JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

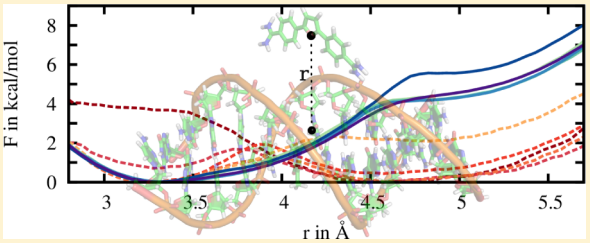
Adaptive Biasing Combined with Hamiltonian Replica Exchange to Improve Umbrella Sampling Free Energy Simulations

Fabian Zeller and Martin Zacharias*

Physik-Department T38, Technische Universität München, James Franck Str. 1, 85748 Garching, Germany

Supporting Information

ABSTRACT: The accurate calculation of potentials of mean force for ligand−receptor binding is one of the most important applications of molecular simulation techniques. Typically, the separation distance between ligand and receptor is chosen as a reaction coordinate along which a PMF can be calculated with the aid of umbrella sampling (US) techniques. In addition, restraints can be applied on the relative position and orientation of the partner molecules to reduce accessible phase space. An approach combining such phase space reduction with flattening of the free energy landscape and configurational exchanges has been developed, which significantly improves the convergence of PMF calculations in comparison with standard umbrella sampling. The free energy surface along the reaction coordinate is smoothened by iteratively adapting biasing potentials corresponding to previously calculated PMFs. Configurations are allowed to exchange between the umbrella simulation windows via the Hamiltonian replica exchange method. The application to a DNA molecule in complex with a minor groove binding ligand indicates significantly improved convergence and complete reversibility of the sampling along the pathway. The calculated binding free energy is in excellent agreement with experimental results. In contrast, the application of standard US resulted in large differences between PMFs calculated for association and dissociation pathways. The approach could be a useful alternative to standard US for computational studies on biomolecular recognition processes.



INTRODUCTION

The determination of binding free energies of biomolecular complexes constitutes an important task in understanding biological processes as it plays a key role in revealing binding mechanisms and ligand functionality.[1] Successive advances in the determination of binding free energies by means of computer simulations are contributing with growing impact to this area.[1−8] Presently established approaches to calculate the absolute binding free energy of biomolecular complexes are the double-decoupling scheme[9,10] and potential of mean force (PMF) based pulling methods.[2,11] In the PMF approach, the separation of the complex along a predefined reaction path is induced during a molecular dynamics simulation, and the associated change in free energy is calculated. Simulations are typically only performed for the dissociation process.

The PMF approach is particularly promising as it imitates the molecular binding process more closely[6] compared to an alchemical transformation and avoids the necessity of sampling of phase space, which is not relevant for the binding process. This is especially advantageous in the case of large solvation free energies[2,6] or binding sites that are small in comparison to the overall system size. However, difficulties may arise when a simple association pathway for the complex cannot be constructed.[6]

While the theoretical background for determining standard absolute binding free energies is well established[2,6,9] and the currently available force fields provide useful atomic models for a wide range of cases,[12,13] sampling issues are still a bottleneck for practical applications. The major sources of inaccuracy are related to an insufficient phase space coverage.[14,15] Convergence and reversibility of the simulations are often uncertain.[6] Although for simple cases this problem can be handled by ordinary umbrella sampling (US),[16,17] the computational costs of brute force approaches generally exceed the currently available resources, and thus more sophisticated sampling techniques might be required.

The PMF approach basically reduces the binding process to determining the changes in free energy along a one-dimensional reaction coordinate. This allows for an especially powerful application and combination of recent developments concerning free energy sampling efficiency. Combinations of umbrella sampling and H-REMD based exchanges along the reaction coordinate have been evaluated.[15] Also, adaptive biasing techniques like metadynamics,[18] the adaptive biasing force method,[15,19] and adaptively biased MD[20] have been used to calculate free energy changes. However, combination of all three approaches has not been explored.

In the present study, we designed a robust method making use of phase space reduction with appropriate restraints as well as adaptive biasing and Hamiltonian replica exchanges along the reaction coordinate. The method was first tested on the separation of a sodium/chloride ion pair in solution and then applied to a DNA−furamidine complex (Figure 1). For the

Received: August 2, 2013
Published: December 23, 2013

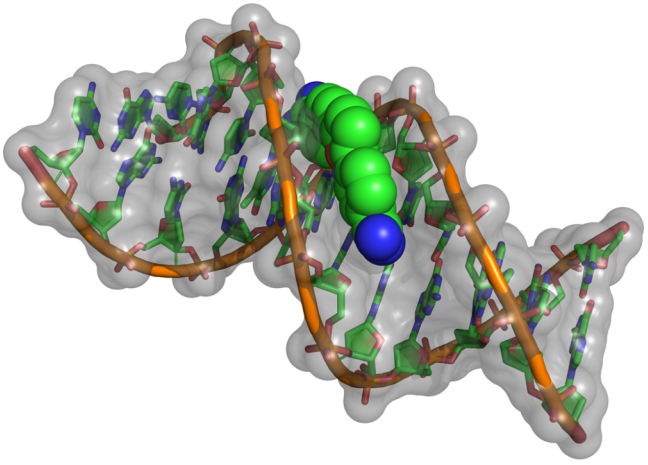


Figure 1. The crystal structure of furamidine in complex with a 12 base pair DNA oligo nucleotide (PDB: 227D) served as the starting structure for the simulations. The minor groove of d-$(CGCGAATTCGCT)_2$ (stick model) provides a tight binding pocket for furamidine (van der Waals spheres).

latter system, standard restrained umbrella sampling resulted in slow convergence of the sampling and in significant differences between the PMFs calculated for dissociation and association pathways, mainly due to trapped water molecules. In contrast, the iterative biasing with PMFs obtained along the reaction coordinate and exchanging of configurations using H-REMD yielded rapid convergence and excellent agreement of the PMFs for dissociation and association of the minor groove ligand. The predicted standard binding free energy was also in very good agreement with experimental results and allowed to relate changes in hydration of the DNA to the binding process.

