JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

RNA/Peptide Binding Driven by Electrostatics—Insight from Bidirectional Pulling Simulations

Trang N. Do,[†] Paolo Carloni,[‡] Gabriele Varani,[§] and Giovanni Bussi*[,†]

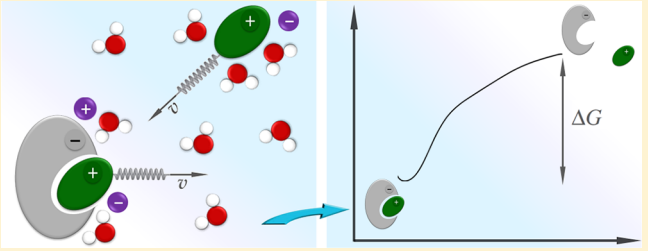
[†]SISSA/ISAS - International School for Advanced Studies, Trieste 34136, Italy

[‡]Computational Biophysics, German Research School for Simulation Sciences, D-52425 Jülich, Germany and Institute for Advanced Simulation IAS-5, Computational Biomedicine, Forschungszentrum Jülich, D-52425 Jülich, Germany

[§]Department of Chemistry and Department of Biochemistry, University of Washington, Seattle, Washington 98195-1700, United States

Supporting Information

ABSTRACT: RNA/protein interactions play crucial roles in controlling gene expression. They are becoming important targets for pharmaceutical applications. Due to RNA flexibility and to the strength of electrostatic interactions, standard docking methods are insufficient. We here present a computational method which allows studying the binding of RNA molecules and charged peptides with atomistic, explicit-solvent molecular dynamics. In our method, a suitable estimate of the electrostatic interaction is used as an order parameter (collective variable) which is then accelerated using bidirectional pulling simulations. Since the electrostatic interaction is only used to enhance the sampling, the approximations used to compute it do not affect the final accuracy. The method is employed to characterize the binding of TAR RNA from HIV-1 and a small cyclic peptide. Our simulation protocol allows blindly predicting the binding pocket and pose as well as the binding affinity. The method is general and could be applied to study other electrostatics-driven binding events.



1. INTRODUCTION

Peptidic ligands are providing increasingly attractive drug leads for many proteins. They have recently been applied to RNA as well.[1−3] Targeting RNA is, however, more challenging than targeting proteins due to the highly charged nature and flexibility of RNA.[4−8] In addition, designing RNA-binding drugs is limited by the poor understanding of RNA/ligand molecular recognition events.[9−12] The latter are intricate processes strongly dominated by a conformational selection and/or induced fit mechanisms.[13,14] Structural information by X-ray and NMR experiments is crucial to provide insight on the binding poses of RNA/peptide complexes.[15] NMR techniques further give useful information on dynamics.[16,17] Computational approaches have proved to be useful for studying nucleic-acid/ligand complexes. On one hand, a docking approach using coarse-grained modeling allows an efficient and systematic search for the native-like binding poses.[18] On the other hand, Molecular Dynamics (MD) simulations with implicit-solvent representations have shown to provide a relatively good agreement with experiments in structural and dynamical properties of the complexes.[12,19,20] Finally, using an explicit representation of solvent and ions could shed light on the RNA structural adaptation and molecular recognition under realistic conditions of the ionic solution, which is known to play important roles in RNA stability and intermolecular electrostatic interactions.[21,22]

However, an explicit representation of solvent and ions implies a significant increment in the number of atoms to be simulated.[23] Therefore, a well-known difficulty of atomistic MD simulations is that they can only follow the system dynamics on the microsecond time scale at most. Although recent developments allowed reaching the millisecond time scale at least for small proteins,[24,25] the study of slower conformational transitions of larger molecules requires some form of acceleration. Several methods have been proposed to alleviate this problem, and many of them are based on the *a priori* choice of proper reaction coordinates, or Collective Variables (CVs), which are biased during the simulation (see, e.g., refs 26−35). A common difficulty of these algorithms is to properly choose an effective CV which is highly dependent on the specific problem. Several authors have proposed to bias the potential energy of the system as a solution.[36−40] However, the potential energy should not be interpreted as a proper CV in solvated systems where large fluctuating contributions arise from the solvent−solvent interactions. Indeed, these methods are efficiently used in solvated systems in a spirit more related to that of parallel tempering,[41] simulated tempering,[42] and multicanonical sampling,[43] i.e., to let the system evolve in an ensemble where effective barriers are decreased and conformational transitions are more likely to occur. In this respect, it appears fascinating to

