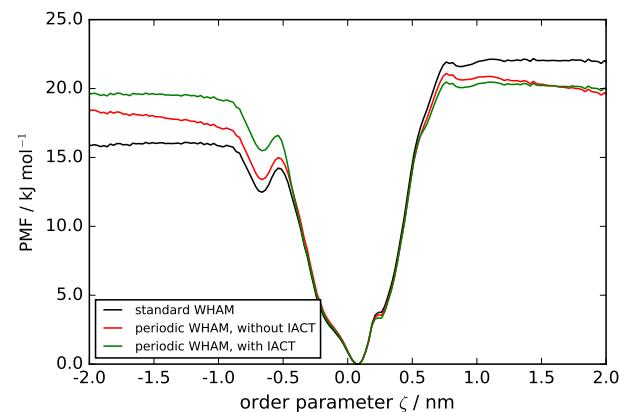
5 Discussion
5.1 Enforcing PMF Periodicity
In Sec. 4.2, 4.4 and 4.5, it was shown that PMF offsets due to an unfortunate choice of restraining reference points or insufficient sampling of ligand configurations in the bound state can be eliminated through application of a restraint acting on the ligand's orientation or Replica Exchange in the simulation protocol. However, to obtain more realistic PMF estimates also in case of existent simulation data, sampled from non-optimized protocols, Hub et al. [33] proposed another workaround. In their approach which focuses on the post processing estimation, a modified version of the WHAM

Table 5. Calculated standard binding free enthalpies $\Delta G^{\circ}_{\text{bind}}$ for the binding of 1-butanol (BTL) and 1-dodecanol (DDL) to αCD (c.f. Sec. 4.5). The results as obtained from the setup with restrained ligand orientation (OR) are compared with corresponding results from Replica Exchange (RE) umbrella sampling which yield a configurationally averaged $\Delta G^{\circ}_{\text{bind}}$. Corresponding PMFs are depicted in Fig. 15. Calculations of l_b, ΔG_V and ΔG_{Ω} were performed as described in Sec. 2. The joint contribution of the translational and orientational restraint in the bound state $\Delta G_{\rho} + \Delta G_{\theta}$ was calculated using the MBAR estimator from a sequence of simulations in the bound state with force constants k_{ρ} and k_{θ} varying from zero to the final values as specified in Tab. 1. The estimate for $\Delta G^{\circ}_{\text{bind}}$ from the setup including an orientational restraint as reported in the last column follows from exponential averaging of the values $\Delta G^{\circ}_{\text{bind,Conf.1}}$ and $\Delta G^{\circ}_{\text{bind,Conf.2}}$ according to Eq. (12). Error estimates refer to the UI result.

System	Setup	Conf. X	ΔW_R [kJ mol^{-1}]	l_b [nm]	ΔG_V [kJ mol^{-1}]	ΔG_{Ω} [kJ mol^{-1}]	$\Delta G_{\rho} + \Delta G_{\theta}$ [kJ mol^{-1}]	$\Delta G^{\circ}_{\text{bind,Conf.X}}$ [kJ mol^{-1}]	$\Delta G^{\circ}_{\text{bind}}$ [kJ mol^{-1}]
BTL	OR	1	-24.17 ± 1.11	0.2218	5.72	14.93	-6.50	-10.02 ± 1.11	-13.86 ± 1.11
	OR	2	-26.92 ± 1.11	0.2017	5.96	14.95	-7.24	-13.25 ± 1.11	
	RE	1,2	-18.39 ± 0.90	0.3002	4.96	0.00	0.00	-	-13.43 ± 0.90
DDL	OR	1	-44.14 ± 1.11	0.2982	4.98	14.93	-7.01	-31.24 ± 1.11	-33.10 ± 1.11
	OR	2	-43.94 ± 1.11	0.2988	4.97	14.95	-7.49	-31.50 ± 1.11	
	RE	1,2	-36.66 ± 0.89	0.4842	3.77	0.00	0.00	-	-32.89 ± 0.89

algorithm was developed. Their method, denoted as `g_wham` as part of the GROMACS program collection, offers the calculation of integrated autocorrelation times (IACT) for reducing the bias from limited sampling as well as constraints for enforced FEC periodicity and / or symmetry. It should be kept in mind that imposing such a constraint will yield a solution for the free energy profile which - by design - satisfies the state function property by preventing an offset. On the other hand, it clearly does not reveal any information about the origin of this artifact, nor does it solve the actual sampling problem. Moreover, such an artificially generated FEC (and in consequence the derived estimate for $\Delta G^{\circ}_{\text{bind}}$) might deviate significantly from the "true" profile, one would obtain in the absence of any sampling issues. To study the effect of the periodicity constraint, the simulation data for butanol / αCD (Conf. 1) without restraining the butanol orientation were reevaluated using `g_wham`. Resulting profiles with and without enforced periodicity are shown in Fig. 16. Estimation using standard WHAM without enforced periodicity yields a significant offset as shown beforehand (c.f. Fig. 15 (a)). As can be seen, the application of the periodicity constraint yields identical values at the end points of the considered order parameter interval but it induces artificial slopes in the bulk water regions. This artifact was also described in the original publication [33] where it was ascribed to the neglect of locally different IACT. Therefore, additional analysis was performed by incorporation of the distribution of local IACT into the analysis in addition to the enforced periodicity. The resulting periodic and IACT-corrected profile indeed shows flat bulk water regions. Estimation of the standard binding free enthalpy from the periodic / periodic and IACT-corrected profiles yields -13.9 / -14.7 kJ mol^{-1}, respectively, compared to -10.0 kJ mol^{-1} as obtained from standard WHAM estimation from the setup including an orientational restraint (c.f. Tab. 5).