METHODS

Theoretical Background and Method Design. In the present work, the sampling efficiency of the umbrella simulations for calculating binding free energies is intended to be enhanced by three orthogonal approaches: reduction of the effective phase space accessible to the system, uniform sampling of the free energy landscape, and overcoming of free energy barriers by configurational exchanges.

The strategy to calculate the standard binding free energy via a PMF along a separation pathway with reduced accessible phase space is based on an approach by Woo and Roux.[2] While relative orientational and positional restraints are applied on the ligand, a radial PMF is generated spanning from a bound state to a bulk state distance. The convergence of the PMF sampling is significantly improved due to the constriction of the relative external degrees of freedom of the system. Woo and Roux provided a theoretical formulation that corrects for the biases caused by the restraints and yields the absolute binding free energy for standard conditions. The overall binding free energy is split into several contributions:[2]

$$G_{\text{bind}} = -G_{\text{o}}^{\text{site}} - G_{\text{a}}^{\text{site}} - k_{\text{B}}T \ln(I^*S^*C^0) + G_{\text{o}}^{\text{bulk}} \quad (1)$$

with

$$I^* = \int_{\text{site}} \mathrm{d}r \; \mathrm{e}^{-\beta[W(r)-W(r^*)]} \quad (2)$$

$$S^* = (r^*)^2 \int_0^{\pi} \sin(\Theta_1)\, \mathrm{d}\Theta_1 \int_0^{2\pi} \mathrm{d}\Phi_1 \, \mathrm{e}^{-\beta u_a(\Theta_1,\Phi_1)} \quad (3)$$

where $\beta = 1/k_{\text{B}}T$. I^* contains the main free energy contribution of the radial PMF at a bulk radius $W(r^*)$ and an integral over the region of the PMF that constitutes the binding site. This bound state region in the PMF is not unambiguously defined. For the purpose of comparability it should be chosen in accordance with what in experiments was measured as the bound state. However, the binding free energy turns out not to be very sensitive to the exact choice. This is mainly due to the steep rise of the PMF around its minimum and the resulting suppression of the corresponding integral contributions by the Boltzmann factor.[15] Variation of PMF values chosen as a bound state threshold between 2 $k_{\text{B}}T$ and 15 $k_{\text{B}}T$ resulted in a binding free energy difference less than 0.1 kcal/mol. For the calculations, the system was assumed to be in the bound state for PMF values smaller than roughly 4 $k_{\text{B}}T$. $G_{\text{o}}^{\text{site}}$ and $G_{\text{o}}^{\text{site}}$ are the contributions related to the loss of free energy due to the introduction of the restraints in the bound state. S^*C^0 corresponds to the release of the ligand from a sphere shell volume element at a bulk separation r^* to a volume that represents the standard concentration. $G_{\text{o}}^{\text{bulk}}$ is related to the release of the orientational restraints from the ligand in the bulk. S^* and $G_{\text{o}}^{\text{bulk}}$ can be calculated as a numerical integral since the energy in the bulk is independent from the ligand's position and orientation. It is worth it to mention that S^* and $G_{\text{o}}^{\text{bulk}}$ do not depend on the reference restraining angles but only on the restraining potential strength as the entropy gain in the bulk is isotropic.

Uniform sampling and flattening of the free energy landscape in the individual umbrella simulations is achieved by the introduction of adaptive biasing potentials (ABPs).[20] Adaptive biasing potentials constitute one of the most recent developments concerning memory based sampling methods.[18,21] While the free energy landscape is explored, intermediate estimates of the free energy landscape are subtracted from the system in order to bias the simulation into energetically unfavored regions. At equilibrium and in case of convergence, the ABP corresponds to the system's real free energy landscape. The effective free energy landscape experienced by the simulation is then constant, and therefore the probability distributions are uniform or only influenced by umbrella potentials, which are used in order to focus the sampling.

In this study, a simple linear interpolation and iteration implementation is used for every individual umbrella window. Initial PMFs are calculated without adaptive biasing potentials performing short standard US simulations. The PMF estimates are used as biasing potentials and are subtracted from the system's free energy along the reaction coordinate in a subsequent set of US simulations. The negative mean force along the reaction coordinate corresponding to the biasing PMFs is obtained by direct numerical differentiation, which is equivalent to a linear interpolation of the PMFs between the evaluation positions. The subsequent US simulations yield new PMF estimates that are expected to be better sampled as the free energy landscape is smoothened by the preceding PMF estimates. This adaptive process is repeated iteratively until the PMF calculations eventually converge. After each iteration step, the probability distributions for the umbrella windows are unbiased from both the harmonic and adaptive potentials and converted into free energies using the weighted histogram analysis method.[22] The final PMF is calculated from the sampling data of all umbrella windows together, again via the WHAM equations.

dx.doi.org/10.1021/ct400689h | J. Chem. Theory Comput. 2014, 10, 703−710

The necessary code changes (code changes available via the Internet at http://www.t38.ph.tum.de/index.php?id=88) were incorporated into the pmemd module of Amber12 and the utilized WHAM program.[23]

Some important aspects of this implementation should be pointed out. The ABPs used for the first iterations do not necessarily need to be converged as any (possibly arbitrary) bias is completely removed via the WHAM equations. The intention of introducing the ABPs is to speed up the possibly complicated equilibration process and to help explore phase space by smoothening the free energy surface along the reaction coordinate. The efficiency of these processes in detail depends on the utilized iteration protocol.

Convergence of the final PMF calculations is only assured when individual ABPs are used for each umbrella simulation window. If only one ABP is constructed from the sampling of all umbrella windows across the whole reaction coordinate range, artificial free energy barriers can be generated. It can then occur that during equilibration, different configurational space is sampled in neighboring windows. In the overlap regions, the resulting biasing potential would not correspond to the respective states of the individual simulations. In this case, misleading biases may evolve that prevent equilibration.