Received: November 12, 2012
Published: January 17, 2013

push this idea further and to use as a CV only the component of the energy which is relevant for the transition of interest such that the solvent fluctuations are averaged out.

In this paper, we propose to use an approximation of the electrostatic interaction free-energy between two molecules as a CV for binding problems. This free energy can be easily computed on the fly within the Debye–Hückel formalism and can be used as a descriptor to distinguish the bound (low free energy) and unbound (high free energy) states. The so-called Debye–Hückel energy (DHEN) is an approximate evaluation of the intermolecular interactions and is here merely used as a CV on top of which a suitable bias is added. In this manner, the accuracy of the free energy calculation is not dictated by the Debye–Hückel approximation.

We here perform bidirectional Steered Molecular Dynamics (SMD) simulations[28,34] with bias applied on our proposed DHEN CV. The case-study system is the HIV Trans-Activation Responsive (TAR) RNA element in complex with a cyclic peptide inhibitor. TAR is a stem-bulge-loop RNA structure, formed by the first 59 nucleotides of the nascent HIV-1 mRNA molecules. Its apical portion (nucleotides G17–G45, see Figure 1a) directly binds to the arginine-rich viral trans-activator of

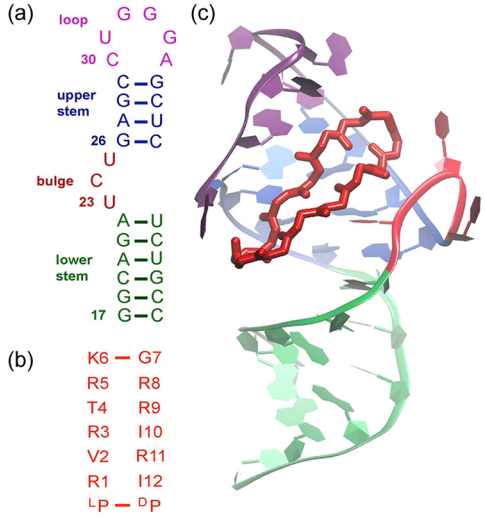


Figure 1. (a) Apical portion (nucleotides G17–G45) of HIV-1 TAR RNA. (b) Sequence of the so-called L22 peptide, a cyclic peptide mimic of Tat protein. (c) NMR structure of the TAR/L22 complex.

transcription protein (Tat), enhancing a key step in the viral replication, i.e., the transcription from the pro-viral DNA into the full-length viral mRNA.[44−47] Therefore, the TAR/Tat interaction is a potential target for developing new anti-HIV drugs. Due to the conserved structure of TAR,[48,49] drug resistance development against TAR-binding inhibitors is expected to be slower than with other conventional HIV-1 protein targets. As a new approach to tackling TAR/Tat interaction, Davidson et al. focused on the development of conformationally constrained mimics of HIV-1 Tat and discovered a family of beta-hairpin cyclic peptides that bind to TAR with nanomolar affinity and greatly improved specificity compared with previous ligands.[50] Among more than 100 peptides in the investigated family, the arginine-rich sequence cyclo-RVRTRKGRRIRIPP (L22 hereafter, see Figure 1b) stands out for its potency. L22 binds to TAR with the affinity of 1 nM. NMR experiments also showed that the L22 peptide binds to the major groove of the upper RNA helix (nucleotides 26–29 and 36–39, see Figure 1c), which is also the binding pocket of Tat.[44,46]

In a recent work, we employed atomistic MD simulations to investigate structural and dynamical properties of the TAR/L22 complex.[51] Our simulations also confirmed that binding between an RNA and a positively charged peptide is a spontaneous process strongly driven by electrostatic interactions. In this work, using bidirectional SMD simulations with our proposed DHEN CV, we are able to predict the correct binding pocket and binding pose without any *a priori* structural information of the complex. We here also generalize a reweighting technique to compute the Potential of Mean Force (PMF) in a bidirectional steering scheme on any *a posteriori* chosen CV, which is not necessarily the steered CV.