Figure 16. Effect of enforced periodicity (periodic) and integrated autocorrelation times (IACT) on the PMF for the system 1-butanol / αCD (Conf. 1). No orientational restraint was applied to the ligand. Calculation was performed using the `g_wham` method [33]. Error bars were neglected for clarity. Standard WHAM calculation (black curve) refers to estimation without IACT correction and without periodicity constraint.

This examples illustrates that in the context of binding free enthalpy calculations, the usage of artificially constrained profiles might give a reasonable estimate for $\Delta G^{\circ}_{\text{bind}}$, however, one should be aware of that such a value does not purely reflect the precision of the force field. If very precise estimates are required (either for the profile itself or $\Delta G^{\circ}_{\text{bind}}$), we advise to focus on the elimination of possible sampling issues in the simulation protocol (if system complexity allows it) and to use non-constrained estimation.

5.2 Influence of the Host's / Ligand's Flexibility

If the host molecule is able to adopt multiple conformations, a bias might be introduced caused by the selection of initial conformations of the host or the method for generating starting configurations of the umbrella windows. As found by You et al. [53] from studies of βCD complexes, significantly different PMF depths can be obtained depending on the initial host conformation unless simulation time was sufficient. For such cases, discrepancies of the estimated binding free enthalpy compared to results from unbiased direct counting might be expected. Due to the insensitivity of the adopted CNT conformations upon ligand binding alongside with the insensitivity of the PMFs towards increasingly restrictive restraining setups (c.f. Sec. 4.1), we do not expect such a bias in this case. For further validation, a modified CNT was studied featuring decreased barriers for proper and improper dihedrals compared to the standard model. Despite increased conformational flexibility, the resulting PMF obtained from the association with hexane (data not shown) was identical with the profile as shown in Fig. 9 (b). For simulations based on αCD, we conclude from the good agreement between the PMF-based estimates for $\Delta G^{\circ}_{\text{bind}}$ and the corresponding results from double decoupling [16, 100] as well as direct counting [19] that simulation time was sufficient in order to remove any possible bias due to the initial host conformations. Moreover, the force field used in the present study does not show multiple conformations for αCD [103]. Considering host molecules which tend to undergo significant conformational changes upon ligand binding, the incorporation of a conformational restraint to bias the host conformation close to the bound state conformation might be advantageous [35]. Moreover, the ligand conformation could be biased analogously which might be of practical value for speeding up convergence, especially for very flexible ligands. The impact of such a conformational restraint with respect to the calculation of $\Delta G^{\circ}_{\text{bind}}$ can be calculated rigorously [35]. In this case, Eq. (2) has to be complemented by the free energy contribution from rigidification of the non-complexed host (and / or unbound ligand) and the contribution from releasing the conformational restraint from the complexed host (and / or bound ligand) again. To obtain accurate results for this process, the force field has to capture the relative energies of the different conformers very accurately [104]. As judged by the good agreement for the $\Delta G^{\circ}_{\text{bind}}$ estimates obtained from the PMF and double decoupling in case of αCD / dodecanol, we conclude that no conformational restraint is required for the flexible ligands considered in this work.

6 Conclusions

In this article, we studied the evaluation of one-dimensional potentials of mean force (PMF) of host-guest system obtained via umbrella sampling. A carbon nanotube (CNT) and α-cyclodextrin (αCD) were chosen as idealized model systems for pore- or channel-like protein host molecules featuring a hydrophobic cavity. A robust simulation protocol for the calculation of standard binding free enthalpies from such a PMF was established. From systematic studies of different CNT / ligand combinations of increasing complexity, we could identify distinct computational artifacts that may occur in the PMF calculation. Such artifacts which show up as PMF offset between the two flat bulk water regions prohibit an unambiguous estimation of the binding free enthalpy and have not been studied in detail so far. It was found that despite an identical manifestation, three different origins for PMF offsets can be distinguished: (i) an unfortunate choice of reference points for the umbrella distance restraint; (ii) a misfit in probability distributions between bound and unbound umbrella windows in case of multiple binding modes; (iii) offsets introduced by the UI estimator due to non-Gaussian-shaped biased distribution functions. It is important to distinguish these origins from possible primary reasons such as insufficient overlap between neighboring umbrella windows (which is especially critical when estimation is performed with WHAM) or insufficient sampling time. Neither the introduction of additional windows nor the extension of simulation time per window will eliminate the PMF artifacts in these cases. It was shown that offsets due to (i) and (ii) can be eliminated by either restraining the ligand orientation close to the bound state orientation or through combination of the umbrella sampling setup with Replica Exchange (RE-US). Application of two-dimensional umbrella sampling by incorporation of a second biased coordinate such as the orientational angle θ, might be an alternative to the application of restraints that may provide insight into the free energy surface at the rim region [31, 32, 105]. Offsets resulting from the analysis method can be identified by comparing PMF results from different estimators (UI, MBAR, WHAM). Such a comparison which serves as consistency check is always recommended. We note that comparative simulations for αCD / alcohol systems conducted with the CHARMM36 all-atom force field also lead to PMF offsets if the ligand orientation was not restrained (c.f. Fig. A1). This illustrates that the detected artifacts are force-field independent. Regarding the influence of the simulation protocol, it can be expected that artifacts due to issues (i) and (ii) also occur for alternative PMF-based protocols such as Forward Flux Sampling if the ligand orientation is not preserved or proper sampling of multiple orientations can not be guaranteed.