Minor discontinuities in the force due to the linear interpolation of the ABP are found to be negligible. However, the PMF estimates should be sampled to such an extent that the ABPs are sufficiently smooth. The resolution of the ABPs is arbitrarily controllable by the bin width used for the PMF calculations. Note, that the iteration approach exclusively requires equilibrium simulations. It can be utilized for any US-simulation application.

Configurational exchanges between neighboring umbrella windows were allowed using the H-REMD implementation in the Amber12 package.[24] Here, the Hamiltonians of the windows differ in the position of the harmonic umbrella potentials and in the individual biasing potentials. The Boltzmann distribution of the canonical configurations is preserved by a Metropolis-like exchange acceptance criterion.[25] The exchange mechanism additionally benefits from the biasing potentials, which lead to an equal distribution of sampling overlap between neighboring windows.

See corresponding sections for details on the simulation procedure and results.

Simulation Setup. All simulations were performed with the AMBER12 software package.[24] The GAFF forcefield and antechamber as part of Amber12 were used to prepare the furamidine ligand. The parmBsc parameters[26] were employed for the DNA force field description. The crystallographic structure of the d(CGCGAATTCGCT)$_2$–furamidine complex was taken from ref 27 (PDB: 227D). The complex was solvated in explicit TIP3P[28] water molecules in a box of dimensions 55 × 55 × 70 Å^3. Sodium ions described by the force field parameters from Joung and Cheatham[29] were added to neutralize the overall charge. The total system size was 16011 atoms. Long range interactions were accounted for by the particle mesh Ewald method, using a grid spacing of ∼1 Å and the default AMBER12 PME parameters. The real space cutoff radius was 9 Å. The internal energy of the resulting system was minimized for 10 000 steps. The equilibration was carried out in the isobaric–isothermal ensemble at a constant pressure of 1 bar. The system was heated up to 100, 200, and 298 K in three 20 ps simulation steps. During this process, the complex was restrained to the crystallographic structure using harmonic positional potentials with a force constant of 50 kcal/Å^2 per atom. The positional restraining force was then consecutively reduced to 24, 12, 6, 2, and finally 0.2 kcal/Å^2 in five simulation steps of 20 ps. Further equilibration of 100 ps was carried out with the positional restraints applied only to the DNA atoms, using a force constant of 0.002 kcal/Å^2. These slight restraints were maintained during the sampling runs in order to prevent drifting issues. The sampling trajectories were generated in the canonical ensemble at 298 K. Simulations on the sodium/chloride ion test system were performed using the same water model, ion force field, and equilibration protocol.

Reaction Coordinate and Restraining of External Degrees of Freedom. In order to construct the binding pathway and to restrain the ligand's position and orientation with respect to the DNA, three centers, each in the ligand and the receptor, were defined (Figure 2), following a procedure

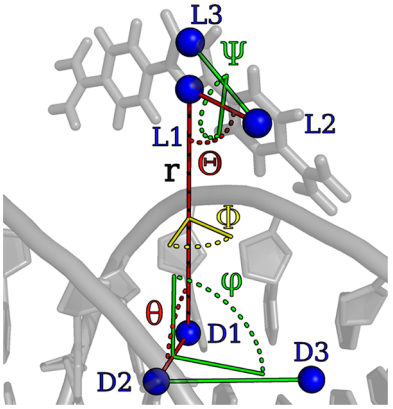


Figure 2. The relative orientation and position of the ligand are restrained during the umbrella simulation pulling process. The three Euler angles Φ, Θ, and Ψ as well as two axial angles ϕ and θ are defined by centers of mass in DNA (D1, D2, D3) and ligand (L1, L2, L3, see Methods section and Figure 1S).

introduced by Woo and Roux.[2] The points D1, D2, D3, L1, L2, and L3 were chosen as the center of mass of the atom groups (5–9,17–20,22@C3′), (5–9@C3′), (17–22@C3′), (25@CA1,CA2,C11,C12,C41,C42), (25@C12,C42), and (25@CB1,CB2), respectively (the atom specifications (corresponding PDB file available via the Internet at http://www.t38.ph.tum.de/index.php?id=88) are outlined in Figure 1S, Supporting Information). The distance r (D1–L1) defines the relative position of the ligand on a pulling axis determined by the angle θ (D2–D1–L1) and the dihedral angle ϕ (D3–D2–D1–L1). The dihedral angle Φ (D2–D1–L1–L2), the angle Θ (D1–L1–L2), and the dihedral angle Ψ (D1–L1–L2–L3) were chosen as the three Euler angles that determine the relative orientation of the ligand. The distance r was chosen as a reaction coordinate and later on varied via the umbrella sampling windows. The remaining previously defined degrees of freedom were restrained by the harmonic potentials $u_o = k_o[(\Phi - \Phi_0)^2 + (\Theta - \Theta_0)^2 + (\Psi - \Psi_0)^2]$ and $u_a = k_a[(\phi - \phi_0)^2 + (\theta - \theta_0)^2]$ with force constants $k_0 = k_a = 30$ kcal/mol/rad^2. For the reference angles θ_0, ϕ_0, Θ_0, Φ_0, and Ψ_0, the most probable values in the bound state of the complex were chosen. They were obtained by analyzing the trajectory of a 2 ns free simulation of the bound complex. The free energy cost of introducing the relative orientational and positional restraints on the ligand was calculated from this trajectory via free energy perturbation (see Table 1).