2. MATERIALS AND METHODS

2.1. Theoretical Approaches. *2.1.1. Electrostatic Interaction as a Collective Variable.* We construct a CV describing in an approximate manner the electrostatic interaction free energy between the ligand *B* and its target *A*. Our CV aims to be highly general, be computationally efficient, and to use a small number of parameters. This obviously comes at the cost of accuracy, which appears to be well justified here because the approximate electrostatic free energy of the system is used only as a CV for guiding the exploration of the conformational space.

To derive our CV, we make several simplifying assumptions: (*i*) The ionic solution is highly dilute such that the relation $e_c\Phi(\mathbf{r}) \ll k_BT$ holds; here, e_c, k_B, T are the constants denoting the electronic charge, Boltzmann constant, and temperature of the system, respectively, and $\Phi(\mathbf{r})$ represents the electrostatic potential at a position $\mathbf{r}$. (*ii*) The difference in electrostatic interactions due to different atomic sizes are negligible. (*iii*) The ionic solution is considered homogeneous; i.e., the decreasing in solution-screening effect at the region close to or inside the molecules, which causes an increasing in strength of electrostatic interactions, is also negligible. These assumptions do not affect the accuracy of free-energy calculations, which is obtained from the work performed along steered MD simulations.

Let us consider the electrostatic potential $\Phi(\mathbf{r})$ of a target/ligand complex system in an aqueous solution containing a dilute 1:1 electrolyte (K^+ and Cl^- featuring an ionic strength of 10 mM in our case). The electrostatic potential can be calculated by the linearized Poisson–Boltzmann equation (PBE) or the Debye–Hückel equation:[52]

$$-\nabla(\varepsilon(\mathbf{r})\,\nabla\Phi(\mathbf{r})) = 4\pi\sum_{i=1}^{N_m} q_i\delta(\mathbf{r}-\mathbf{r}_i) - \bar{\kappa}^2(\mathbf{r})\,\Phi(\mathbf{r}) \quad (1)$$

where the permittivity $\varepsilon(\mathbf{r})$ adopts appropriate dielectric constant values in different regions of the Debye–Hückel model, the molecule is composed of N_m atoms i at positions $\mathbf{r}_i$ with point charges q_i represented by the Dirac delta functions, $\bar{\kappa}(\mathbf{r})$ is the modified dielectric-independent Debye–Hückel parameter defined as $\bar{\kappa}(\mathbf{r}) = (\varepsilon_w)^{1/2}\kappa$ if $\mathbf{r}$ belongs to the solvent region of Debye–Hückel model and $\bar{\kappa}(\mathbf{r}) = 0$ elsewhere, ε_w is the dielectric constant of water, and κ is the usual Debye–Hückel parameter whose value is given by

$$\kappa = \left(\frac{8\pi N_A e_c^2}{1000\varepsilon_w k_B T}\right)^{1/2} I_s^{1/2}$$