7 Author Contributions

All authors designed the simulations, interpreted data and contributed to the writing of the paper. SJ and NH performed preliminary simulations leading to this project. DM carried out all simulations and analyses discussed in the manuscript.

8 Other Contributions

The authors gratefully acknowledge Dr. Jens Krüger for technical support as well as Prof. A. E. Mark and Prof. P. H. Hünenberger for helpful discussion regarding the occurrence of PMF offsets. Moreover, the authors thank Prof. J. Pleiss from the University of Stuttgart who organized together with the author NH a seminar for the graduate school in Simulation Technology (GS SimTech) during the winter semester 2014 which gave rise to fruitful discussions and inspiring ideas for the establishment of robust simulation protocols. We thank Catharina Kleist and Felix Geisendörfer who, as students at the Hamburg University of Technology, performed some preliminary simulations for the present work. Finally, the authors thank Prof. A. Grossfield for his comments which helped to improve the clarity of the manuscript.

9 Potentially Conflicting Interests

The authors declare no competing financial interest.

10 Funding Information

Author Information

ORCID:
Daniel Markthaler: 0000-0003-3281-436X
Sven Jakobtorweihen: 0000-0001-5492-8304
Niels Hansen: 0000-0003-4366-6120

References

[1] **Leelananda SP**, Lindert S. Computational Methods in Drug Discovery. Beilstein J Org Chem. 2016; 12(1):2694–2718. https://doi.org/10.3762/bjoc.12.267.

[2] **Gilson MK**, Given JA, Bush BL, McCammon JA. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: a Critical Review. Biophys J. 1997; 72(3):1047–1069. https://doi.org/10.1016/S0006-3495(97)78756-3.

[3] **Boresch S**, Tettinger F, Leitgeb M, Karplus M. Absolute Binding Free Energies: A Quantitative Approach for their Calculation. J Phys Chem B. 2003; 107(35):9535–9551. https://doi.org/10.1021/jp0217839.

[4] **Pohorille A**, Jarzynski C, Chipot C. Good Practices in Free-Energy Calculations. J Phys Chem B. 2010; 114(32):10235–10253. https://doi.org/10.1021/jp102971x.

[5] **IUPAP**. Symbols, Units and Nomenclature in Physics. Physica A. 1978; 93(1–2):1–60. https://doi.org/10.1016/0378-4371(78)90209-1.

[6] **Renner T**. Quantities, Units and Symbols in Physical Chemistry. Cohen ER, Cvitas T, Frey JG, Holström B, Kuchitsu K, Marquardt R, Mills I, Pavese F, Quack M, Stohner J, Strauss HL, Takami M, Thor AJ, editors, The Royal Society of Chemistry; 2007. https://doi.org/10.1039/9781847557889.

[7] **Pople JA**. Nobel Lecture: Quantum Chemical Models. Rev Mod Phys. 1999; 71(5):1267. https://doi.org/10.1103/RevModPhys.71.1267.

[8] **Rocklin GJ**, Mobley DL, Dill KA. Calculating the Sensitivity and Robustness of Binding Free Energy Calculations to Force Field Parameters. J Chem Theory Comput. 2013; 9(7):3072–3083. https://doi.org/10.1021/ct400315q.

[9] **Yin J**, Fenley AT, Henriksen NM, Gilson MK. Toward Improved Force-Field Accuracy through Sensitivity Analysis of Host-Guest Binding Thermodynamics. J Phys Chem B. 2015; 119(32):10145–10155. https://doi.org/10.1021/acs.jpcb.5b04262.

[10] **Zuckerman DM**. Equilibrium Sampling in Biomolecular Simulations. Annu Rev Biophys. 2011; 40(1):41–62. https://doi.org/10.1146/annurev-biophys-042910-155255.

[11] **Christ CD**, Mark AE, van Gunsteren WF. Basic Ingredients of Free Energy Calculations: A Review. J Comput Chem. 2010; 31(8):1569–1582. https://doi.org/10.1002/jcc.21450.

[12] **van Gunsteren WF**, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NFA, Yu HB. Biomolecular Modeling: Goals, Problems, Perspectives. Angew Chem Int Ed. 2006; 45(25):4064–4092. https://doi.org/10.1002/anie.200502655.

[13] **König G**, Brooks BR. Predicting Binding Affinities of Host-Guest Systems in the SAMPL3 Blind Challenge: The Performance of Relative Free Energy Calculations. J Comput Aided Mol Des. 2012; 26(5):543–550. https://doi.org/10.1007/s10822-011-9525-y.

[14] **Wong-ekkabut J**, Karttunen M. The Good, the Bad and the User in Soft Matter Simulations. Biochim Biophys Acta. 2016; 1858(10):2529–2538. https://doi.org/10.1016/j.bbamem.2016.02.004.

[15] **Deng Y**, Roux B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. J Phys Chem B. 2009; 113(8):2234–2246. https://doi.org/10.1021/jp807701h.

[16] **Markthaler D**, Gebhardt J, Jakobtorweihen S, Hansen N. Molecular Simulations of Thermodynamic Properties for the System α-Cyclodextrin/Alcohol in Aqueous Solution. Chem Ing Tech. 2017; 89(10):1306–1314. https://doi.org/10.1002/cite.201700057.

[17] **Gumbart JC**, Roux B, Chipot C. Standard Binding Free Energies from Computer Simulations: What is the Best Strategy? J Chem Theory Comput. 2012; 9(1):794–802. https://doi.org/10.1021/ct3008099.

[18] **Pan AC**, Xu H, Palpant T, Shaw DE. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. J Chem Theory Comput. 2017; 13(7):3372–3377. https://doi.org/10.1021/acs.jctc.7b00172.