Table 1. Calculated Free Energy Contributions (r^* = 14.9 Å) and Experimental Binding Free Energy Values

$\Delta G^{site}_{o,a}$	+1.15 ± 0.07 kcal/mol
$-k_BT\ \ln(S^*I^*C^0)$	−11.82 ± 0.67 kcal/mol
ΔG^{bulk}_{o}	+5.10 ± 0.00 kcal/mol
$\Delta G_{el.longrange(0.2M\ NaCl)}$	−1.27 ± 0.16 kcal/mol
$\Delta G_{bind,calc}$	−9.14 ± 0.90 kcal/mol
$\Delta G_{bind,exp(0.2M\ NaCl)}$ [30]	−9.0 kcal/mol
$\Delta G_{bind,exp(0.2M\ NaCl)}$ [27]	−9.3 kcal/mol

Simulation Procedure. Umbrella simulations were set up to sample canonical configurations along the reaction coordinate r. A set of 40 umbrella windows (indexed by i) biased by the harmonic potentials $u^i_{UMB}(r) = k_{UMB}(r - r_{0,i})^2$ with an intermediate spacing of 0.3 Å was used to cover a range of distances from 3.2 Å to 14.9 Å. The force constant k_{UMB} was 8 kcal/Å^2 for all windows. The individual umbrella simulations were additionally biased by adaptive potentials $u^{i,j}_{ABP}$ in order to gradually flatten the free energy landscape. The adaptive potentials were constructed by linear interpolation of iteratively (indexed by j) calculated discrete PMFs (see Theoretical Background). The pathway along the distance r was divided into three regions (I, II, and III) that were treated with different sampling efforts, corresponding to the respective complexity of the free energy landscape. The eight windows covering the closest distances (I, 3.2−5.3 Å) were allowed to exchange configurations every 10 ps via the pmemd H-REMD routine (see Theoretical Background). The two remaining regions (windows 9−20 (II, 5.6−8.9 Å) and 21−40 (III, 9.2−14.9 Å)) were treated with adaptively biased simulations only. Starting configurations for the sets of umbrella configurations were generated using the bound complex system resulting from equilibration and consecutively driving it to the target distances by applying the harmonic potentials u^i_{UMB} for 0.5 ns each. The reverse pathway (association) was simulated as a convergence verification for the 20 closest windows (regions I and II) where free energy sampling was found to be most complicated. Starting configurations for the reverse convergence test run were generated starting from a bulk configuration previously generated by the 8.9 Å distance setup window. For the close range region, 14 iterations were performed for the dissociation and 16 iterations for the association pathway, with simulation times ranging from 0.5 to 5 ns. The detailed iteration protocol for all regions is shown in Table 1S (Supporting Information). The probability distributions obtained by the setup runs (without adaptive biasing) were used for calculating the initial adaptive biasing potentials. As long as the resulting PMFs were still changing significantly, the adaptive biasing potentials were completely replaced by the recent PMF estimates (iteration runs of 2 ns or shorter). Once the largest PMF change per iteration was below 0.5 kcal/mol, the trajectories were stored for the final PMF calculations and the biasing potentials were consecutively updated by the sampling data from previous runs (iteration runs of 5 ns or longer). After each adaptation of the biasing potential, a 50 ps equilibration simulation was performed before starting data gathering. The PMF resolution was 0.1 Å for the initial estimates of the biasing potentials generated by the short setup runs and 0.05 Å for all following iterations.

In order to estimate the errors of the PMF calculations, we divided the sampling data of each iteration run into six successive subsets and calculated individual PMFs for them. At each evaluation point, we used the subset PMF value with the largest deviation from the total PMF as lower/upper error estimates. This allows not only estimation of the statistical significance of the free energy differences at every data bin but also of the error propagation across the bins. The obtained error estimates do not contain any information about to what extent overall phase space is sampled.

Long Range Electrostatics and Salt Concentration Effects. For the highly charged DNA−ligand system, long-range electrostatic interactions beyond the largest US distance can be relevant for binding. Note that this contribution is only relevant for comparison with the experiment and not for the performance of the presented US approach. At large distances between the ligand and DNA, a Debye−Hückel approach is reasonable since the ligand and DNA are then only influenced by screened DNA and ligand charges. The radial PMF was calculated up to a bulk separation of r^* = 14.9 Å. For distances r > 12.0 Å, the following model was used to approximate the long-range charge−charge interactions:

$$F(r) = -A \sum_{i}^{\text{charges DNA}} \sum_{j}^{\text{charges ligand}} \frac{e^{-\kappa[d_{ij}(r)-d_0]}}{d_{ij}(r)(1 + d_0\kappa)} - k_BT\ \ln(r^2) + C \quad (4)$$

Each DNA−ligand charge pair interaction is represented by a Debye−Hückel term with a common factor A, a charge radius d_0, and the Debye-screening length $1/\kappa$. The distances $d_{ij}(r)$ denote the separations of the charge pairs. The contribution $-k_BT\ \ln(r^2)$ corresponds to the entropy that is associated with the volume element the ligand can access at a DNA−ligand separation r. C is a constant free energy offset parameter. The salt concentration was taken into account by choosing the Debye-screening length $1/\kappa$ in accordance with the experimental conditions ($1/\kappa = \sqrt{(\varepsilon_0\varepsilon_r k_BT/2N_Ae^2c)}$, c = 0.2 M). The free parameters A, d_0, and C were determined by fitting the model into the radial PMF between r = 12.0 Å and r = 14.9 Å. The distances $d_{ij}(r)$ were obtained from the simulation trajectories as average distances corresponding to the DNA−ligand separation r. The negative DNA charges were assumed to be located at the phosphorus atoms; the two positive ligand charges were assumed to be located at the center of mass of the nitrogen atoms of the amidine groups. The fit yielded A = 5.75 ± 0.74 Å kcal/mol, C = 19.48 ± 0.53 kcal/mol, and $d_0 = 2.00^{+2.38}_{-0.64}$ Å. Note that d_0 is on the order of magnitude of the radii of the phosphate and charged ligand groups. With the parameters resulting from the fit, the long-range free energy contribution was calculated by evaluating the model at r^* = 14.9 Å. This approximation neglects the salt concentration influences on the binding free energy in the close range region, as in the simulations only counterions were present in order to neutralize the system's overall charge.