dx.doi.org/10.1021/ct3009914 | J. Chem. Theory Comput. 2013, 9, 1720–1730

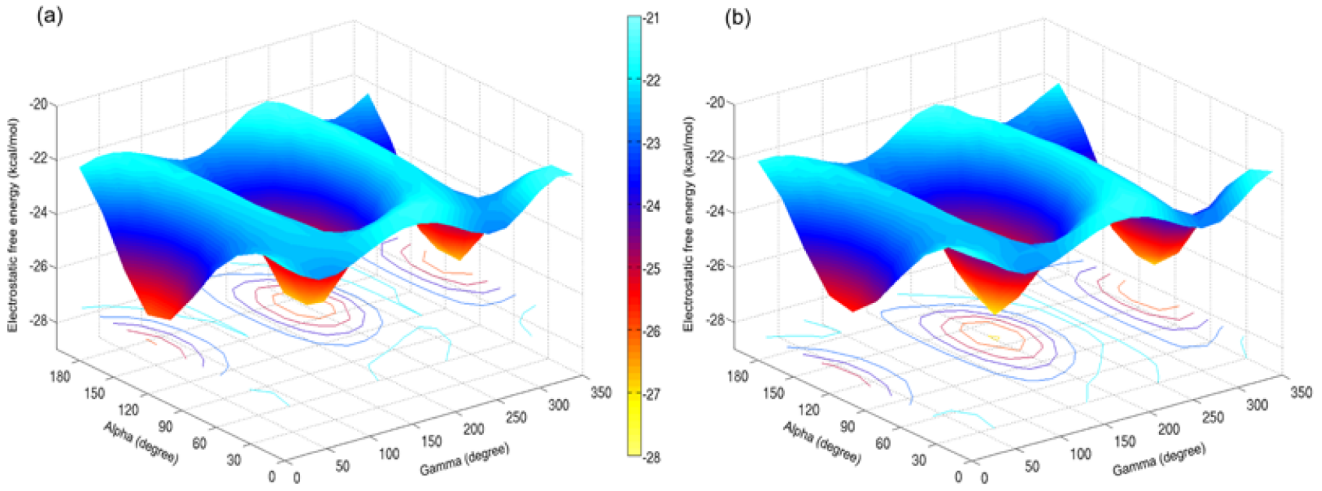


Figure 2. Electrostatic interaction free energy as a function of TAR's orientations calculated by both methods: solving nonlinear PBE (a) and linearized PBE (b). Both calculations agree on two energy minima corresponding to two orientations of TAR at which L22 faces both the upper and lower major groove of TAR.

where N_A is Avogadro's number and $I_s = 1/2\sum_{i=1}^{N_I} c_i z_i^2$ is the ionic strength of the solution determined by N_I types of ions; each type has a charge $q_i = z_i e_c$ and a concentration c_i.

The general solution of eq 1 in the solvent region is given by

$$\Phi(\mathbf{r}) = \frac{1}{k_B T \varepsilon_w} \sum_{i=1}^{N_m} \frac{q_i e^{-\kappa|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|} \quad (2)$$

From the electrostatic potential in eq 2, one can easily derive the electrostatic-interaction term in the free energy of a two-nonoverlapping-molecule system as

$$G^{DH} = \sum_{j\in B} q_j \Phi(\mathbf{r}_j) = \frac{1}{k_B T \varepsilon_w} \sum_{j\in B} \sum_{i\in A} q_i q_j \frac{e^{-\kappa|\mathbf{r}_{ij}|}}{|\mathbf{r}_{ij}|} \quad (3)$$

where A (B) is the set of all the atoms of the target (ligand) molecule, i and j are the atom indexes in the two sets A and B, and $|\mathbf{r}_{ij}| = |\mathbf{r}_i - \mathbf{r}_j|$ denotes the distance between atoms i and j. Here, we introduce a further simplification by assuming that only the intermolecular electrostatic interactions contribute significantly to the free-energy difference between the bound and unbound states of a complex. We thus ignore the intramolecular relaxation in eq 3.

2.1.2. Validation against Poisson–Boltzmann Equation. To compare the estimated electrostatic free energy using nonlinear and linearized PBEs, we performed electrostatic calculations on our case study, the L22–TAR complex system. To build our initial molecular configuration, we first placed L22 at the origin of an arbitrarily selected coordinate system and then placed TAR at 40 Å on an axis, e.g., the x axis in our setup. This choice of distance ensured that L22 and TAR were far enough to have no intermolecular contacts. The dipole moments of both molecules were aligned with another direction, e.g., the z axis in our setup. We next progressively rotated both L22 and TAR molecules about the x, y, and z axes. The combination of these rotations is equivalent to placing the ligand in all three-dimensional rotations everywhere on the surface of a sphere with a radius of 40 Å around the RNA. The "angle-step" of the rotation was 1°. At every step, the electrostatic-interaction free energy was calculated by two methods: (*i*) numerically solving the nonlinear PBE using APBS 1.3[53] and (*ii*) using eq 3, which is resulted from the analytical solution of the linearized PBE. These calculations provided the orientation dependence of the electrostatic interaction.