[19] **Baz J**, Gebhardt J, Kraus H, Markthaler D, Hansen N. Insights into Noncovalent Binding Obtained from Molecular Dynamics Simulations. Chem Ing Tech. 2018; 90(11):1864–1875. https://doi.org/10.1002/cite.201800050.

[20] **Torrie GM**, Valleau JP. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-critical Lennard-Jones Fluid. Chem Phys Lett. 1974; 28(4):578–581. https://doi.org/10.1016/0009-2614(74)80109-0.

[21] **Torrie GM**, Valleau JP. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. J Comput Phys. 1977; 23(2):187–199. https://doi.org/10.1016/0021-9991(77)90121-8.

[22] **Huber T**, Torda AE, Van Gunsteren WF. Local Elevation: A Method for Improving the Searching Properties of Molecular Dynamics Simulation. J Comput Aided Mol Des. 1994; 8(6):695–708. https://doi.org/10.1007/BF00124016.

[23] **Laio A**, Parrinello M. Escaping Free-Energy Minima. Proc Natl Acad Sci USA. 2002; 99(20):12562–12566. https://doi.org/10.1073/pnas.202427399.

[24] **Darve E**, Rodríguez-Gómez D, Pohorille A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. J Chem Phys. 2008; 128(14):144120. https://doi.org/10.1063/1.2829861.

[25] **Allen RJ**, Valeriani C, ten Wolde PR. Forward Flux Sampling for Rare Event Simulations. J Phys Condens Matter. 2009; 21(46):463102. https://doi.org/10.1088/0953-8984/21/46/463102.

[26] **Isralewitz B**, Gao M, Schulten K. Steered Molecular Dynamics and Mechanical Functions of Proteins. Curr Opin Cell Biol. 2001; 11(2):224–230. https://doi.org/10.1016/S0959-440X(00)00194-9.

[27] **Del Valle EM**. Cyclodextrins and their Uses: A Review. Process Biochem. 2004; 39(9):1033–1046. https://doi.org/10.1016/S0032-9592(03)00258-9.

[28] **Velez-Vega C**, Gilson MK. Force and Stress along Simulated Dissociation Pathways of Cucurbituril–Guest Systems. J Chem Theory Comput. 2012; 8(3):966–976. https://doi.org/10.1021/ct2006902.

[29] **Velez-Vega C**, Gilson MK. Overcoming Dissipation in the Calculation of Standard Binding Free Energies by Ligand Extraction. J Comput Chem. 2013; 34(27):2360–2371. https://doi.org/10.1002/jcc.23398.

[30] **Yordanova D**, Ritter E, Gerlach T, Jensen JH, Smirnova I, Jakobtorweihen S. Solute Partitioning in Micelles: Combining Molecular Dynamics Simulations, COSMOmic, and Experiments. J Phys Chem B. 2017; 121(23):5794–5809. https://doi.org/10.1021/acs.jpcb.7b03147.

[31] **Allen TW**, Andersen OS, Roux B. Ion Permeation through a Narrow Channel: Using Gramicidin to Ascertain All-Atom Molecular Dynamics Potential of Mean Force Methodology and Biomolecular Force Fields. Biophys J. 2006; 90(10):3447–3468. https://doi.org/10.1529/biophysj.105.077073.

[32] **Allen TW**, Andersen OS, Roux B. Molecular Dynamics—Potential of Mean Force Calculations as a Tool for Understanding Ion Permeation and Selectivity in Narrow Channels. Biophys Chem. 2006; 124(3):251–267. https://doi.org/10.1016/j.bpc.2006.04.015.

[33] **Hub JS**, De Groot BL, Van Der Spoel D. g_wham - a Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J Chem Theory Comput. 2010; 6(12):3713–3720. https://doi.org/10.1021/ct100494z.

[34] **Flood E**, Boiteux C, Lev B, Vorobyov I, Allen TW. Atomistic Simulations of Membrane Ion Channel Conduction, Gating, and Modulation. Chem Rev. 2019; https://doi.org/10.1021/acs.chemrev.8b00630.

[35] **Woo HJ**, Roux B. Calculation of Absolute Protein–Ligand Binding Free Energy from Computer Simulations. Proc Natl Acad Sci USA. 2005; 102(19):6825–6830. https://doi.org/10.1073/pnas.0409005102.

[36] **Filippini G**, Goujon F, Bonal C, Malfreyt P. Energetic Competition Effects on Thermodynamic Properties of Association Between β-CD and Fc Group: A Potential of Mean Force Approach. J Phys Chem C. 2012; 116(42):22350–22358. https://doi.org/10.1021/jp3057724.

[37] **Filippini G**, Bonal C, Malfreyt P. Why is the Association of Supramolecular Assemblies Different under Homogeneous and Heterogeneous Conditions? Phys Chem Chem Phys. 2012; 14(29):10122–10124. https://doi.org/10.1039/c2cp41661a.

[38] **Baştuğ T**, Chen PC, Patra SM, Kuyucak S. Potential of Mean Force Calculations of Ligand Binding to Ion Channels from Jarzynski's Equality and Umbrella Sampling. J Chem Phys. 2008; 128(15):155104. https://doi.org/10.1063/1.2904461.

[39] **Krüger J**, Fels G. Potential of Mean Force of Ion Permeation through alpha7 nAChRIon Channel. In: *International Workshop on Portals for Life Sciences*, vol. 513 of CEUR Workshop Proceedings; 2009. http://ceur-ws.org/Vol-513/paper06.pdf.