RESULTS AND DISCUSSION

The combination of phase space restraints, H-REMD-US, and iterative biasing along the reaction coordinate was first applied to a quickly converging example (sodium/chloride ion pair in explicit solvent) and verified by comparison with standard umbrella sampling (illustrated in Figure 3). Sampling time for ABP-H-REMD-US was 1 and 10 ns per window for the first and second iteration, respectively, and 10 ns per window using standard US.

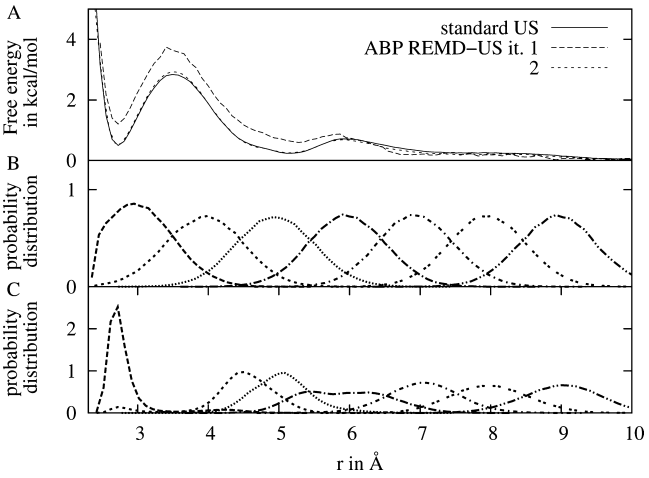


Figure 3. (A) Na–Cl dissociation PMFs, calculated with standard umbrella sampling (solid line) and adaptively biased replica exchange umbrella sampling (dashed lines). The second iteration of the ABP-REMD-US implementation converges to the PMF obtained by standard US. (B) Probability distributions of the separation r from the second iteration of the ABP-REMD umbrella simulations. The free energy landscape is flattened by previous PMF estimates. The distributions approximate Gaussians according to the harmonic umbrella potentials. (C) Probability distribution of the separation r for standard US. Sparsely sampled high free energy regions affect the sampling efficiency.

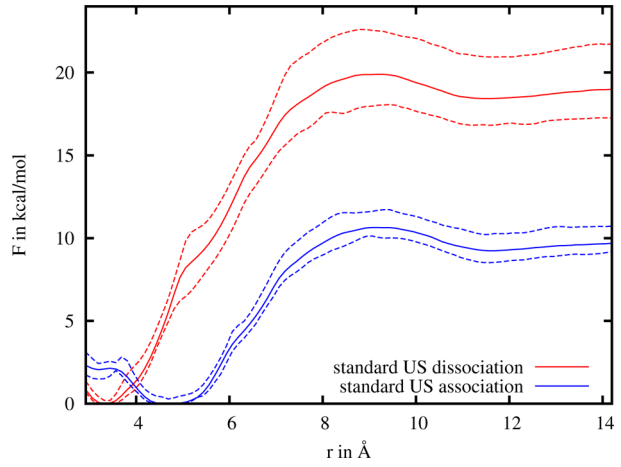


Figure 4. Standard umbrella simulations with restraints on the ligand did not yield converged PMFs for the d(CGCGAATTCGCT)$_2$–furamidine complex. PMFs for dissociation (red) and association (blue) were obtained from a total simulation time of 160 ns each. Lower and upper error estimates are depicted as dashed lines.

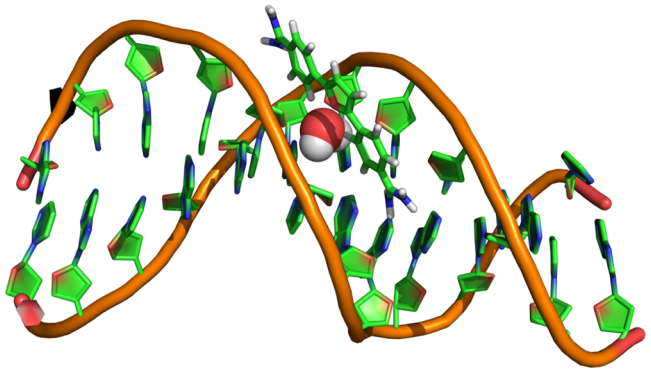


Figure 5. Example of a water molecule (van der Waals spheres) trapped between the DNA and the furamidine ligand frequently observed at close distance umbrella sampling windows.

Within a very small residual difference and statistical error (compare dotted and continuous line in Figure 3A), both methods result in the same PMF, and the probability distributions from the adaptively biased runs approximate Gaussian distributions centered at each reference distance as expected (Figure 3B). Note that even in this case standard US yields reaction coordinate distributions that do not overlap very well in certain distance regimes (Figure 3C). This may limit the accuracy of a calculated PMF.

Furamidine in complex with the minor groove of DNA forms a much more challenging system, in particular because the unbound DNA is strongly hydrated and the minor groove binding region includes bound water molecules. Here, we considered both dissociation and association of the furamidine. Starting configurations for the umbrella simulations were either the bound complex or a configuration with the ligand located in the bulk (without any contacts to the DNA besides the long-range electrostatics, see Methods). Association and dissociation processes are expected to result in the same free energy change; the purpose of the comparison is to verify the convergence of the free energy simulations and the ability of the applied method to reproduce the reversibility of the pathway.

A first attempt to obtain the binding free energy using standard umbrella sampling with phase space restraints (as illustrated in Figure 2) failed to yield converged PMFs (Figure 4). Despite a total sampling time of 4 ns per US window, the dissociation and the association PMFs are far from coinciding even within the relatively large error bars. Inspection of the sampled structures indicated that water molecules are trapped between ligand and DNA during the binding process (example shown in Figure 5), strongly disfavoring the bound configuration.

A main source of limited convergence is free energy barriers along the reaction coordinate that are rarely overcome during the simulation. High free energy barriers might occur during the equilibration process, e.g., in the form of a trapped water molecule, and prevent relaxation of the system. They can also constitute a characteristic of the free energy landscape of the equilibrated system, which may lead to undersampling of parts of the reaction path (as shown for the above ion-separation case).

