

# Toward Automated Free Energy Calculation with Accelerated Enveloping Distribution Sampling (A-EDS)

Jan Walther Perthold, Dražen Petrov, and Chris Oostenbrink\*

Cite This: *J. Chem. Inf. Model.* 2020, 60, 5395–5406

Read Online

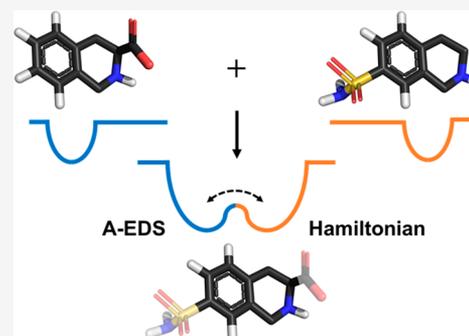
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Free-energy perturbation (FEP) methods are commonly used in drug design to calculate relative binding free energies of different ligands to a common host protein. Alchemical ligand transformations are usually performed in multiple steps which need to be chosen carefully to ensure sufficient phase-space overlap between neighboring states. With one-step or single-step FEP techniques, a single reference state is designed that samples phase-space not only representative of a full transformation but also ideally resembles multiple ligand end states and hence allows for efficient multistate perturbations. Enveloping distribution sampling (EDS) is one example for such a method in which the reference state is created by a mathematical combination of the different ligand end states based on solid statistical mechanics. We have recently proposed a novel approach to EDS which enables efficient barrier crossing between the different end states, termed accelerated EDS (A-EDS). In this work, we further simplify the parametrization of the A-EDS reference state and demonstrate the automated calculation of multiple free-energy differences between different ligands from a single simulation in three different well-described drug design model systems.



## INTRODUCTION

With the paradigm shift from the rigid lock-and-key hypothesis of receptor–ligand recognition to more dynamic conceptions of molecular interactions involving structural ensembles rather than single conformations and recent advances in computer technology, molecular dynamics (MD) simulations are being used more and more in the field of drug design. Specifically, MD-based methods to efficiently calculate relative drug–target binding free energies are commonly used.<sup>1</sup> Calculations involving transformations of one ligand into other ligands via unphysical pathways are often referred to as “alchemical” changes. If such transformations are conducted when the ligand is bound to the receptor and when free in solution, free-energy differences obtained from these transformations can be translated into relative binding free energies using thermodynamic cycles.<sup>2</sup> A plethora of methods is available to perform alchemical transformations;<sup>3</sup> however, free-energy perturbation (FEP) approaches involving many unphysical intermediate states are most commonly applied.<sup>1,4</sup> Recently, FEP-based methods have been shown to allow for the calculation of relative ligand binding free energies with remarkable accuracy on large scales.<sup>5–7</sup> FEP approaches are based on Zwanzig’s equation described more than six decades ago,<sup>8</sup> which relates the free-energy differences between two states,  $\Delta F_{2,1}$ , to the exponential average of the energy difference given by their Hamiltonians  $H_1(\vec{r})$  and  $H_2(\vec{r})$ , obtained from an ensemble of the positions  $\vec{r}$  of one of the states ( $R$  is the ideal gas constant and  $T$  is the temperature):

$$\Delta F_{2,1} = -RT \ln \langle e^{-(H_2(\vec{r}) - H_1(\vec{r}))/RT} \rangle_1 \quad (1)$$

Zwanzig’s equation only gives unbiased free-energy estimates if there is sufficient overlap in the phase-space that is sampled by both states, which is the reason for the introduction of unphysical intermediate states in practice. However, it is intriguing to think about possibilities to design a single reference state which does not only allow for single-step perturbations, but also for the accurate representation of multiple end states simultaneously, thereby greatly enhancing the efficiency of the FEP approach. One-step perturbation (OSP) methods make use of empirically tuned reference-state Hamiltonians  $H_R(\vec{r})$  which often involve “soft” atoms to accurately represent multiple end states.<sup>9–14</sup> However, OSP approaches suffer from limited transferability of the designed Hamiltonians, i.e. a different perturbation might require empirical reparametrization of the reference-state Hamiltonian.<sup>15</sup> Moreover, the soft-core approach does not work very well for larger changes in the charge distributions<sup>16,17</sup> and additional end state simulations are often needed to calculate correct contributions of the electrostatic potential energy to the free-energy differences.<sup>18–20</sup> A different

**Special Issue:** Novel Directions in Free Energy Methods and Applications

**Received:** April 30, 2020

**Published:** June 3, 2020



approach to elegantly calculate a reference Hamiltonian is enveloping distribution sampling (EDS).<sup>21</sup> Here,  $n$  different Boltzmann-weighted end-state Hamiltonians  $H_i(\vec{r})$  are combined in the reference Hamiltonian  $H_R(\vec{r})$ , the partition function of which is now the sum of the partition functions of the end states. Additionally, to ensure equal sampling of the end states, constant free-energy offset parameters  $\Delta F_i^R$  are added:<sup>21–24</sup>

$$H_R(\vec{r}) = -RT \ln \left( \sum_{i=1}^n e^{-(H_i(\vec{r}) - \Delta F_i^R)/RT} \right) \quad (2)$$

Here, we call the offset parameters  $\Delta F_i^R$  *free-energy* offset parameters, because they correspond to the relative free-energies of the end states if all states are sampled with equal probability. However, eq 2 was found to result in high energy-barriers, preventing efficient sampling of the minima of all end states. An additional smoothening parameter  $s$  was proposed to smoothen the energy landscape of the reference Hamiltonian.<sup>22,25–29</sup> However, for the small values of  $s$  that are often used, the energy minima of the reference Hamiltonian no longer correspond to the energy minima of the original end states, reducing the phase-space overlap and further complicating the search for the optimal values of  $s$  and  $\Delta F_i^R$ .<sup>28,30–34</sup> A computationally expensive approach to bypass this problem uses replica-exchange simulations, in which multiple replicas are performed at different values of  $s$ .<sup>35,36</sup>

We have very recently proposed an alternative EDS formulation in which the EDS reference-state Hamiltonian given in eq 2 is smoothed with a harmonic potential energy function, similar to the accelerated MD approach proposed by McCammon and co-workers.<sup>37,38</sup> Our accelerated EDS (A-EDS) approach notably preserves local energy minima and the energy landscape around them, while simultaneously decreasing the energy barriers between the end states. Moreover, it allows for intuitive and potentially automated tuning of the EDS reference-state Hamiltonian parameters by setting an anticipated value for the maximum energy barrier between the end-states as the only search parameter.<sup>39</sup>

In this work, we propose an improved algorithm for the A-EDS Hamiltonian parameter search which further simplifies the construction of the reference-state Hamiltonian. With our

updated search scheme, it is no longer necessary to give a certain value for the anticipated maximum energy barrier between the end states. The search for the optimal acceleration parameters is instead based on a given acceleration factor. This acceleration factor is translated to the probability of reaching the maximum transition-state energy  $E^\ddagger$ , estimated from the average energy and energy fluctuation of each end state. Hence, the user specifies the likelihood of a state transition as search parameter and does not need to have any knowledge about the structure and magnitudes of the energy landscape of the nonaccelerated EDS Hamiltonian before the simulation. Additionally, we improved the fully automated search for the free-energy offset parameters  $\Delta F_i^R$  by considering the deformation (flattening) of the original end-state Hamiltonians upon acceleration and by using a memory relaxation time for their calculation, which allows for faster adjustment of the parameters.

To demonstrate the applicability of our proposed method to automatically calculate multiple free-energy differences from a single simulation, three protein–ligand systems were chosen. All three systems are already established in the literature and the drug–target interactions were examined both using experimental and computational methods:

- 1 glutamate receptor A2 (GRA2) allosteric modulators<sup>14,39</sup>
- 2 trypsin (TRP) inhibitors<sup>19,40–47</sup>
- 3 phenylethanolamine *N*-methyltransferase (PNMT) inhibitors<sup>32,35,36,48</sup>

Notably, systems 2 and 3 both include net-charge changes in the ligand sets, which were previously considered too challenging for multistate EDS approaches.<sup>35,36</sup>

## ■ THEORY

**A-EDS Hamiltonian.** In A-EDS the reference Hamiltonian  $H_R(\vec{r})$  given in eq 2 is smoothed by an approach that is similar to Gaussian accelerated MD (GAMD),<sup>38</sup> which is a variant of the general accelerated MD approach.<sup>37</sup> A harmonic potential energy term is added to the EDS reference Hamiltonian, pulling the energy landscape down to a minimum energy value  $E_{\min}$ . The continuous accelerated EDS (A-EDS) reference Hamiltonian  $H_R^*(\vec{r})$  is defined as

$$H_R^*(\vec{r}) = \begin{cases} H_R(\vec{r}) - \frac{E_{\max} - E_{\min}}{2} & H_R(\vec{r}) \geq E_{\max} \\ H_R(\vec{r}) - \frac{1}{2(E_{\max} - E_{\min})} (H_R(\vec{r}) - E_{\min})^2 & E_{\min} < H_R(\vec{r}) < E_{\max} \\ H_R(\vec{r}) & H_R(\vec{r}) \leq E_{\min} \end{cases} \quad (3)$$

Where,  $E_{\min}$  and  $E_{\max}$  are tunable parameters, in addition to the  $\Delta F_i^R$  in eq 2. Effectively,  $H_R(\vec{r})$  is only modified in the region between  $E_{\min}$  and  $E_{\max}$ . The A-EDS functional form preserves the energy landscape around local energy minima.