[40] **Neale C**, Hsu JC, Yip CM, Pomès R. Indolicidin Binding Induces Thinning of a Lipid Bilayer. Biophys J. 2014; 106(8):L29–L31. https://doi.org/10.1016/j.bpj.2014.02.031.

[41] **Sun R**, Han Y, Swanson JM, Tan JS, Rose JP, Voth GA. Molecular Transport through Membranes: Accurate Permeability Coefficients from Multidimensional Potentials of Mean Force and Local Diffusion Constants. J Chem Phys. 2018; 149(7):072310. https://doi.org/10.1063/1.5027004.

[42] **Zhou HX**, Gilson MK. Theory of Free Energy and Entropy in Noncovalent Binding. Chem Rev. 2009; 109(9):4092–4107. https://doi.org/10.1021/cr800551w.

[43] **General IJ**. A Note on the Standard State's Binding Free Energy. J Chem Theory Comput. 2010; 6(8):2520–2524. https://doi.org/10.1021/ct100255z.

[44] **Doudou S**, Burton NA, Henchman RH. Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. J Chem Theory Comput. 2009; 5(4):909–918. https://doi.org/10.1021/ct8002354.

[45] **Zwanzig RW**. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J Chem Phys. 1954; 22(8):1420–1426. https://doi.org/10.1063/1.1740409.

[46] **Kirkwood JG**. Statistical Mechanics of Fluid Mixtures. J Chem Phys. 1935; 3(5):300–313. https://doi.org/10.1063/1.1749657.

[47] **Shirts MR**, Chodera JD. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. J Chem Phys. 2008; 129(12):124105. https://doi.org/10.1063/1.2978177.

[48] **Hermans J**, Wang L. Inclusion of Loss of Translational and Rotational Freedom in Theoretical Estimates of Free Energies of Binding. Application to a Complex of Benzene and Mutant T4 Lysozyme. J Am Chem Soc. 1997; 119(11):2707–2714. https://doi.org/10.1021/ja963568+.

[49] **Steinberg IZ**, Scheraga HA. Entropy Changes Accompanying Association Reactions of Proteins. J Biol Chem. 1963; 238(1):172–181. http://www.jbc.org/content/238/1/172.short.

[50] **Mobley DL**, Gilson MK. Predicting Binding Free Energies: Frontiers and Benchmarks. Annu Rev Biophys. 2017; 46:531–558. https://doi.org/10.1146/annurev-biophys-070816-033654.

[51] **Weiß RG**, Chudoba R, Setny P, Dzubiella J. Affinity, Kinetics, and Pathways of Anisotropic Ligands Binding to Hydrophobic Model Pockets. J Chem Phys. 2018; 149(9):094902. https://doi.org/10.1063/1.5025118.

[52] **IUPAC-IUB**. Symbols for Specifying the Conformation of Polysaccharide Chains, Recommendations 1981. Eur J Biochem. 1983; 131:5–7. https://doi.org/10.1111/j.1432-1033.1983.tb07224.x.

[53] **You W**, Tang Z, Chang CeA. Potential Mean Force from Umbrella Sampling Simulations: What can We Learn and What is Missed? J Chem Theory Comput. 2019; 15(4):2433–2443. https://doi.org/10.1021/acs.jctc.8b01142.

[54] **Trzesniak D**, Kunz APE, van Gunsteren WF. A Comparison of Methods to Compute the Potential of Mean Force. ChemPhysChem. 2007; 8(1):162–169. https://doi.org/10.1002/cphc.200600527.

[55] **Schmid N**, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. Eur Biophys J. 2011; 40(7):843. https://doi.org/10.1007/s00249-011-0700-9.

[56] **Pol-Fachin L**, Rusu VH, Verli H, Lins RD. GROMOS 53A6GLYC, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. J Chem Theory Comput. 2012; 8(11):4681–4690. https://doi.org/10.1021/ct300479h.

[57] **Berendsen HJC**, Postma JPM, van Gunsteren WF, Hermans J. Interaction Models for Water in Relation to Protein Hydration. In: Pullmann B, editor. *Intermolecular Forces* Dordrecht, the Netherlands: Reidel; 1981.p. 331–342. https://doi.org/10.1007/978-94-015-7658-1_21.

[58] **Eastwood JW**, Hockney RW. Computer Simulation Using Particles. N. Y.: McGraw-Hill; 1981.

[59] **Berendsen HJC**, van der Spoel D, van Drunen R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. Comput Phys Commun. 1995; 91(1-3):43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

[60] **Hess B**, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput. 2008; 4(3):435–447. https://doi.org/10.1021/ct700301q.

[61] **Abraham MJ**, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. SoftwareX. 2015; 1:19–25. https://doi.org/10.1016/j.softx.2015.06.001.

[62] **Silva TF**, Vila-Viçosa D, Reis PB, Victor BL, Diem M, Oostenbrink C, Machuqueiro M. The Impact of Using Single Atomistic Long-Range Cutoff Schemes with the GROMOS 54A7 Force Field. J Chem Theory Comput. 2018; 14(11):5823–5833. https://doi.org/10.1021/acs.jctc.8b00758.

[63] **Gonçalves YM**, Senac C, Fuchs PFJ, Hünenberger PH, Horta BA. Influence of the Treatment of Non-Bonded Interactions on Thermodynamic and Transport Properties of Pure Liquids Calculated Using the 2016H66 Force Field. J Chem Theory Comput. 2019; 15(3):1806–1826. https://doi.org/10.1021/acs.jctc.8b00425.

[64] **Reißer S**, Poger D, Stroet M, Mark AE. Real Cost of Speed: The Effect of a Time-Saving Multiple-Time-Stepping Algorithm on the Accuracy of Molecular Dynamics Simulations. J Chem Theory Comput. 2017; 13(6):2367–2372. https://doi.org/10.1021/acs.jctc.7b00178.