In the present study, biasing potentials were added in order to address these issues by flattening the free energy landscape. The biasing potentials correspond to adapting PMF estimates that are obtained during an iterative US simulation process. PMF estimates obtained by umbrella simulations with gradually smoother free energy landscapes and therefore improved sampling behavior are calculated until convergence is achieved. This iterative procedure was carried out for both the association and dissociation simulations independently and is in detail described in the Methods section.

A second convergence issue consists in the restricted mobility of the ligand and water molecules within a US window. This may inhibit equilibration of the hydration structure within reasonable sampling time and was observed particularly in the close range region (region I) of the reaction pathway.

To improve sampling in this respect, exchanges of configurations between neighboring US windows were included

dx.doi.org/10.1021/ct400689h | J. Chem. Theory Comput. 2014, 10, 703–710

for the close range region using the H-REMD-US method (see Methods). Via the replica exchanges, the mobility of ligand and specific water molecules is effectively extended from the range of one umbrella window only to the whole range of windows connected via H-REMD. The mean relative exchange rates from the last five dissociation and association iterations lie between 0.24 and 0.37. A detailed documentation of the exchange rates for the individual iteration steps can be found in Table 2S in the Supporting Information.

The distance distributions from the dissociation and association standard US simulations show substantial differences only for distances below 8.9 Å. Therefore, the association process was started from a distance of 8.9 Å (covering regions I and II) in subsequent simulations.

After 16 ABP iterations for the association and 14 iterations for the dissociation process (see Table 1S SI for the detailed iteration protocol), excellent convergence of the PMFs and agreement between both binding pathways was achieved (Figure 6).

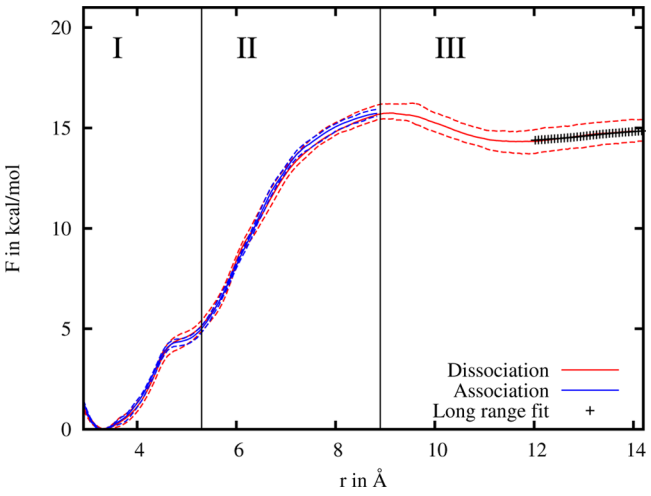


Figure 6. Calculated PMFs for the dissociation (red) and the association (starting from r = 8.9 Å, blue) pathways. Lower and upper error estimates are depicted as dashed lines. The three different regions of the reaction coordinate distinguished in the Methods section are indicated. A Debye−Hückel like model (see Methods) was fitted to the PMF values for distances greater than 12.0 Å in order to estimate the electrostatic long-range effects (crosses).

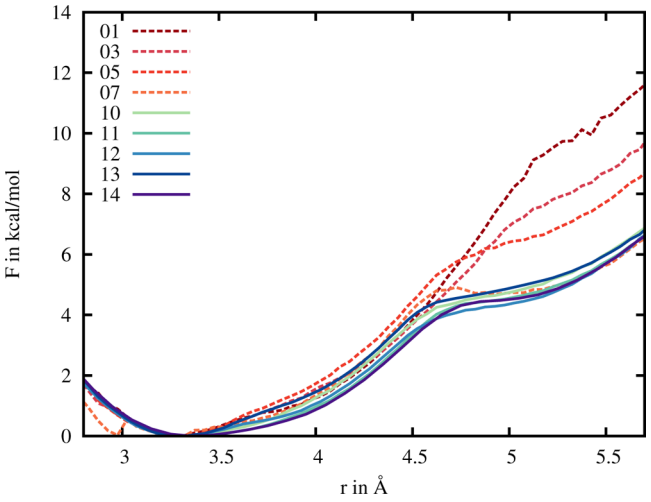


Figure 7. Selected adaptive biasing iterations for the dissociation at close distances (region I) using ABP-REMD umbrella simulations, starting from the bound configuration. The sampling from the last five iterations (solid lines) was used for the final PMF calculation.

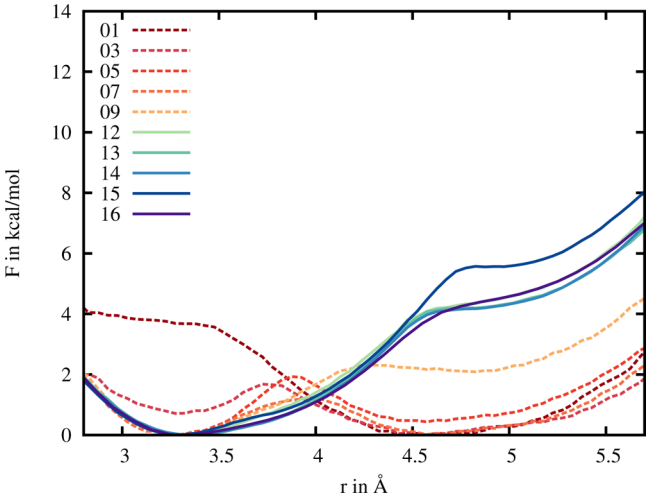


Figure 8. Selected adaptive biasing iterations for the association at close distances (region I) using ABP-REMD umbrella simulations, starting from a bulk configuration. The sampling from the last five iterations (solid lines) was used for the final PMF calculation.