**A-EDS Parameter Search.** In our earlier work, we have proposed an intuitive parameter search,<sup>39</sup> which is simplified further in the current work. During a parameter-search simulation, the end state  $i$  which is currently being sampled in a simulation of the EDS reference state is defined as the state with the lowest energy  $H_i(\vec{r}) - \Delta F_i^R$ . We monitor the average

value of  $H_i(\vec{r}) - \Delta F_i^R$  while the state is sampled, denoted as  $\bar{E}_i$ , as well as the value of  $H_R(\vec{r})$  when a new state is sampled, indicative of the barrier height and denoted as  $H_R^\ddagger(\vec{r})$ . From the values of  $H_R^\ddagger(\vec{r})$  we compute  $\bar{E}_{\max}^\ddagger$  as the average of the maximum transition energy between states within a state round-trip (we define a round-trip as having visited all states at least once). With these properties we define the maximum energy barrier  $\Delta E_{\max}$  between the end states in the unmodified EDS reference Hamiltonian  $H_R(\vec{r})$ :

$$\begin{aligned}\bar{E}_i &= \langle H_i(\bar{r}) - \Delta F_i^R \rangle_{R,i} \\ \bar{E}_{\max}^\ddagger &= \langle \max(H_R^\ddagger(\bar{r})) \rangle_{R, \text{round-trips}} \\ \Delta E_{\max} &= \bar{E}_{\max}^\ddagger - \min(\bar{E}_i)\end{aligned}\quad (4)$$

$H_R^\ddagger(\bar{r})$  is the value of  $H_R(\bar{r})$  when a new state is sampled. The unmodified EDS Hamiltonian  $H_R(\bar{r})$  can now be accelerated by setting  $E_{\min}$  and  $E_{\max}$  such that the maximum energy barrier between the states,  $\Delta E_{\max}$ , is reduced to the target value  $\Delta E_{\max}^*$  in  $H_R^*(\bar{r})$ . The target value  $\Delta E_{\max}^*$  is set to a user-defined multiple  $z$  of the standard deviation,  $\sigma_{H_i(\bar{r})}$ , of the energy  $H_i(\bar{r})$  of the state with the lowest average energy  $\bar{E}_i$ . Like for the average end state energies  $\bar{E}_i$ , all  $\sigma_{H_i(\bar{r})}$  are only calculated while the respective state is sampled. By setting the acceleration as sigma-level  $z$ , the minimum probability of reaching the maximum transition-energy,  $P_{\min}(\bar{E}_{\max}^\ddagger)$ , becomes for Gaussian end state energy distributions:

$$P_{\min}(\bar{E}_{\max}^\ddagger) = \frac{1}{2} \left[ 1 - \operatorname{erf} \left( \frac{z}{\sqrt{2}} \right) \right] \quad (5)$$

where  $\operatorname{erf}()$  is the error function. Hence, the probability of reaching the maximum transitions-energy is decreased exponentially with increasing sigma-level  $z$ .

The value of  $\Delta E_{\max}^*$  that is obtained in this way, is subsequently used to determine  $E_{\min}$  and  $E_{\max}$  in eq 3.  $E_{\max}$  simply corresponds to  $\bar{E}_{\max}^\ddagger$ , and  $E_{\min}$  is calculated as

$$E_{\max} = \bar{E}_{\max}^\ddagger$$

$$E_{\min} = \begin{cases} E_{\min} = E_{\max} & \Delta E_{\max}^* \geq \Delta E_{\max} \\ E_{\min} = 2(\min(\bar{E}_i) + \Delta E_{\max}^*) - E_{\max} & \Delta E_{\max}^* < \Delta E_{\max} \end{cases} \quad (6)$$

If the resulting value of  $E_{\min}$  is less than  $\min(\bar{E}_i)$ , it is adjusted to

$$H_i^*(\bar{r}) = \begin{cases} H_i(\bar{r}) - \frac{E_{\max} - E_{\min}}{2} & H_i(\bar{r}) \geq E_{\max} + \Delta F_i^R \\ H_i(\bar{r}) - \frac{1}{2(E_{\max} - E_{\min})} (H_R(\bar{r}) - E_{\min} - \Delta F_i^R)^2 & E_{\min} + \Delta F_i^R < H_i(\bar{r}) < E_{\max} + \Delta F_i^R \\ H_i(\bar{r}) & H_i(\bar{r}) \leq E_{\min} + \Delta F_i^R \end{cases} \quad (9)$$

To allow for faster adjustment of the free-energy offset parameters  $\Delta F_i^R$  during the parameter-search simulation, we implemented a memory relaxation time  $\tau$  for the exponential averages of the energy differences in analogy to time-averaged distance restraining functions.<sup>49</sup> Accordingly, the exponential energy differences with a characteristic memory decay time  $\tau$  read ( $t$  is the simulation time and  $\Delta t$  the time-step):

$$\begin{aligned}\langle e^{-(H_i^*(\bar{r}) - H_R^*(\bar{r}))/RT} \rangle_t &= \\ (1 - e^{-\Delta t/\tau}) e^{-(H_i^*(\bar{r}) - H_R^*(\bar{r}))/RT}(t) + e^{-\Delta t/\tau} & \\ \langle e^{-(H_i^*(\bar{r}) - H_R^*(\bar{r}))/RT} \rangle_{t-\Delta t} &\end{aligned} \quad (10)$$

Additionally, we increase  $\tau$  linearly over the course of the simulation from a value  $\tau_A$  to a value  $\tau_B$  ( $t_{\text{tot}}$  is the total simulation time of the nonequilibrium parameter search):

$$E_{\min} = \frac{E_{\max} \Delta E_{\max}^* + E_{\max} \min(\bar{E}_i) - \frac{1}{2} E_{\max}^2 - \frac{1}{2} \min(\bar{E}_i)^2}{\Delta E_{\max}^*} \quad (7)$$

The proposed optimization scheme for the acceleration parameters of the A-EDS Hamiltonian is used in a non-equilibrium parameter-search simulation in which the user only specifies the desired maximum barrier height  $\Delta E_{\max}^*$  as sigma-level  $z$  which gives a multiple of  $\sigma_{H_i(\bar{r})}$  of the end state with the lowest average energy. No initial values for  $E_{\min}$  and  $E_{\max}$  have to be set, as they are adjusted while the reference state explores the energy landscape. To enable sampling of all states in the beginning of the search simulation, the value of  $E_{\max}$  is set to the maximum transition energy  $E_{\max}^\ddagger$  of the unmodified Hamiltonian  $H_R(\bar{r})$  as the system explores the energy landscape until all states have been visited at least once. After all states have been visited at least once,  $E_{\max}$  is calculated as described above.

In addition to the A-EDS parameters  $E_{\min}$  and  $E_{\max}$ , the free-energy offset parameters  $\Delta F_i^R$  are optimized simultaneously during the parameter-search simulation. The free-energy offset parameter of the first state is arbitrarily set to zero and the other free-energy offset parameters are calculated relative to the first state by simply using the perturbation formula, eq 1:

$$\begin{aligned}\Delta F_1^R &= 0 \\ \Delta F_{i \neq 1}^R &= \Delta F_{i,1} = \Delta F_{i,R} - \Delta F_{1,R} \\ &= -RT \ln \langle e^{-(H_i^*(\bar{r}) - H_R^*(\bar{r}))/RT} \rangle_R \\ &\quad + RT \ln \langle e^{-(H_1^*(\bar{r}) - H_R^*(\bar{r}))/RT} \rangle_R\end{aligned} \quad (8)$$

Note that in contrast to the previously suggested approach,<sup>39</sup> we here consider the accelerated end-state Hamiltonians  $H_i^*(\bar{r})$  instead of the unmodified end-state Hamiltonians to account for deformation (flattening) of the original energy landscapes:

$$\tau = \tau_A + (\tau_B - \tau_A) \frac{t}{t_{\text{tot}}} \quad (11)$$

In systems in which the free energy offset parameters  $\Delta F_i^R$  are very large,  $\tau_A$  can be set to a small value in beginning of the parameter-search simulation to allow for rapid adjustment of all  $\Delta F_i^R$ , while more statistics are used upon convergence toward the end of the parameter-search simulation. Using this approach, the free energy offset parameters  $\Delta F_i^R$  fluctuate around their optimal values which ensure equal sampling of all end states.

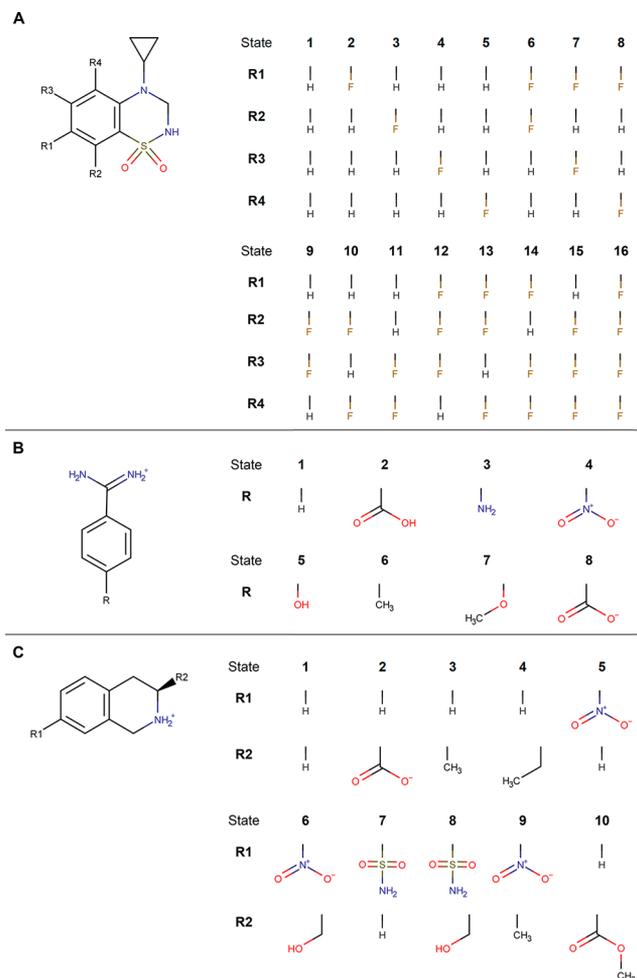
After the nonequilibrium parameter-search simulation, the values of  $E_{\min}$ ,  $E_{\max}$ , and  $\Delta F_i^R$  are fixed and an equilibrium simulation is performed to determine the final free energy differences between the states.