[65] **Loeffler HH**, Bosisio S, Duarte Ramos Matos G, Suh D, Roux B, Mobley DL, Michel J. Reproducibility of Free Energy Calculations across Different Molecular Simulation Software Packages. J Chem Theory Comput. 2018; 14(11):5567–5582. https://doi.org/10.1021/acs.jctc.8b00544.

[66] **Kunz AE**, Allison JR, Geerke DP, Horta BAC, Hünenberger PH, Riniker S, Schmid N, van Gunsteren WF. New Functionalities in the GROMOS Biomolecular Simulation Software. J Comput Chem. 2012; 33(3):340–353. https://doi.org/10.1002/jcc.21954.

[67] **Riniker S**, Christ CD, Hansen HS, Hünenberger PH, Oostenbrink C, Steiner D, van Gunsteren WF. Calculation of Relative Free Energies for Ligand-Protein Binding, Solvation, and Conformational Transitions Using the GROMOS Software. J Phys Chem B. 2011; 115(46):13570–13577. https://doi.org/10.1021/jp204303a.

[68] **Schmid N**, Christ CD, Christen M, Eichenberger AP, van Gunsteren WF. Architecture, Implementation and Parallelisation of the GROMOS Software for Biomolecular Simulation. Comput Phys Commun. 2012; 183(4):890–903. https://doi.org/10.1016/j.cpc.2011.12.014.

[69] **Papadourakis M**, Bosisio S, Michel J. Blinded Predictions of Standard Binding Free Energies: Lessons Learned from the SAMPL6 Challenge. J Comput Aided Mol Des. 2018; 32(10):1047–1058. https://doi.org/10.1007/s1082.

[70] **Åqvist J**. Cold Adaptation of Triosephosphate Isomerase. Biochemistry. 2017; 56(32):4169–4176. https://doi.org/10.1021/acs.biochem.7b00523.

[71] **Darden T**, York D, Pedersen L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. J Chem Phys. 1993; 98(12):10089–10092. https://doi.org/10.1063/1.464397.

[72] **Essmann U**, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A Smooth Particle Mesh Ewald Method. J Chem Phys. 1995; 103(19):8577–8593. https://doi.org/10.1063/1.470117.

[73] **Tribello GA**, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New Feathers for an Old Bird. Comput Phys Commun. 2014; 185(2):604–613. https://doi.org/10.1016/j.cpc.2013.09.018.

[74] **Hess B**, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A Linear Constraint Solver for Molecular Simulations. J Comput Chem. 1997; 18(12):1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[75] **Hess B**. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J Chem Theory Comput. 2008; 4(1):116–122. https://doi.org/10.1021/ct700200b.

[76] **Miyamoto S**, Kollman PA. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. J Comput Chem. 1992; 13(8):952–962. https://doi.org/10.1002/jcc.540130805.

[77] **Berendsen HJC**, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular Dynamics with Coupling to an External Bath. J Chem Phys. 1984; 81(8):3684–3690. https://doi.org/10.1063/1.448118.

[78] **Kell G**. Precise Representation of Volume Properties of Water at One Atmosphere. J Chem Eng Data. 1967; 12(1):66–69. https://doi.org/10.1021/je60032a018.

[79] **Nosé S**. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. Mol Phys. 1984; 52(2):255–268. https://doi.org/10.1080/00268978400101201.

[80] **Hoover WG**. Canonical Dynamics: Equilibrium Phase-Space Distributions. Phys Rev A. 1985; 31(3):1695. https://doi.org/10.1103/PhysRevA.31.1695.

[81] **Martyna GJ**, Tuckerman ME, Tobias DJ, Klein ML. Explicit Reversible Integrators for Extended Systems Dynamics. Mol Phys. 1996; 87(5):1117–1157. https://doi.org/10.1080/00268979600100761.

[82] **Parrinello M**, Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. J Appl Phys. 1981; 52(12):7182–7190. https://doi.org/10.1063/1.328693.

[83] **Nosé S**, Klein ML. Constant Pressure Molecular Dynamics for Molecular Systems. Mol Phys. 1983; 50(5):1055–1076. https://doi.org/10.1080/00268978300102851.

[84] **Páll S**, Hess B. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures. Comput Phys Commun. 2013; 184(12):2641–2650. https://doi.org/10.1016/j.cpc.2013.06.003.

[85] **Barker J**, Watts R. Monte Carlo Studies of the Dielectric Properties of Water-Like Models. Mol Phys. 1973; 26(3):789–792. https://doi.org/10.1080/00268977300102101.

[86] **Ryckaert JP**, Ciccotti G, Berendsen HJ. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. J Comput Phys. 1977; 23(3):327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

[87] **Heinz TN**, van Gunsteren WF, Hünenberger PH. Comparison of Four Methods to Compute the Dielectric Permittivity of Liquids from Molecular Dynamics Simulations. J Chem Phys. 2001; 115(3):1125–1136. https://doi.org/10.1063/1.1379764.

[88] **Zuiderweg ER**, Scheek RM, Boelens R, van Gunsteren WF, Kaptein R. Determination of Protein Structures from Nuclear Magnetic Resonance Data Using a Restrained Molecular Dynamics Approach: The Lac Repressor DNA Binding Domain. Biochimie. 1985; 67(7-8):707–715. https://doi.org/10.1016/S0300-9084(85)80158-9.