It is of interest to inspect the evolution of the PMFs during the iteration process. During the first iterations (dashed lines in Figures 7 and 8), both association and dissociation PMFs showed significant differences especially at close distances between the ligand and DNA. The reverse run curves in particular are heavily influenced by water molecules trapped in the tight binding pocket and do not predict the correct free energy minimum position in the bound state. The first PMF estimates for the association show a free energy minimum at r = 4.5 Å which does not coincide with the native binding distance of 3.2 Å. This is similar to what was observed for the standard US simulations and leads to a far too low binding free energy prediction at the early stages of the iterative process.

During the iteration procedure, the PMFs for association and dissociation converged to coinciding curves. The main structural barriers are overcome within roughly 5.5 ns of simulation time per window for the dissociation and 8.5 ns per window for the association pathway. In comparison, the standard US simulation PMFs did not show any tendency to coincide during 4.0 ns of sampling time per window.

The association PMF converged more slowly than the dissociation PMF. This is to be expected because the bound configuration is not given by the starting configurations of the dissociation simulations and needs to be explored during the simulation.

All binding free energy contributions are listed in Table 1. I^*, which includes the radial PMF, constitutes the main contribution. The long-range electrostatic contribution was calculated assuming the validity of a Debye−Hückel screening model (see Methods) beyond r = 14.9 Å and amounted to $F(r^*) = -1.27 \pm 0.16$ kcal/mol. It should be emphasized that the determination of the long-range electrostatic effects is considered an estimate, utilized in order to enable direct comparison of the simulation data to experimental results. The approach could yield useful long-range contribution estimates also for other systems. The finally calculated binding free

dx.doi.org/10.1021/ct400689h | J. Chem. Theory Comput. 2014, 10, 703−710

energy as a sum of all contributions is in excellent agreement with experimental results (Table 1).

In addition to the free binding energy, the minor groove width during the DNA−furamidine complex formation was investigated. The width was measured as the distance between the centers of mass of the phosphorus atom groups (7,8,9@P) and (20,21@P, see Figure 1S, Supporting Information) flanking the binding site (Figure 9). The minor groove width vs corresponding DNA−furamidine separation r was extracted from the trajectories generated in the US windows. The sampling errors were estimated by subrun comparison (see Methods). Note that the binding pathway dictated by the restraints may not necessarily correspond to the natural binding pathway.

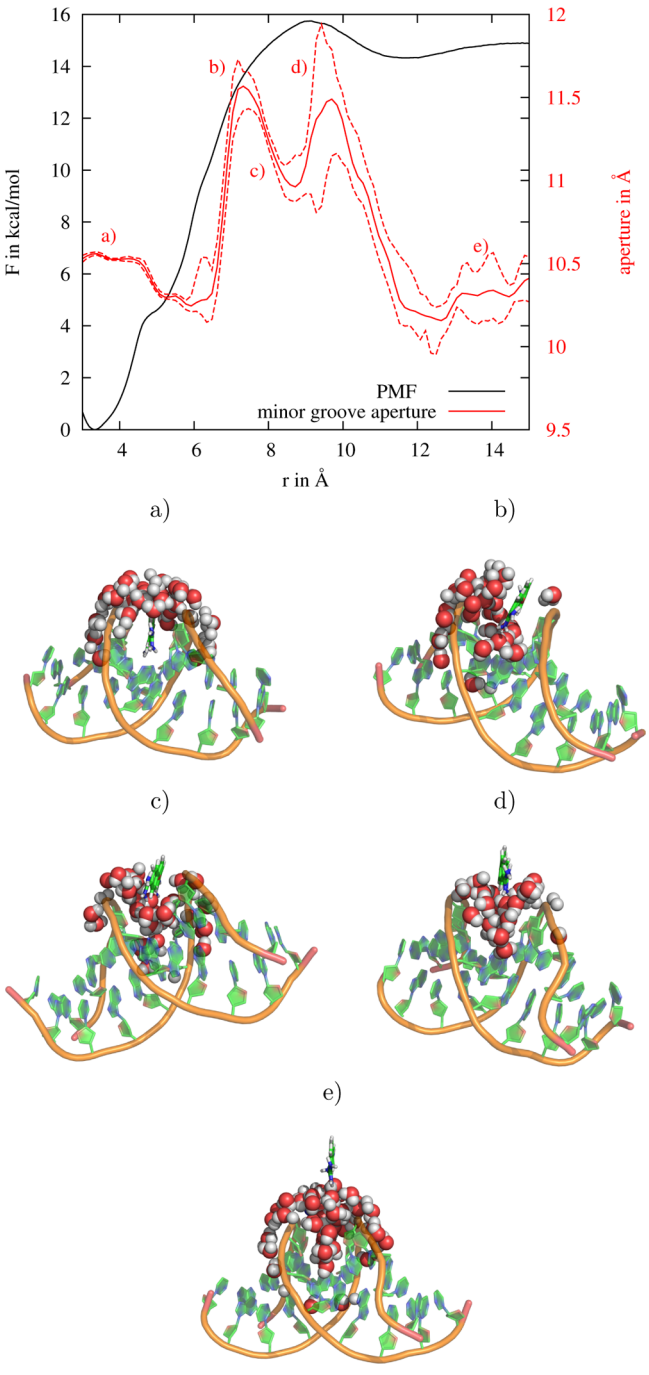


Figure 9. Binding PMF and width (aperture) of the minor groove (upper panel), measured as the distance between the centers of mass of the phosphorus atoms (7,8,9@P) and (20,21@P). Lower and upper error estimates are depicted as dashed lines. The sampling snapshots (a−e) correspond to different DNA (cartoon)−ligand (sticks) separations r and are indicated in the plot. Water molecules in the vicinity of the binding site are represented as van der Waals spheres.

Remarkably, at close and large distances between ligand and binding site, the minor groove width is small and very similar. The width is enlarged by ∼1.2 Å at intermediate distances and shows a double peak with a peak separation of ∼2.5 Å (labeled b and d in Figure 9). This separation corresponds to approximately one hydration water layer.