## METHODS

**Simulation Settings.** All MD simulations were performed using the GROMOS MD++ 1.5.0 simulation engine<sup>50</sup> to which the A-EDS functionality was added. All systems were simulated in NVT ensembles at 300 K using Nosé–Hoover chains<sup>51</sup> to ensure canonical energy distributions with 5 coupled thermostats and relaxation times of 0.1 ps. The solute and solvent degrees of freedom were coupled to different heat-baths. The equations of motion were solved using a leapfrog integration scheme<sup>52</sup> with a time-step of 2 fs. Covalent bond lengths of solute molecules were constrained using the SHAKE<sup>53</sup> algorithm, and covalent bond lengths and angles of solvent water molecules were constrained using the SETTLE<sup>54</sup> algorithm. The center-of-mass motions of the systems were removed every 10 000 steps. Neighbor searching was performed every 10 fs using a group-based cutoff scheme within a cutoff sphere of 1.4 nm. For the calculation of nonbonded interactions, a twin-range cutoff scheme was employed. Within a cutoff of 0.8 nm, nonbonded interactions were calculated every step. Between 0.8 and 1.4 nm, nonbonded interactions were evaluated every 10 fs and held constant between the updates. For the calculation of electrostatic interactions, a reaction-field contribution<sup>55</sup> with a relative dielectric permittivity of 61<sup>56</sup> beyond the cutoff sphere was added. For parametrization and preparation of the drug-target model systems, the reader is referred to previous literature.<sup>14,19,36</sup> For the PNMT ligands, small updates were made to the parameters. These involved the introduction of explicit aromatic hydrogens and the more precise definition of integer-charged charge groups. The updated molecular building blocks are available in the [Supporting Information](#). It is noted that for PNMT ligands and TRP ligands additional weak distance restraints with force constants of  $250 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$  were introduced to prevent the reference-state ligands from leaving the binding site during the sampling of end states with unfavorable binding affinity. The additional protein–ligand distance restraint definitions for these systems is given in [Tables S1 and S2](#) in the [Supporting Information](#).

**Perturbation Topology Setup.** For each of the protein systems, multiple end states were considered in the multistate A-EDS calculations, as shown in [Figure 1](#). For GRA2, 16 distinct states were used, which is more than the five end states used in our previous work.<sup>39</sup> For TRP, eight end states were used, including the protonated (state 2) and deprotonated (state 8) carboxylic acid. For PNMT, ten different states were considered, including state 2, which bears a different net-charge from all the other ligands. To combine the different ligand end states given in [Figure 1](#) into a single reference state, a single-topology approach was chosen. This was also shown to work best in [ref 32](#), while a dual topology approach was used in [refs 35 and 36](#).

To simplify the complex setup of such molecules, a recently developed automated approach was used for systems TRP and PNMT, based on maximal common substructure (MCS) search for a pair of molecules. The search involves an iterative procedure in which in each step a pair of atoms, each belonging to one of the compounds, is added to the current common substructure (current solution). Upon adding a pair of atoms, common parts of molecular topologies are checked for mismatching parameters of bonded interactions. As the search was primarily aimed at maximizing the number of common atoms, while keeping all bonded interactions unperturbed, only added pairs without any mismatched parameters of bonded

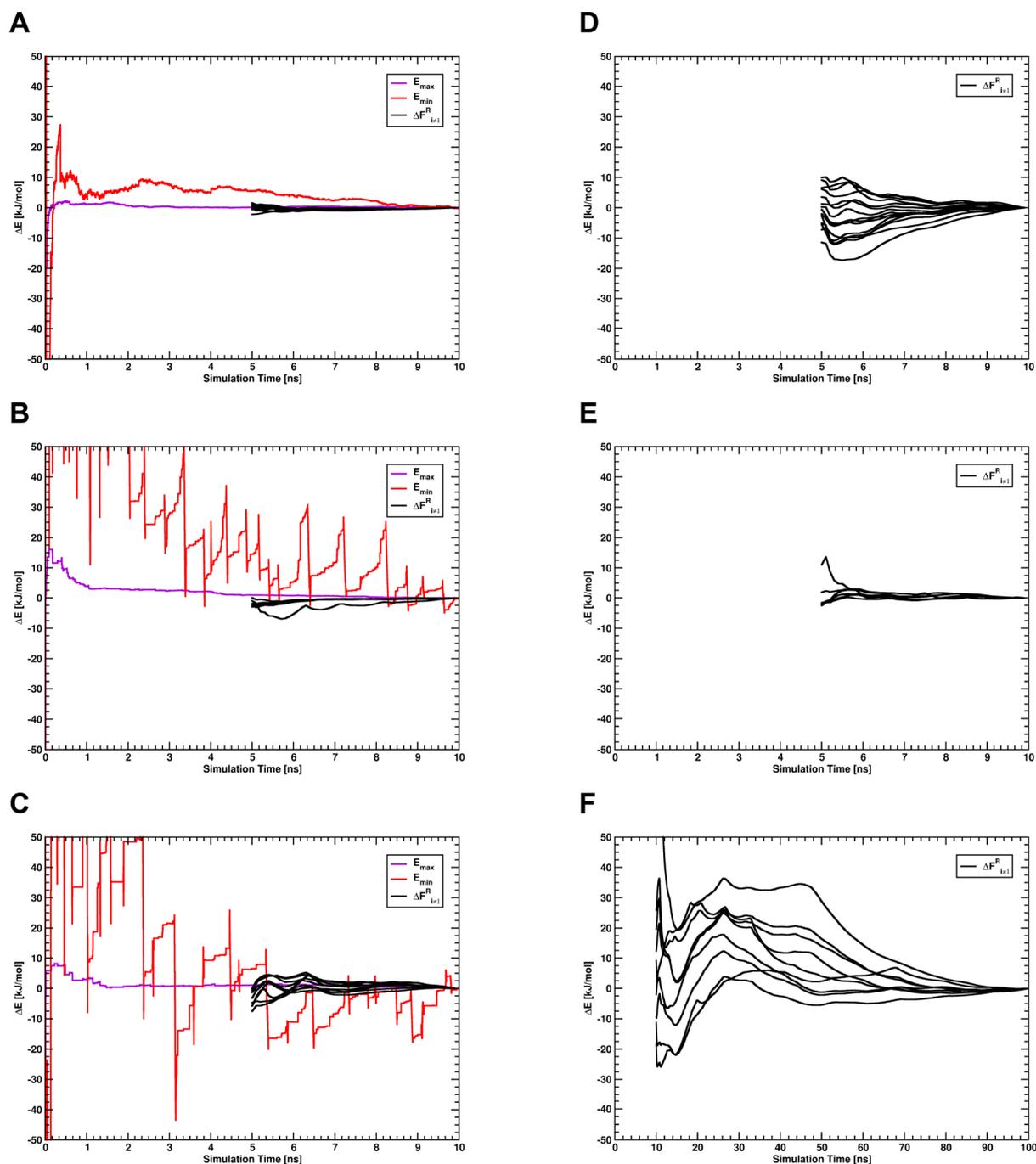


**Figure 1.** Different ligand end states for which relative binding free energies were calculated. Different benzothiadiazine dioxide ligands of system GRA2 (A), different benzamidines ligands of system TRP (B), and the different tetrahydroisoquinoline ligands of system PNMT (C) are shown with their corresponding end state numbering.

interactions were taken into account. To create a perturbation topology from the resulting MCS, atom pairs from the solution common substructure were merged into single atoms with interaction parameters assigned to their respective end-state parameters, while the remaining atoms from both topologies not belonging to the solution common substructure were assigned as noninteracting dummy atoms transforming into interacting atoms (end-state parameters) and the other way around.

To generate a combined multistate single topology from a set of multiple compounds ( $n$  topologies), MCS search (as described above) was performed  $n - 1$  times. An initial MCS search was performed on two arbitrarily chosen compounds from the set, producing a first perturbation topology (representative of the first two end states). Each of the following ( $n - 2$ ) MCS searches were performed on the resulting perturbation topology from the previous MSC step and one of the remaining compounds (topologies) from the set, appending an additional end state in each of the iterations and finally producing the multistate single topology.