[89] **Ferrenberg AM**, Swendsen RH. Optimized Monte Carlo Data Analysis. Phys Rev Lett. 1989; 63(12):1195–1198. https://doi.org/10.1103/PhysRevLett.63.1195.

[90] **Kumar S**, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. J Comput Chem. 1992; 13(8):1011–1021. https://doi.org/10.1002/jcc.540130812.

[91] **Roux B**. The Calculation of the Potential of Mean Force Using Computer Simulations. Comput Phys Commun. 1995; 91(1-3):275–282. https://doi.org/10.1016/0010-4655(95)00053-I.

[92] **Billeter S**, van Gunsteren W. Computer Simulation of Proton Transfers of Small Acids in Water. J Phys Chem A. 2000; 104(15):3276–3286. https://doi.org/10.1021/jp994127q.

[93] **Kästner J**, Thiel W. Bridging the Gap Between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: "Umbrella Integration". J Chem Phys. 2005; 123(14):144104. https://doi.org/10.1063/1.2052648.

[94] **Kästner J**, Thiel W. Analysis of the Statistical Error in Umbrella Sampling Simulations by Umbrella Integration. J Chem Phys. 2006; 124(23):234106. https://doi.org/10.1063/1.2206775.

[95] Python Implementation of the Umbrella Integration Method;. Accessed: 2018-01-08. https://github.com/martinstroet/umbrella_integration.

[96] Python Implementation of the Multistate Bennett Acceptance Ratio;. Accessed: 2018-01-08. https://github.com/choderalab/pymbar.

[97] **Baumketner A**. Removing Systematic Errors in Interionic Potentials of Mean Force Computed in Molecular Simulations Using Reaction-Field-Based Electrostatics. J Chem Phys. 2009; 130(10):104106. https://doi.org/10.1063/1.3081138.

[98] **Gilson MK**, Irikura KK. Correction to "Symmetry Numbers for Rigid, Flexible, and Fluxional Molecules: Theory and Applications". J Phys Chem B. 2013; 117(10):3061–3061. https://doi.org/10.1021/jp401194k.

[99] **Reif MM**, Hünenberger PH. Computation of Methodology-Independent Single-Ion Solvation Properties from Molecular Simulations. IV. Optimized Lennard-Jones Interaction Parameter Sets for the Alkali and Halide Ions in Water. J Chem Phys. 2011; 134(14):144104. https://doi.org/10.1063/1.3567022.

[100] **Gebhardt J**, Hansen N. Calculation of Binding Affinities for Linear Alcohols to α-Cyclodextrin by Twin-System Enveloping Distribution Sampling Simulations. Fluid Phase Equilib. 2016; 422:1–17. https://doi.org/10.1016/j.fluid.2016.02.001.

[101] **Sugita Y**, Kitao A, Okamoto Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. J Chem Phys. 2000; 113(15):6042–6051. https://doi.org/10.1063/1.1308516.

[102] **Mobley DL**, Chodera JD, Dill KA. On the Use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. J Chem Phys. 2006; 125(8):084902. https://doi.org/10.1063/1.2221683.

[103] **Gebhardt J**, Kleist C, Jakobtorweihen S, Hansen N. Validation and Comparison of Force Fields for Native Cyclodextrins in Aqueous Solution. J Phys Chem B. 2018; 122(5):1608–1626. https://doi.org/10.1021/acs.jpcb.7b11808.

[104] **Tirado-Rives J**, Jorgensen WL. Contribution of Conformer Focusing to the Uncertainty in Predicting Free Energies for Protein-Ligand Binding. J Med Chem. 2006; 49(20):5880–5884. https://doi.org/10.1021/jm060763i.

[105] **Singh G**, Chamberlin AC, Zhekova HR, Noskov SY, Tieleman DP. Two-Dimensional Potentials of Mean Force of Nile Red in Intact and Damaged Model Bilayers. Application to Calculations of Fluorescence Spectra. J Chem Theory Comput. 2015; 12(1):364–371. https://doi.org/10.1021/acs.jctc.5b00520.
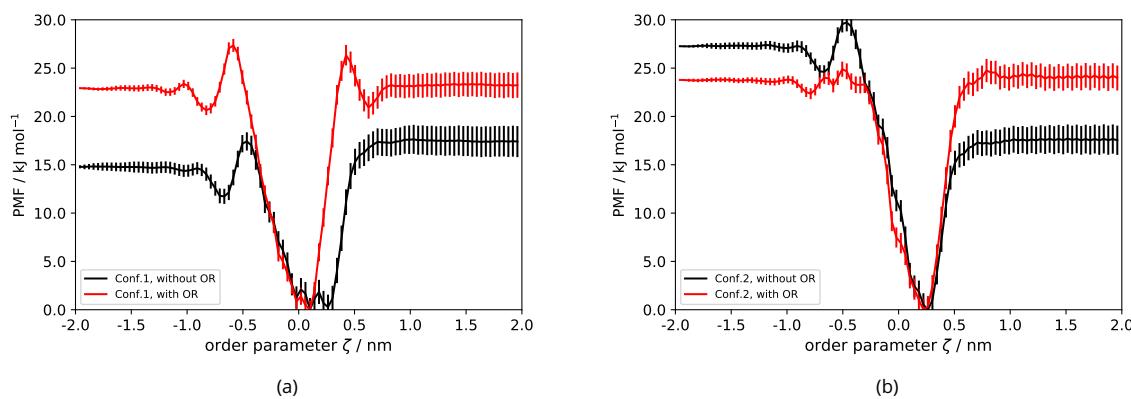
Appendix