Starting from the ligand-bound state, the first maximum of the minor groove width occurs at a ligand−DNA separation that is slightly larger than the size of one hydration layer. A comparison of representative snapshots taken of a bound ligand−DNA configuration (labeled a in Figure 9) and of a configuration with a separation corresponding to the first maximum of the minor groove width (Figure 9b) indeed indicates a water molecule layer between the ligand and one DNA strand leading to a slight widening of the minor groove. With the ligand being further pulled out, the minor groove begins to close because the single water layer cannot completely fill the available space between ligand and DNA (Figure 9c). At a distance r of roughly 10 Å, a more expanded water layer (Figure 9d) can be formed between the ligand and both of the DNA strands, and the minor groove is widened again. In the unbound state (Figure 9e), the minor groove is hydrated by a thin water layer which leads to a minor groove width slightly closer than the ligand-bound state width.

CONCLUSION

Restrained umbrella sampling supported by H-REMD and the inclusion of adaptive biasing potentials was used to calculate a radial PMF for an example of DNA−ligand binding. The approach showed significantly improved convergence in comparison to standard restrained US and excellent agreement of PMFs for dissociation and association pathways. The inclusion of both the ABP and the H-REMD-US techniques contributed to the sampling enhancement. The resulting binding free energy of −9.14 ± 0.90 kcal/mol is confirmed by experimental measurements (−9.0/−9.3 kcal/mol).

Phase space reduction due to the introduction of restraints generally leads to higher free energy barriers in US. This can be a major drawback in standard US methods but is accounted for by the sampling robustness of the combined approach. The addition of biasing potentials along the reaction coordinate comes with the beneficial side effect of increased exchange rates in H-REMD-US due to augmented overlap of sampled states in neighboring US windows.

Different techniques of calculating absolute binding free energies have recently been compared on an example indicating that H-REMD-US does not improve sampling efficiency.[15] In the present system, persistent trapping of water molecules between the ligand and DNA was found to be one reason for limited convergence and could be avoided using the H-REMD-US method.

The metadynamics method[18] constitutes an approach to directly obtain a free energy change along a reaction coordinate by accumulating a biasing potential. Biasing potentials are however only added locally around a sampled region in order to iteratively fill a free energy basin. Such an approach may converge slowly because it depends on efficient diffusion of the

system along the reaction coordinate. Using the method presented here, adaptive biasing potentials are calculated from scans along the entire reaction coordinate because the system is driven by the US potentials. The iterative procedure allows controlling the calculated PMFs at different stages of the simulation over the complete range of the reaction coordinate. It is generally applicable for any other ligand−receptor system.

The investigated minor-groove ligand is a relatively rigid molecule. As demonstrated, even in such a case, the accurate calculation of a PMF can be challenging. The application of the approach to ligands that undergo significant configurational transitions remains to be tested.

■ ASSOCIATED CONTENT

Supporting Information

Detailed iteration protocol, H-REMD exchange rates, and a figure with atom specifications. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +49-89-289-12335. Fax: +49-89-289-12444. E-mail: zacharias@tum.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Sonderforschungsbereich 863 of Deutsche Forschungsgemeinschaft and performed using supercomputer resources at the Leibniz Center provided by a grant to PRACE project pr89tu.

■ REFERENCES

(1) *Free energy calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer: New York, 2007.

(2) Woo, H.-J.; Roux, B. Calculation of absolute protein−ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825−6830.

(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*, 9535−9551.

(4) Rodinger, T.; Howell, P. L.; Pomes, R. Absolute free energy calculations by thermodynamic integration in four spatial dimensions. *J. Chem. Phys.* **2005**, *123*, 034−104.

(5) Ytreberg, F. M.; Zuckerman, D. M. Simple estimation of absolute free energies for biomolecules. *J. Chem. Phys.* **2006**, *124*, 104−105.

(6) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.

(7) Buelens, F. P.; Grubmüller, H. Linear-scaling soft-core scheme for alchemical free energy calculations. *J. Comput. Chem.* **2012**, *33*, 25−33.

(8) Luitz, M. P.; Zacharias, M. Role of tyrosine hot-spot residues at the interface of colicin E9 and immunity protein 9: A comparative free energy simulation study. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 461−468.

(9) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047−1069.

(10) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005**, *123*, 084108.

(11) Lee, M. S.; Olson, M. A. Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophys. J.* **2006**, *90*, 864−877.

(12) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(13) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.

(14) Jiang, W.; Roux, B. Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2559−2565.

(15) Gumbart, J. C.; Roux, B.; Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013**, *9*, 794−802.

(16) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(17) Jorgensen, W. L. Interactions between amides in solution and the thermodynamics of weak binding. *J. Am. Chem. Soc.* **1989**, *111*, 3770−3771.

(18) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(19) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169−9183.

(20) Babin, V.; Roland, C.; Sagui, C. Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* **2008**, *128*, 134101.

(21) Huber, T.; Torda, A.; Gunsteren, W. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695−708.

(22) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(23) Grossfield, A. *WHAM: the Weighted Histogram Analysis Method 2.0.7*; Grossfield Lab: Rochester, NY, 2013.

(24) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber12*; University of California: San Francisco, CA, 2012.

(25) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.

(26) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.

(27) Laughton, C. A.; Tanious, F.; Nunn, C. M.; Boykin, D. W.; Wilson, W. D.; Neidle, S. A Crystallographic and Spectroscopic Study of the Complex between d(CGCGAATTCGCG)2 and 2,5-Bis(4-guanylphenyl)furan, an Analogue of Berenil. Structural Origins of Enhanced DNA-Binding Affinity. *Biochemistry* **1996**, *35*, 5655−5661.

(28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(29) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.

(30) Mazur, S.; Tanious, F. A.; Ding, D.; Kumar, A.; Boykin, D. W.; Simpson, I. J.; Neidle, S.; Wilson, W. D. A thermodynamic and structural analysis of DNA minor-groove complex formation. *J. Mol. Biol.* **2000**, *300*, 321−337.