**A-EDS Parameter-Search Simulations.** For the calculation of the A-EDS acceleration parameters, nonequilibrium A-EDS parameter-search simulations were first performed for the unbound ligands in water with three different acceleration  $\sigma$ -



**Figure 2.** Convergence of the A-EDS acceleration parameters  $E_{\max}$  and  $E_{\min}$  and the free-energy offset parameters  $\Delta F_{i \neq 1}^R$  relative to their final values for the A-EDS parameter-search simulation with an acceleration  $\sigma$ -level of  $2\sigma$ , for the unbound ligands in water of system GRA2 (A), TRP (B), and PNMT (C), and for the ligands bound to the protein of system GRA2 (D), TRP (E), and PNMT (F). The convergence of the free-energy offset parameters  $\Delta F_{i \neq 1}^R$  is shown as a forward cumulative average over the free-energy offset trajectory with the first 5 ns discarded as the nonequilibrated region. Note that in panels B and C the A-EDS parameter  $E_{\min}$  (red solid line) falls off the plot in the beginning of the trajectory.

levels,  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$ , to calculate the anticipated maximum energy barrier between the end states,  $\Delta E_{\max}^*$  (i.e.,  $z = 1, 2$ , and  $3$ ). In these simulations, A-EDS parameters  $E_{\max}$ ,  $E_{\min}$ , and all free-energy offset parameters  $\Delta F_i^R$  were determined simultaneously, with an increase in the memory relaxation time for the calculation of the free-energy offset parameters  $\Delta F_i^R$  from  $\tau_A = 1$  ps to  $\tau_B = 100$  ps within the parameter-search simulation time of 10 ns. For the parameter-search simulations of the ligands bound to the proteins, A-EDS parameters  $E_{\max}$  and  $E_{\min}$  previously determined for the ligands in water for the three different

acceleration levels were used and held fixed throughout the parameter search. Only the free-energy offset parameters  $\Delta F_{i \neq 1}^R$  were optimized during these parameter-search simulations, with the same relaxation time parameters as used for the ligands in water. For the PNMT system the parameter-search simulations were prolonged to 100 ns with a relaxation time  $\tau_B$  of 1000 ps because only few state transitions were observed during the first 10 ns (see Table S3 in the Supporting Information) and free-energy offset parameters converged more slowly (see Figure 2). Since the free-energy offset parameters fluctuate around their

optimal values during the parameter-search simulations, the values used for the subsequent equilibrium simulations were calculated by averaging over the free-energy offset parameter trajectories of the parameter-search simulations. For systems GRA2 and TRP, the first 5 ns of the simulations were discarded for the averaging as nonequilibrated region. For system PNMT, the first 10 ns of the simulations were discarded for the averaging as nonequilibrated region.

**A-EDS Equilibrium Sampling Simulations.** A-EDS equilibrium simulations for the calculation of relative binding free energies were started from the last snapshots of the parameter-search simulations in water and for the ligands bound to the proteins, respectively. The A-EDS parameters determined during the parameter-search simulations were held fixed and sampling was performed for another 10 ns in water and for 100 ns (systems GRA2 and TRP) or 200 ns (system PNMT) for the protein systems.

## RESULTS AND DISCUSSION

**A-EDS Parameter-Search Simulations.** To determine the A-EDS acceleration parameters  $E_{\max}$  and  $E_{\min}$ , and the free-energy offset parameters  $\Delta F_i^R$ , parameter-search simulations were performed for 10 ns for the ligands solvated in water. The resulting acceleration parameters are given in Table 1 for all

**Table 1. A-EDS Acceleration Parameter-Search Results Obtained from the Simulations of the Unbound Ligands in Water for the Three Different Acceleration  $\sigma$ -Levels<sup>a</sup>**

system	parameter	1 $\sigma$	2 $\sigma$	3 $\sigma$
GRA2	$E_{\max}$	158.5	97.2	93.6
	$\Delta E_{\max}^*$ (i.e., $z\sigma$ )	19.2	27.0	34.5
	$E_{\max}$	119.2	94.7	81.4
	$E_{\min}$	-535.0	-150.6	-44.1
	$\Delta E_{\max}^*$	391.4	472.2	486.2
TRP	$\Delta E_{\max}^*$ (i.e., $z\sigma$ )	132.1	209.5	240.2
	$E_{\max}$	33.7	36.1	36.5
	$E_{\min}$	-546.1	-496.1	-455.6
	$\Delta E_{\max}^*$	300.8	506.9	538.4
	$\Delta E_{\max}^*$ (i.e., $z\sigma$ )	35.6	216.5	261.6
PNMT	$E_{\max}$	30.6	32.9	31.0
	$E_{\min}$	-1240.8	-560.4	-523.2
	$\Delta E_{\max}^*$			
	$\Delta E_{\max}^*$ (i.e., $z\sigma$ )			
	$E_{\max}$			

<sup>a</sup> $\Delta E_{\max}^*$  and  $\Delta E_{\min}^*$  are determined from the simulation and define the values of  $E_{\max}$  and  $E_{\min}$ . Energy units are kilojoules per mole.

systems. Depending on the chosen acceleration level the resulting A-EDS parameter  $E_{\min}$ , controlling the range over which the energy landscape is flattened, varies greatly. The parameters  $E_{\max}$ , which correspond to the maximum energy barrier between the states, are comparably independent of the chosen acceleration level. The convergence of all parameters for all ligand systems in water for an acceleration  $\sigma$ -level of 2 $\sigma$  are given in Figure 2A–C, and for the other acceleration levels in the Supporting Information, Figure S1 and S2. Parameters mostly are converged or only change little with simulation time. The greatest fluctuation occurs for the acceleration parameter  $E_{\min}$ , which can be explained by the dependence of this parameter on the slower-converging energy fluctuation estimate for each end state. However, this fluctuation does not have a big influence on the calculated free-energy offset parameters  $\Delta F_i^R$ .

The data in Table 1 nicely demonstrates how the selected acceleration level determines the values of  $E_{\min}$  and  $E_{\max}$ . Setting the acceleration level to 1 $\sigma$  implies requesting a large number of

transitions between the states. Accordingly, the acceleration is strongest, leading to the largest range of acceleration (difference  $E_{\max} - E_{\min}$  is largest). In contrast, setting the acceleration level to 3 $\sigma$  allows for fewer transitions, higher energy barriers and hence to a reduced range of acceleration. As the fluctuations in the energy basins are system dependent, the resulting values of  $E_{\max} - E_{\min}$  differ considerably between the different cases.

To calculate the free-energy offset parameters  $\Delta F_i^R$  for the end states of the ligands bound to the proteins, A-EDS acceleration parameters  $E_{\max}$  and  $E_{\min}$  were held fixed during the parameter-search simulation in the protein systems, as convergence of these parameters was very slow in preliminary simulations, given the generally complicated and possibly non-Gaussian energy landscapes of host–guest systems, where the EDS region is not surrounded by a homogeneous and fast-relaxing medium, but by the protein environment. In our previous work, we determined that the free-energy differences are quite robust with respect to the exact choice of  $E_{\min}$  and  $E_{\max}$ .<sup>39</sup> Therefore, the acceleration parameters for each acceleration level were set to the values determined during the parameter-search simulations in water. The convergence of the free-energy offset parameters for all systems with an acceleration  $\sigma$ -level of 2 $\sigma$  is shown in Figure 2D–F and for the other acceleration levels in the Supporting Information, Figure S1 and S2. For systems GRA2 and PNMT, parameter-search simulations in the protein with an acceleration level of 1 $\sigma$  (very strong acceleration) became unstable. In particular, the SHAKE algorithm failed at maintaining the bond-length constraints, indicative of the occurrence of large forces due to too strong acceleration. Subsequent simulations in the protein using this acceleration level were not performed for these systems. Moreover, for system PNMT, the convergence of the free-energy offset parameters  $\Delta F_i^R$  was slower than for the other systems, and the parameter search was extended to 100 ns. All final free-energy offset parameters  $\Delta F_i^R$  are given in the Supporting Information, Tables S7–S9. The free-energy offset values can depend strongly on the chosen acceleration level, as the flattening of the energy landscape contributes to the free-energy differences between the accelerated end states (eq 9) that are combined to the single reference state in the simulation.

**A-EDS Equilibrium Simulations.** To calculate relative binding free energies for the ligands, A-EDS equilibrium simulations were performed both in water and in the protein-bound state for all systems with the A-EDS parameters determined for each acceleration level in the parameter-search simulations. The free-energy differences between the ligands in each state were calculated from the free-energy difference of each state to the A-EDS reference state obtained from Zwanzig's equation (eq 1). The effects of any distance restraints added in the simulations (Tables S1, S2) on the free energy were removed in this stage, by adding the restraining energies to the reference state Hamiltonians.