Obviously, the performance of numerical PBE solving depends on the software, grid spacing, and other procedures as well as input parameters. Although we did not optimize our PBE-solving procedures, the evaluation of our DHEN CV has a clear computational advantage compared to solving PBE numerically. The results, however, were not considerably different from one another. That can be observed in Figure 2, which shows the dependence of electrostatic free energy on the two angles α and γ describing the rotation of TAR about the x and y axes. Both calculation methods agreed on the two minima representing two orientations of TAR at which the electrostatic free energy of the system had the lowest values. In these configurations, L22 faced both the upper and lower major groove of TAR. This result further suggested that there are two possible low-energy funnels in which L22 can approach TAR (see Figure 3). This agreement can become weaker as the two molecules get closer and their low dielectric interiors approach each other. However, under this condition, the full solution of the nonlinear PBE could underestimate the change in molecular flexibility, thus giving a poor estimate of the interaction free energy anyway. We therefore believe that our DHEN provides a good compromise between accuracy and computational cost. These observations provided us with more confidence when using the electrostatic free energy in eq 3 as a CV to accelerate the simulations.

We therefore propose to use the expression (eq 3), hereafter referred to as Debye–Hückel free energy (DHEN), as a CV. The DHEN CV was implemented in an in-house version of PLUMED 1.3.[54]

2.1.3. Reconstruction of the PMF from Bidirectional Pullings. PMF as a Function of the Steered CV. To reconstruct the PMF as a function of the steered CV, we employed the bidirectional PMF estimator developed by Minh and Adib:[55]

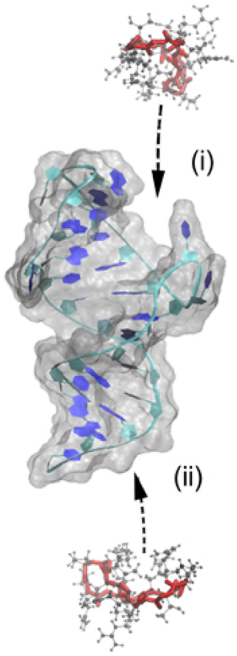


Figure 3. Possible approaches of L22 to TAR predicted by electrostatic-free-energy calculations: (i) upper major groove, which is as well the binding site of Tat, and (ii) lower major groove.

$$G_0(z) = -\beta^{-1} \ln \frac{\sum_t \left[\left\langle \frac{n_F \delta(z - z_t) e^{-\beta W_0^t}}{n_F + n_B e^{-\beta(W-\Delta F)}} \right\rangle_F + \left\langle \frac{n_B \delta(z - z_{\tau-t}) e^{\beta W_{\tau-t}^{\tau}}}{n_F + n_B e^{\beta(W+\Delta F)}} \right\rangle_B \right] e^{\beta \Delta F_t}}{\sum_t e^{-\beta[V(z;t) - \Delta F_t]}} \quad (4)$$

where $G_0(z)$ denotes the PMF as a function of the CV z, τ is the total time of each pulling, z_t is the value of the CV z at time t, $\langle\cdot\rangle_F$ and $\langle\cdot\rangle_B$ denote the averages taken over all forward and backward realizations, respectively, n_F and n_B are the number of realizations in forward and backward pullings, W_0^t is the cumulative pulling work at time t, W is the total work at time τ, $V(z;t) = k[z - z_{restr}(t)]^2/2$ is the harmonic potential acting on the CV at time t, where $z_{restr}(t)$ denotes the value of the restrained CV, and ΔF_t is the free-energy difference between the equilibrium state at time t and the initial equilibrium state of the forward process, whose value is given by