Double Decoupling

According to the double decoupling method (DDM), the calculation of the standard binding free enthalpy reads as [15]:

$$\Delta G^{\circ}_{\text{bind}} = \Delta G^{\text{L}\rightarrow\text{D}}_{\text{u}} - \Delta G^{\text{L}}_{\text{b},0\rightarrow\text{R}} - \Delta G^{\text{L}\rightarrow\text{D}}_{\text{b,R}} - \Delta G^{\text{D}}_{\text{b,R}\rightarrow 0} \qquad (13)$$

$\Delta G^{\text{L}\rightarrow\text{D}}_{\text{u}}$ refers to the free energy contribution for transforming the fully interacting unbound ligand (L) into its ideal gas or decoupled (D) state. $\Delta G^{\text{L}\rightarrow\text{D}}_{\text{b,R}}$ represents the analogue contribution for the ligand bound to the CNT host. To prevent drifting of the decoupled ligand, an auxiliary translational (and possibly orientational) restraint (R) has to be applied. The translational and orientational restraints were implemented as harmonic potentials acting on the host-ligand COM-COM radial distance and the orientational angle θ, respectively. $\Delta G^{\text{L}}_{\text{b},0\rightarrow\text{R}}$ and $\Delta G^{\text{D}}_{\text{b,R}\rightarrow 0}$ refer to the contributions due to application and release of the auxiliary restraints for the fully interacting and decoupled ligand in the bound state, respectively. Decoupling of the ligand from the bulk solvent and host was conducted in a sequence of 20 discrete steps as controlled by the coupling parameter λ, equally distributed between $\lambda = 0$ (fully interacting) and $\lambda = 1$ (decoupled state). It should be stressed that since DDM was applied for systems unpolar CNT / unpolar ligand, the scaling with λ solely affects the dispersion interactions with the environment. Activation of the translational restraint in case of the fully interacting bound ligand was performed in 11 distinct simulations using uniformly increasing values for the force constant between 0 and 500 kJ mol^{-1} nm^{-2}. In case of an additional orientational restraint, it was activated simultaneously with the translational restraint using uniformly increasing values for the force constant between 0 and 500 kJ mol^{-1} rad^{-2}. The MBAR free energy estimator was used in all cases. The contribution $\Delta G^{\text{D}}_{\text{b,R}\rightarrow 0}$ was calculated analytically according to [48]:

$$\Delta G^{\text{D}}_{\text{b,R}\rightarrow 0} = -RT \ln\left(\frac{V^{\circ}\, 8\pi^2}{V_{\text{tr}}\,\Omega}\right) \qquad (14)$$

with the accessible translational and rotational volumes of

$$V_{\text{tr}} = \left(\frac{2\pi RT}{k_{\text{tr}}}\right)^{\frac{3}{2}} \qquad (15)$$

$$\frac{\Omega}{8\pi^2} = \frac{1}{2}\int_{0}^{\pi} e^{-U_\theta(\theta)/RT}\,\sin\theta\,\mathrm{d}\theta \qquad (16)$$

In case of a harmonic potential $U_\theta(\theta)$ according to Eq. (11), the rotational volume V_{rot} was calculated numerically while it reduces to unity in the absence of an orientational restraint. Calculated binding free enthalpies from DDM for systems unpolar ligand / unpolar CNT are summarized in Tab. A1. In all simulations, long-range electrostatics were treated with the particle-mesh Ewald (PME) method.

Appendix Table A1. Calculated standard binding free enthalpies ΔG°_{bind} from double decoupling for unpolar methane, ethane and elongated ethane binding to unpolar CNT. For each ethane model two data sets are presented, corresponding to the setup with and without orientational restraint (OR). Detailed description of the double decoupling approach can be found in the appendix. $\Delta G^{L\to D}_u$, $\Delta G^{L\to D}_{b,R}$ and $\Delta G^L_{b,0\to R}$ were calculated using the MBAR estimator. The contribution for removing the restraints from the decoupled ligand ($\Delta G^D_{b,R\to 0}$) was calculated analytically according to Eq. (14). The estimate of $\Delta G^\circ_{bind,Conf.1}$ as obtained from the setup including an orientational restraint corresponds to one distinct binding configuration and has to be corrected by an entropic symmetry term of $-RT \ln 2$ [48, 98] to obtain ΔG°_{bind} in case of elongated ethane. Estimates for statistical uncertainties of ΔG°_{bind} as obtained from application of standard error propagation to Eq. 13 are below 0.5 kJ mol^{-1} where the statistical uncertainties of the individual free energy terms are delivered by the MBAR estimator [47].

System	Setup	$\Delta G^{L\to D}_u$ [kJ mol^{-1}]	$\Delta G^{L\to D}_{b,R}$ [kJ mol^{-1}]	$\Delta G^L_{b,0\to R}$ [kJ mol^{-1}]	$\Delta G^D_{b,R\to 0}$ [kJ mol^{-1}]	$\Delta G^\circ_{bind,Conf.1}$ [kJ mol^{-1}]	ΔG°_{bind} [kJ mol^{-1}]
Methane		-9.60	13.78	3.86	-14.22	-	-13.00
Ethane	No OR	-7.37	30.82	2.38	-14.22	-	-26.35
	OR	-7.37	31.99	17.03	-29.15	-27.23	-27.23
Long Ethane	No OR	-22.67	14.32	2.32	-14.22	-	-25.09
	OR	-22.67	22.23	8.31	-29.15	-24.06	-25.79

(a)   (b)

Appendix Figure A1. PMFs for 1-butanol binding to αCD in Conf. 1 (a) and Conf. 2 (b). Red and black profiles refer to the setup with and without imposed orientational restraint (OR), respectively. Simulations are based on the CHARMM36 all-atom force field.