For system GRA2, all calculated free-energy differences for the unbound ligands were very similar for all three acceleration levels (Table 2) and converged very rapidly (Figure S3A–C in the Supporting Information). However, for the bound ligands, different results were obtained for acceleration levels of 2 $\sigma$  and 3 $\sigma$ , despite apparent convergence of the calculated free energy values (Figure S3D, E in the Supporting Information). Simulations with an acceleration level of 1 $\sigma$  were not performed in the protein, as simulations became unstable during the parameter-search simulation with this acceleration level. Analysis of the sampled states and state transitions in these

**Table 2. Relative Free-Energy Differences  $\Delta F_{i\neq 1,1}$  between the Ligand End States for the Three Different Acceleration  $\sigma$ -Levels for System GRA2<sup>a</sup>**

	$\Delta F_{i\neq 1,1}$ unbound ligand			$\Delta F_{i\neq 1,1}$ two bound ligands			calculated $\Delta\Delta F_{i\neq 1,1}$		
	$1\sigma$	$2\sigma$	$3\sigma$	$1\sigma$	$2\sigma$	$3\sigma$	$1\sigma$	$2\sigma$	$3\sigma$
$\Delta F_{2,1}$	-111.2 ± 0.3	-111.0 ± 0.2	-110.8 ± 0.1	n.a.	-224.4 ± 1.6	-231.4 ± 1.2	n.a.	-2.4 ± 1.7	-9.8 ± 1.0
$\Delta F_{3,1}$	22.6 ± 0.2	22.5 ± 0.2	22.4 ± 0.1	n.a.	45.5 ± 1.5	39.6 ± 1.4	n.a.	0.5 ± 1.5	-5.3 ± 1.2
$\Delta F_{4,1}$	8.9 ± 0.2	8.6 ± 0.1	8.7 ± 0.1	n.a.	7.3 ± 1.7	24.8 ± 2.3	n.a.	-9.9 ± 1.7	7.5 ± 2.1
$\Delta F_{5,1}$	-53.6 ± 0.3	-54.3 ± 0.2	-54.4 ± 0.1	n.a.	-116.5 ± 1.5	-118.0 ± 0.6	n.a.	-7.8 ± 1.6	-9.3 ± 0.4
$\Delta F_{6,1}$	-90.6 ± 0.3	-90.9 ± 0.2	-90.7 ± 0.1	n.a.	-187.5 ± 1.6	-197.0 ± 1.3	n.a.	-5.6 ± 1.6	-15.6 ± 1.1
$\Delta F_{7,1}$	-104.3 ± 0.2	-104.8 ± 0.2	-104.5 ± 0.1	n.a.	-218.1 ± 1.7	-207.1 ± 1.8	n.a.	-8.6 ± 1.7	2.0 ± 1.6
$\Delta F_{8,1}$	-135.0 ± 0.2	-135.7 ± 0.2	-135.8 ± 0.1	n.a.	-289.5 ± 1.6	-292.5 ± 1.3	n.a.	-18.1 ± 1.6	-20.9 ± 1.1
$\Delta F_{9,1}$	60.3 ± 0.3	60.5 ± 0.2	60.5 ± 0.1	n.a.	108.8 ± 1.5	124.8 ± 1.9	n.a.	-12.2 ± 1.6	3.9 ± 1.7
$\Delta F_{10,1}$	-73.0 ± 0.3	-73.4 ± 0.2	-73.3 ± 0.1	n.a.	-162.6 ± 1.5	-160.0 ± 1.4	n.a.	-15.8 ± 1.6	-13.3 ± 1.2
$\Delta F_{11,1}$	-47.8 ± 0.3	-48.0 ± 0.2	-48.2 ± 0.1	n.a.	-117.2 ± 1.5	-94.3 ± 1.4	n.a.	-21.1 ± 1.6	2.1 ± 1.2
$\Delta F_{12,1}$	-55.3 ± 0.4	-54.9 ± 0.2	-54.8 ± 0.2	n.a.	-123.3 ± 1.6	-113.1 ± 1.9	n.a.	-13.5 ± 1.6	-3.6 ± 1.5
$\Delta F_{13,1}$	-156.0 ± 0.3	-157.0 ± 0.2	-156.9 ± 0.1	n.a.	-340.2 ± 1.6	-338.5 ± 1.5	n.a.	-26.2 ± 1.7	-24.8 ± 1.3
$\Delta F_{14,1}$	-130.8 ± 0.3	-131.6 ± 0.2	-131.6 ± 0.1	n.a.	-286.9 ± 1.6	-270.5 ± 1.7	n.a.	-23.7 ± 1.6	-7.4 ± 1.5
$\Delta F_{15,1}$	-38.0 ± 0.5	-37.7 ± 0.2	-37.7 ± 0.1	n.a.	-102.3 ± 1.7	-79.1 ± 2.5	n.a.	-27.0 ± 1.7	-3.7 ± 2.3
$\Delta F_{16,1}$	-123.1 ± 0.4	-123.2 ± 0.2	-123.3 ± 0.2	n.a.	-277.7 ± 2.1	-260.3 ± 2.4	n.a.	-31.2 ± 2.1	-13.6 ± 2.0

<sup>a</sup>Energy units are kilojoules per mole. n.a. = not available.**Table 3. Relative Free-Energy Differences  $\Delta F_{i\neq 1,1}$  between the Ligand End States for the Three Different Acceleration  $\sigma$ -Levels for System TRP<sup>a</sup>**

	$\Delta F_{i\neq 1,1}$ unbound ligand			$\Delta F_{i\neq 1,1}$ bound ligand			calculated $\Delta\Delta F_{i\neq 1,1}$		
	$1\sigma$	$2\sigma$	$3\sigma$	$1\sigma$	$2\sigma$	$3\sigma$	$1\sigma$	$2\sigma$	$3\sigma$
$\Delta F_{2,1}$	-160.2 ± 0.8	-163.4 ± 1.2	-161.4 ± 0.8	-150.6 ± 1.5	-148.5 ± 1.2	-160.6 ± 1.3	9.6 ± 1.7	14.9 ± 1.7	0.8 ± 1.5
$\Delta F_{3,1}$	-143.8 ± 0.9	-142.5 ± 0.3	-146.2 ± 1.7	-139.1 ± 0.8	-136.6 ± 0.6	-136.0 ± 0.8	4.7 ± 1.2	5.9 ± 0.7	10.2 ± 1.9
$\Delta F_{4,1}$	70.7 ± 0.3	70.3 ± 0.3	71.0 ± 0.3	68.6 ± 0.3	66.6 ± 0.5	63.3 ± 1.4	-2.1 ± 0.4	-3.7 ± 0.6	-7.7 ± 1.4
$\Delta F_{5,1}$	-171.5 ± 2.5	-166.2 ± 1.0	-166.0 ± 0.9	-158.9 ± 1.2	-161.5 ± 2.1	-158.3 ± 2.5	12.6 ± 2.8	4.7 ± 2.3	7.7 ± 2.7
$\Delta F_{6,1}$	-25.2 ± 0.1	-25.1 ± 0.1	-25.3 ± 0.3	-24.1 ± 0.9	-25.4 ± 1.3	-19.6 ± 0.9	1.1 ± 0.9	-0.3 ± 1.3	5.7 ± 0.9
$\Delta F_{7,1}$	-97.1 ± 0.5	-97.7 ± 0.8	-97.1 ± 0.6	-89.8 ± 0.6	-90.7 ± 0.6	-84.7 ± 1.0	7.3 ± 0.8	7.0 ± 1.0	12.4 ± 1.2
$\Delta F_{8,1}$	-427.4 ± 0.6	-425.8 ± 1.0	-430.6 ± 1.4	-402.6 ± 0.6	-401.1 ± 1.6	-395.6 ± 1.9	24.8 ± 0.8	24.7 ± 1.9	35.0 ± 2.4

<sup>a</sup>Energy units are kilojoules per mole.**Table 4. Relative Free-Energy Differences  $\Delta F_{i\neq 1,1}$  between the Ligand End States for the Three Different Acceleration  $\sigma$ -Levels for System PNMT<sup>a</sup>**

	$\Delta F_{i\neq 1,1}$ unbound ligand			$\Delta F_{i\neq 1,1}$ bound ligand			calculated $\Delta\Delta F_{i\neq 1,1}$		
	$1\sigma$	$2\sigma$	$3\sigma$	$1\sigma$	$2\sigma$	$3\sigma$	$1\sigma$	$2\sigma$	$3\sigma$
$\Delta F_{2,1}$	-376.3 ± 1.7	-439.3 ± 1.4	-440.8 ± 2.3	n.a.	-216.9 ± 2.8	-217.6 ± 2.8	n.a.	222.4 ± 3.1	223.2 ± 3.6
$\Delta F_{3,1}$	5.6 ± 0.2	5.8 ± 0.3	6.1 ± 0.3	n.a.	0.8 ± 2.6	4.2 ± 1.6	n.a.	-5.0 ± 2.6	-1.9 ± 1.6
$\Delta F_{4,1}$	8.7 ± 0.3	7.6 ± 0.3	8.4 ± 0.3	n.a.	-6.4 ± 2.9	4.7 ± 1.6	n.a.	-14.0 ± 2.9	-3.7 ± 1.6
$\Delta F_{5,1}$	7.8 ± 0.4	7.9 ± 0.3	8.1 ± 0.3	n.a.	-2.0 ± 2.5	-2.4 ± 1.4	n.a.	-9.9 ± 2.5	-10.5 ± 1.4
$\Delta F_{6,1}$	-24.5 ± 0.9	-29.1 ± 0.8	-30.0 ± 0.9	n.a.	-50.8 ± 2.2	-35 ± 1.5	n.a.	-21.7 ± 2.3	-5.0 ± 1.7
$\Delta F_{7,1}$	-448.4 ± 0.9	-463.0 ± 1.7	-461.3 ± 1.3	n.a.	-465.4 ± 1.9	-475.8 ± 1.5	n.a.	-2.4 ± 2.5	-14.5 ± 2.0
$\Delta F_{8,1}$	-482.1 ± 1.0	-498.0 ± 1.1	-501.0 ± 1.2	n.a.	-526.7 ± 2.0	-510.6 ± 1.3	n.a.	-28.7 ± 2.3	-9.6 ± 1.8
$\Delta F_{9,1}$	13.9 ± 0.4	13.9 ± 0.3	13.8 ± 0.5	n.a.	-0.5 ± 2.1	5.6 ± 1.5	n.a.	-14.4 ± 2.1	-8.2 ± 1.6
$\Delta F_{10,1}$	-161.9 ± 1.3	-163.1 ± 0.6	-163.8 ± 0.5	n.a.	-173.2 ± 2.8	-167.3 ± 2.6	n.a.	-10.1 ± 2.9	-3.5 ± 2.6

<sup>a</sup>Energy units are kilojoules per mole. n.a. = not available.

equilibrium simulations revealed that for an acceleration level of  $3\sigma$ , many states were not sampled adequately and state transitions to these states occurred only rarely (Table 5 and Figure S6 in the Supporting Information).

For system TRP, the calculated free-energy differences for all ligands in the unbound state were again similar for all three acceleration levels. However, for the bound ligands more different values were obtained, especially for the acceleration level of  $3\sigma$  compared to  $1\sigma$  or  $2\sigma$  (Table 3). For  $3\sigma$ , convergence of the calculated values was slower both in water and in the

protein (Figure S4 in the Supporting Information) and much less state transition was observed (Table 5 and Figure S7 in the Supporting Information).

For system PNMT, the acceleration level of  $1\sigma$  (very strong acceleration) lead to partially different calculated free-energy differences between the ligand end states in water compared to the other acceleration levels (Table 4). The difference was especially pronounced for the end state involving a charge-change (ligand 2), but also present for ligands 7 and 8, which have a large sulfonamide group at the R1 position. Possibly, the

**Table 5. Number of Unique Visits of All Ligand End States and of the Ligand End State with the Longest Average Residence Time in Water for the Three Different Acceleration  $\sigma$ -Levels in the Equilibrium A-EDS Simulations**

system	unbound ligands			bound ligands		
	$1\sigma$	$2\sigma$	$3\sigma$	$1\sigma$	$2\sigma$	$3\sigma$
GRA2-all	113081	121863	129663	n.a. <sup>a</sup>	78277	14396
GRA2-1	7084	6366	6802	n.a.	2636	21
TRP-all	97020	99912	83688	249119	198880	75610
TRP-8	1328	427	128	2821	1446	15
PNMT-all	31382	18133	17138	n.a.	18741	10678
PNMT-2	5004	135	48	n.a.	11	2

<sup>a</sup>n.a. = not available.

strong acceleration parameters determined for  $1\sigma$  lead to overflattening of the energy landscape, which prohibits proper sampling of the most important regions of phase-space for these end states. For the ligands bound to the protein, sometimes different free-energy differences between the states were obtained for acceleration levels of  $2\sigma$  and  $3\sigma$ .

Except for the differences described above, the calculated free-energy differences between the end-states are independent of the chosen acceleration level. Notably, they do differ significantly from the corresponding free-energy offset parameters  $\Delta F_i^R$  (Tables S7–S9 in the Supporting Information), as the offset parameters describe the free-energy differences between the accelerated end states in the reference state (eq 9), while the free-energy differences of the end states,  $\Delta F_{i \neq 1,1}$ , are calculated for the original (i.e., nonaccelerated) Hamiltonians.

Table 5 collects the total number of unique visits of the individual states for all simulations. The state with the longest average residence time (the fewest transitions) is also listed, to give a lower bound to these values. It is most relevant to focus on the compounds that are sampled least to judge if the sampling is sufficient. A large number of unique samples is necessary to ensure proper sampling of each of the states. For the compounds in water, the total number of unique visits of all compounds is relatively independent of the acceleration level. Only for PNMT, with the largest differences between states, we observe a significantly lower value at  $2\sigma$  and  $3\sigma$ . The differences are more pronounced for the states that are sampled least, with a clearer drop of the number of visits for TRP and PNMT. In system GRA2, the ligands are arguably the most similar and there is no large difference observed. In the protein-bound states, the number of unique visits is much lower throughout, especially considering that the simulations are a factor 10 (GRA2, TRP) or 20 (PNMT) longer than the ones of the ligands in water. This is representative of the fact that the sampling is much more complex in a protein environment. During the parameter search for GRA2 and PNMT, the simulations at  $1\sigma$  became unstable, suggesting a too strong acceleration. On the other hand, the number of unique visits at  $3\sigma$  drops down considerably, suggesting that this acceleration level, which is most conservative and therefore possibly preferred, should only be used in conjunction with longer simulation times.

**Comparison to Experiment and Other Methods.** The binding free energies in Tables 2–4 are reported relative to the arbitrarily chosen first compound of the set. To compare results for the computed relative binding free energies independent of an arbitrary selection, an estimate of the binding free energy of the A-EDS reference state (average difference of the relative binding free energies to the experimental binding free energies) was used to calculate ligand binding free energies which can directly be compared to experimentally determined values; i.e.,

the relative binding free energy values were shifted by the average deviation to the experimental binding free energies to account for the binding free energy of the reference state.<sup>19,57</sup> This corresponds to a single fitting parameter to compare relative binding free energies to the experimental values. Tables S3–S5 report the individual binding affinities. The overall performance is summarized by computing the root-mean-square error (RMSE) with respect to the experimental values (Table 6).

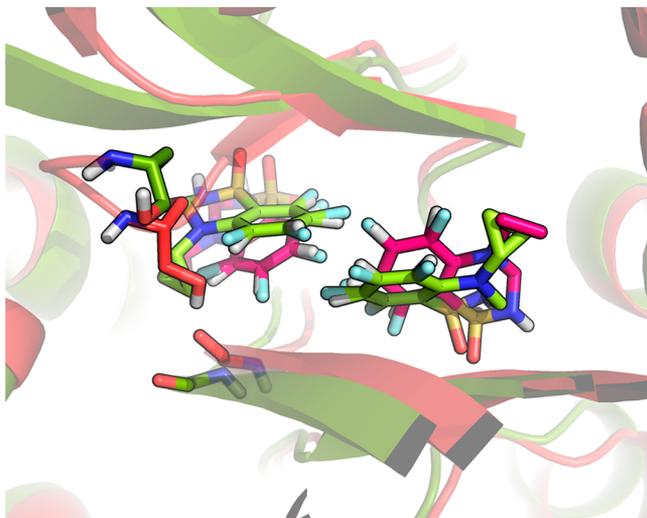
**Table 6. Root-Mean-Square Errors ( $\text{kJ}\cdot\text{mol}^{-1}$ ) Relative to the Experimental Values with Error Estimates Calculated by Leave-One-Out Cross Validation<sup>a</sup>**

	GRA2	TRP	PNMT
A-EDS $1\sigma$	n.a.	$6.1 \pm 0.1$	n.a.
A-EDS $2\sigma$	$6.3 \pm 0.2$	$6.0 \pm 0.1$	$8.8 \pm 0.1$
A-EDS $3\sigma$	$6.0 \pm 0.3$	$8.5 \pm 0.2$	$3.8 \pm 0.1$
EDS	$2.2^b$	n.a.	$5.2^c/4.9^f$
TI	$3.9^c$	$6.6^d$	3.7
OSP	$1.9^c$	$2.8^d$	n.a.

<sup>a</sup>The individual values are reported in Tables S3–S5 in the Supporting Information. n.a. = not available. <sup>b</sup>A-EDS with fixed  $\Delta E_{\text{max}}^*$ ; ref 39. <sup>c</sup>Reference 14. <sup>d</sup>Reference 19. <sup>e</sup>EDS with  $s$ -parameter; ref 32. <sup>f</sup>RE-EDS; ref 36.

For the GRA2 ligands, both simulations with acceleration levels of  $2\sigma$  and  $3\sigma$  performed similarly in terms of agreement with the experimental data, with RMSEs of about 6 kJ/mol. A-EDS simulations previously performed using a fixed set energy barrier<sup>39</sup> using only the five ligands for which experimental data was available showed a lower RMSE of 2.2 kJ/mol. The RMSE of the calculations in this work to the previously reported A-EDS calculations was 9.0 and 6.6 kJ/mol for acceleration levels of  $2\sigma$  and  $3\sigma$ , respectively. The acceleration level in the previous simulations was between the acceleration for  $2\sigma$  and  $3\sigma$  used here. Furthermore, the previous simulations were significantly shorter (11 ns vs 110 ns for parameter search and equilibrium simulations in the protein system). However, the total number of visits to the five first end states lies between 927 (state 2) and 6982 (state 3) and is slightly better comparable to what we observed previously (51 to 2548 unique visits). Obviously, the longer simulation time is largely offset by the larger amount of states that needs to be sampled. The discrepancies between the two sets of simulations can rather be explained due to larger drifts in the overall conformation that are sampled. In system GRA2, two benzothiadiazine dioxide ligands are binding simultaneously to the protein–protein interface of the GRA2 dimer. At the acceleration level of  $2\sigma$ , the distance between the centers of mass of the aromatic rings of the two ligands in the active site moved from initially 0.55 nm to an average value of

0.62 nm. Furthermore, the side chain of Ser108 was seen to change conformation, leading to a different environment of fluorine substituents (Figure 3).



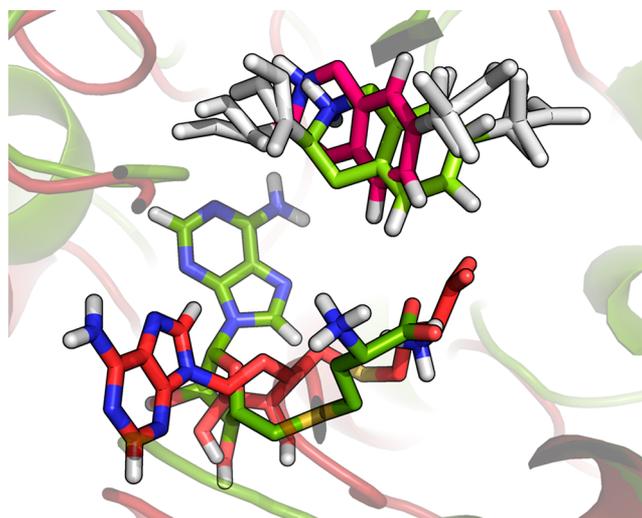
**Figure 3.** Different conformations of the side chain of Ser108 hydrogen-bonding to the carbonyl oxygen of Phe105 in system GRA2 at the end of A-EDS equilibrium simulations of the ligand end states bound to the protein, for an acceleration level of  $2\sigma$  (red, GRA2 ligands in magenta) and  $3\sigma$  (green, GRA2 ligands in light green). Also note the shift in ligand positions at an acceleration level of  $2\sigma$ . The positions of the ligands in the simulation at  $3\sigma$  largely correspond to the ones observed in the initial X-ray structure.

Previously reported calculations using the same force-field but using the thermodynamic integration (TI) or the OSP approach yielded better agreement with experiment, with RMSE values of 3.9 and 1.9 kJ/mol, respectively.<sup>14</sup> These simulations were likely also too short to observe the conformational changes in the active site described above. The calculations using the TI method amounted to a total simulation time of about 50 ns, while the previous simulations with the OSP method was based on 20 ns of simulation time.

For the TRP systems, calculated binding free energies were in better agreement with experimental data for acceleration levels of  $1\sigma$  and  $2\sigma$  than with an acceleration level of  $3\sigma$  (Table S4 in the Supporting Information), possibly caused by insufficient state transitions for the given simulation time with  $3\sigma$ . Moreover, simulations with  $1\sigma$  and  $2\sigma$  performed slightly better than previous calculations employing the same force-field and TI; however, a previously reported method employing OSP and the third-power fitting (TPF) approach for electrostatic interactions performed better than the A-EDS simulations.<sup>19</sup> Except for compound 2, where the TI calculations overestimated the binding affinity, the values obtained for A-EDS with  $2\sigma$  agree closely to the TI data with a root-mean-square difference of 2.9 kJ/mol. The total simulation of the A-EDS simulation for each acceleration level was 130 ns, compared to 245 ns for the previously reported TI method and 52 ns for the previously reported OSP/TPF method.

For the PNMT ligands, the free-energy values computed with an acceleration of  $3\sigma$  fit experimental values much better than those computed with an acceleration of  $2\sigma$  (see also Table S5 in the Supporting Information). However, more state transitions occurred, as expected, for an acceleration level of  $2\sigma$  (Table 5 and Figure S8 in the Supporting Information). Visual inspection

of the trajectories revealed that for the simulation with acceleration parameters determined for  $2\sigma$ , the *S*-adenosyl-*L*-methionine (SAM) cofactor in the PNMT enzyme was in a different conformation, thereby changing the direct environment for the PNMT ligands bound to the protein (see Figure 4),



**Figure 4.** Different conformations of the *S*-adenosyl-*L*-methionine (SAM) cofactor in system PNMT at the end of A-EDS equilibrium simulations of the ligand end states bound to the protein, for an acceleration level of  $2\sigma$  (red, PNMT ligand in magenta) and  $3\sigma$  (green, PNMT ligand in light green).

explaining differences in calculated affinities of the end state for this simulation. A-EDS simulations with an acceleration level of  $3\sigma$  performed similar in terms of comparison to experiment as previously reported values obtained with the same force-field and TI<sup>48</sup> and better than other EDS-based methods.<sup>32,36</sup> It is, moreover, noted that in the A-EDS simulations performed in this work ligand 2, to which the perturbation includes a net-charge change of the molecule, was included in the reference state, while it was omitted in previously reported EDS calculations. While the number of visits is very limited for this state (Table 5) and the binding affinities are quite different from the TI values (Table S5), the A-EDS approach is able to include this state and correctly predicts it to be a nonbinder. For each A-EDS acceleration level a total simulation time of 320 ns was used, while for the simulations employing TI sampling was done for  $\sim 1.5 \mu\text{s}$ , and for the most recently reported approach employing EDS (RE-EDS) sampling was done for  $\sim 670$  ns.

For all systems, either simulations with acceleration parameters of  $2\sigma$  or  $3\sigma$  lead to the best agreement of the calculated binding free energies with experimental values. For system TRP, very little state transitions occurred with an acceleration level of  $3\sigma$ , leading to undersampling and less good agreement with experimental values. For systems GRA2 and PNMT ( $2\sigma$ ) conformations were sampled that differ from the experimentally determined structures, possibly explaining less good agreement with experimental values for these simulations. This highlights a so-far unexplored challenge for longer simulations sampling many different states. Due to the sampling of unphysical reference states and the inclusion of compounds with low affinity, the active site conformation may diverge to irrelevant conformations. Possibly adding further slight restraints on the active site conformations could already be sufficient to avoid this behavior.<sup>58</sup> Simulations with an

acceleration level of  $3\sigma$  might perform generally better if enough simulation time is generated, as the energy landscape is less flattened and important energy minima are more pronounced.

## CONCLUSION

In this work, we presented a novel automated parameter search scheme for A-EDS. The search scheme only requires a single chosen input-parameter, the anticipated acceleration level for flattening of the energy barriers between the EDS end states ( $\sigma$ -level). All other required parameters for equilibrium A-EDS simulations (A-EDS acceleration parameters  $E_{\max}$  and  $E_{\min}$ , and the free-energy offset parameters  $\Delta F_{i \neq 1}^R$ ) are determined using this search scheme in a nonequilibrium parameter-search simulation prior to the A-EDS equilibrium simulation. We demonstrated the applicability of this scheme with three well-established protein–ligand model systems and showed that using this method, free-energy differences between multiple ligands can be computed from a single simulation in an automated way. The full automation of the parameter-search simulations represents an important step forward compared to previously described multistate EDS simulations. All ligand end states were sampled, including ligands to which the perturbation involves net-charge changes. We tested three different acceleration levels for the parameter-search simulations and conclude that in general it is better to choose a more conservative acceleration level, as important energy minima might be lost otherwise. However, a more conservative choice of acceleration levels requires longer simulation times, and in practice, both the importance of energy minima and available simulation time have to be considered when choosing the acceleration level. A remaining challenge is the occurrence of conformational changes in the protein–ligand systems due to unfavorable ligand end states and the unphysical nature of the EDS reference state. EDS ligands that are strongly accelerated and incorporate unfavorable end-states need to be restrained to the protein to prevent them from leaving the binding site and to sample irrelevant binding modes.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.0c00456>.

Definition of the protein–ligand distance restraints for systems TRP and PNMT; calculated and experimentally determined binding affinities of all the compounds and systems; convergence of the A-EDS parameters during the parameter-search simulations for acceleration levels of  $1\sigma$  and  $3\sigma$ ; convergence of the calculated relative free-energy differences  $\Delta F_{i \neq 1,1}$  between the ligand end states relative to their final values from the equilibrium A-EDS simulations for all systems; ligand end state time series in the equilibrium A-EDS simulations for all systems; number of transitions between end states during the first 10 ns of the A-EDS parameter-search simulations with the ligand reference state bound to the protein; A-EDS free-energy offset parameter search results for the three different acceleration  $\sigma$ -levels for all systems (PDF)

GROMOS building block definitions for the different PNMT ligands used in this work (ZIP)

## AUTHOR INFORMATION

### Corresponding Author

Chris Oostenbrink – Institute for Molecular Modeling and Simulation, Department for Material Sciences and Process Engineering, University of Natural Resources and Life Sciences (BOKU), Vienna, 1190 Vienna, Austria; [orcid.org/0000-0002-4232-2556](https://orcid.org/0000-0002-4232-2556); Email: [chris.oostenbrink@boku.ac.at](mailto:chris.oostenbrink@boku.ac.at)

### Authors

Jan Walther Perthold – Institute for Molecular Modeling and Simulation, Department for Material Sciences and Process Engineering, University of Natural Resources and Life Sciences (BOKU), Vienna, 1190 Vienna, Austria; [orcid.org/0000-0002-8575-0278](https://orcid.org/0000-0002-8575-0278)

Drazen Petrov – Institute for Molecular Modeling and Simulation, Department for Material Sciences and Process Engineering, University of Natural Resources and Life Sciences (BOKU), Vienna, 1190 Vienna, Austria

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.0c00456>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank Sereina Riniker and Dominik Sidler for making the input data of the PNMT system available. J.W.P. is a recipient of a DOC Fellowship of the Austrian Academy of Sciences (ÖAW) at the Institute for Molecular Modeling and Simulation at the University of Natural Resources and Life Sciences, Vienna (Grant No. 24987). The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC).

## REFERENCES

- (1) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061.
- (2) Tembre, B. L.; Mc Cammon, J. A. Ligand-receptor interactions. *Comput. Chem.* **1984**, *8*, 281–283.
- (3) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632–2647.
- (4) Jorgensen, W. L.; Thomas, L. L. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869–876.
- (5) Christ, C. D.; Fox, T. Accuracy assessment and automation of free energy calculations for drug design. *J. Chem. Inf. Model.* **2014**, *54*, 108–120.
- (6) Mikulskis, P.; Genheden, S.; Ryde, U. A large-scale test of free-energy simulation estimates of protein-ligand binding affinities. *J. Chem. Inf. Model.* **2014**, *54*, 2794–2806.
- (7) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.
- (8) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

- (9) Mark, A. E.; Xu, Y.; Liu, H.; van Gunsteren, W. F. Rapid non-empirical approaches for estimating relative binding free energies. *Acta Biochim Pol* **1995**, *42*, 525–535.
- (10) Liu, H.; Mark, A. E.; van Gunsteren, W. F. Estimating the Relative Free Energy of Different Molecular States with Respect to a Single Reference State. *J. Phys. Chem.* **1996**, *100*, 9485–9494.
- (11) Oostenbrink, B. C.; Pitera, J. W.; van Lipzig, M. M. H.; Meerman, J. H. N.; van Gunsteren, W. F. Simulations of the Estrogen Receptor Ligand-Binding Domain: Affinity of Natural Ligands and Xenostrogens. *J. Med. Chem.* **2000**, *43*, 4594–4605.
- (12) Oostenbrink, C. Free energy calculations from one-step perturbations. *Methods Mol. Biol.* **2012**, *819*, 487–499.
- (13) Raman, E. P.; Vanommelaeghe, K.; Mackerell, A. D., Jr. Site-Specific Fragment Identification Guided by Single-Step Free Energy Perturbation Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3513–3525.
- (14) Norholm, A. B.; Francotte, P.; Goffin, E.; Botez, I.; Danober, L.; Lestage, P.; Pirotte, B.; Kastrop, J. S.; Olsen, L.; Oostenbrink, C. Thermodynamic characterization of new positive allosteric modulators binding to the glutamate receptor A2 ligand-binding domain: combining experimental and computational methods unravels differences in driving forces. *J. Chem. Inf. Model.* **2014**, *54*, 3404–3416.
- (15) Graf, M. M.; Zhixiong, L.; Bren, U.; Haltrich, D.; van Gunsteren, W. F.; Oostenbrink, C. Pyranose dehydrogenase ligand promiscuity: a generalized approach to simulate monosaccharide solvation, binding, and product formation. *PLoS Comput. Biol.* **2014**, *10*, No. e1003995.
- (16) Pitera, J. W.; van Gunsteren, W. F. One-Step Perturbation Methods for Solvation Free Energies of Polar Solutes. *J. Phys. Chem. B* **2001**, *105*, 11264–11274.
- (17) Oostenbrink, C.; van Gunsteren, W. F. Single-step perturbations to calculate free energy differences from unphysical reference states: limits on size, flexibility, and character. *J. Comput. Chem.* **2003**, *24*, 1730–1739.
- (18) Oostenbrink, C. Efficient free energy calculations on small molecule host-guest systems - a combined linear interaction energy/one-step perturbation approach. *J. Comput. Chem.* **2009**, *30*, 212–221.
- (19) de Ruiter, A.; Oostenbrink, C. Efficient and Accurate Free Energy Calculations on Trypsin Inhibitors. *J. Chem. Theory Comput.* **2012**, *8*, 3686–3695.
- (20) Jandova, Z.; Fast, D.; Setz, M.; Pechlaner, M.; Oostenbrink, C. Saturation Mutagenesis by Efficient Free-Energy Calculation. *J. Chem. Theory Comput.* **2018**, *14*, 894–904.
- (21) Christ, C. D.; van Gunsteren, W. F. Enveloping distribution sampling: a method to calculate free energy differences from a single simulation. *J. Chem. Phys.* **2007**, *126*, 184110.
- (22) Han, K.-K. A new Monte Carlo method for estimating free energy and chemical potential. *Phys. Lett. A* **1992**, *165*, 28–32.
- (23) Wu, D.; Kofke, D. A. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *J. Chem. Phys.* **2005**, *123*, 084109.
- (24) Chen, Y. G.; Hummer, G. Slow conformational dynamics and unfolding of the calmodulin C-terminal domain. *J. Am. Chem. Soc.* **2007**, *129*, 2414–2415.
- (25) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (26) Han, K. K. Multiensemble sampling: An alternative efficient Monte Carlo technique. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1996**, *54*, 6906–6910.
- (27) Lu, N.; Wu, D.; Woolf, T. B.; Kofke, D. A. Using overlap and funnel sampling to obtain accurate free energies from nonequilibrium work measurements. *Phys. Rev. E* **2004**, *69*, 057702.
- (28) Christ, C. D.; van Gunsteren, W. F. Multiple free energies from a single simulation: extending enveloping distribution sampling to nonoverlapping phase-space distributions. *J. Chem. Phys.* **2008**, *128*, 174112.
- (29) Reinhardt, M.; Grubmüller, H. Determining Free Energy Differences Through Non-Linear Morphing. *J. Chem. Theory Comput.* **2020**, *16*, 3504.
- (30) Christ, C. D.; van Gunsteren, W. F. Simple, Efficient, and Reliable Computation of Multiple Free Energy Differences from a Single Simulation: A Reference Hamiltonian Parameter Update Scheme for Enveloping Distribution Sampling (EDS). *J. Chem. Theory Comput.* **2009**, *5*, 276–286.
- (31) Christ, C. D.; Van Gunsteren, W. F. Comparison of three enveloping distribution sampling Hamiltonians for the estimation of multiple free energy differences from a single simulation. *J. Comput. Chem.* **2009**, *30*, 1664–1679.
- (32) Riniker, S.; Christ, C. D.; Hansen, N.; Mark, A. E.; Nair, P. C.; van Gunsteren, W. F. Comparison of enveloping distribution sampling and thermodynamic integration to calculate binding free energies of phenylethanolamine N-methyltransferase inhibitors. *J. Chem. Phys.* **2011**, *135*, 024105.
- (33) Hansen, N.; Dolenc, J.; Knecht, M.; Riniker, S.; van Gunsteren, W. F. Assessment of enveloping distribution sampling to calculate relative free enthalpies of binding for eight netropsin-DNA duplex complexes in aqueous solution. *J. Comput. Chem.* **2012**, *33*, 640–651.
- (34) Maurer, M.; Hansen, N.; Oostenbrink, C. Comparison of free-energy methods using a tripeptide-water model system. *J. Comput. Chem.* **2018**, *39*, 2226–2242.
- (35) Sidler, D.; Schwaninger, A.; Riniker, S. Replica exchange enveloping distribution sampling (RE-EDS): A robust method to estimate multiple free-energy differences from a single simulation. *J. Chem. Phys.* **2016**, *145*, 154114.
- (36) Sidler, D.; Cristofol-Clough, M.; Riniker, S. Efficient Round-Trip Time Optimization for Replica-Exchange Enveloping Distribution Sampling (RE-EDS). *J. Chem. Theory Comput.* **2017**, *13*, 3020–3030.
- (37) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–29.
- (38) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584–3595.
- (39) Perthold, J. W.; Oostenbrink, C. Accelerated Enveloping Distribution Sampling: Enabling Sampling of Multiple End States while Preserving Local Energy Minima. *J. Phys. Chem. B* **2018**, *122*, 5030–5037.
- (40) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (41) Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **1996**, *17*, 1587–1597.
- (42) Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. Monte Carlo Simulations for Proteins: Binding Affinities for Trypsin-Benzamide Complexes via Free-Energy Perturbations. *J. Phys. Chem. B* **1997**, *101*, 9663–9669.
- (43) Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient  $\beta$  in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins: Struct., Funct., Genet.* **1999**, *34*, 395–402.
- (44) Brandsdal, B. O.; Smalas, A. O.; Aqvist, J. Electrostatic effects play a central role in cold adaptation of trypsin. *FEBS Lett.* **2001**, *499*, 171–175.
- (45) Brandsdal, B. O.; Aqvist, J.; Smalas, A. O. Computational analysis of binding of P1 variants to trypsin. *Protein Sci.* **2001**, *10*, 1584–1595.
- (46) Leiros, H. K.; Brandsdal, B. O.; Andersen, O. A.; Os, V.; Leiros, L.; Helland, R.; Otlewski, J.; Willassen, N. P.; Smalas, A. O. Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements. *Protein Sci.* **2004**, *13*, 1056–1070.
- (47) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6290–6295.
- (48) Nair, P. C.; Malde, A. K.; Mark, A. E. Using Theory to Reconcile Experiment: The Structural and Thermodynamic Basis of Ligand Recognition by Phenylethanolamine N-Methyltransferase (PNMT). *J. Chem. Theory Comput.* **2011**, *7*, 1458–1468.

(49) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. Time-dependent distance restraints in molecular dynamics simulations. *Chem. Phys. Lett.* **1989**, *157*, 289–294.

(50) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput. Phys. Commun.* **2012**, *183*, 890–903.

(51) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(52) Hockney, R. W. *The Potential Calculation and some Applications*; Academic Press Inc.: New York, 1970; p 77 p.

(53) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(54) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(55) Tironi, I.; Sperb, R.; Smith, P.; van Gunsteren, W. F. A Generalized Reaction Field Method for Molecular Dynamics Simulation. *J. Chem. Phys.* **1995**, *102*, 5451.

(56) Heinz, T. N.; van Gunsteren, W. F.; Hünenberger, P. H. Comparison of Four Methods to Compute the Dielectric Permittivity of Liquids from Molecular Dynamics Simulations. *J. Chem. Phys.* **2001**, *115*, 1125–1136.

(57) Oostenbrink, C.; van Gunsteren, W. F. Free energies of binding of polychlorinated biphenyls to the estrogen receptor from a single simulation. *Proteins: Struct., Funct., Genet.* **2004**, *54*, 237–246.

(58) Menzer, W. M.; Xie, B.; Minh, D. D. L. On Restraints in End-Point Protein-Ligand Binding Free Energy Calculations. *J. Comput. Chem.* **2020**, *41*, 573–586.