$$e^{-\beta \Delta F_t} = \left\langle \frac{n_F e^{-\beta W_0^t}}{n_F + n_B e^{-\beta(W-\Delta F)}} \right\rangle_F + \left\langle \frac{n_B e^{\beta W_{\tau-t}^{\tau}}}{n_F + n_B e^{\beta(W+\Delta F)}} \right\rangle_B \quad (5)$$

in which $\Delta F = \Delta F_\tau$ is the free-energy difference between the initial and final equilibrium states of the pulling, which can be calculated from the Bennett acceptance ratio (BAR) method.[56] Equation 5 can also be rearranged to give the BAR formula in the particular case where $t = \tau$:

$$e^{-\beta \Delta F} = \left\langle \frac{n_F e^{-\beta W}}{n_F + n_B e^{-\beta(W-\Delta F)}} \right\rangle_F + \left\langle \frac{n_B e^{\beta W}}{n_F + n_B e^{\beta(W+\Delta F)}} \right\rangle_B \quad (6)$$

PMF as a Function of an Arbitrary CV—Reweighting Scheme for Bidirectional Pullings. The PMF computed as a function of the steered CV does not necessarily provide a good picture of the investigated transition, as the steered CV is not guaranteed to properly distinguish all the relevant states. Moreover, in many cases it is instructive to look at the same result from a different perspective, i.e., computing the PMF as a function of a different, *a posteriori* chosen CV. Such a task can be performed by employing a reweighting scheme. Suitable schemes have been proposed for other kinds of nonequilibrium simulations including metadynamics.[57] For SMD simulations, a reweighting algorithm for unidirectional pullings was introduced by some of us in a recent work.[58] Here, we extend this scheme to bidirectional pullings. As introduced in ref 58 at each frame along the ith trajectory, a so-called *weighting factor* can be derived as

$$w_i(t) \propto \frac{e^{-\beta(W_i(t) - \Delta F_t)}}{\sum_{t'} e^{-\beta(V(z_t,t') - \Delta F_{t'})}} \quad (7)$$

where t denotes the time frame, $W_i(t)$ is the work at time t done on the ith pulling, $V(z_t,t') = k/2[z_t - (z_0 + vt')]^2$ is the harmonic potential with a spring constant k and the center of oscillation at $z_0 + vt'$ acting on the CV at time t whose value is z_t, and ΔF_t is a normalization factor that adopts the value of the free-energy difference between the equilibrium state at time t and the initial equilibrium state and can be calculated as in eq 5.

For bidirectional pulling, the weighting factor for both forward and backward trajectories can be written as

$$w_i^F(t) \propto \frac{e^{-\beta(W_i(t) - \Delta F_t)}}{\sum_{t'} e^{-\beta(V(z_t,t') - \Delta F_{t'})}} \quad (8)$$

$$w_i^B(t) \propto \frac{e^{-\beta(W_i(t) - \Delta F)_{\tau-t}}}{\sum_{t'} e^{-\beta(V(z_t,t') - \Delta F_{t'})}} \quad (9)$$

Note that ΔF_t has been adjusted to $\Delta F_{\tau-t}$ in eq 9 since the backward process is actually a reverse of the forward one.

We notice that eqs 8 and 9 do not fix the relative weight of forward and backward trajectories. In the same spirit as in ref 55, this can be optimally done using the weighted-histogram analysis method (WHAM),[59,60] which provides the optimal relative weight. Therefore, the weighting factors in eqs 8 and 9 can be rewritten as

$$w_i^F(t) \propto \frac{e^{-\beta(W_i(t) - \Delta F_t)}}{\sum_{t'} e^{-\beta(V(z_t,t') - \Delta F_{t'})}} \times \frac{1}{n_F + n_B e^{-\beta(W_i(\tau) - \Delta F)}} \quad (10)$$

$$w_i^B(t) \propto \frac{e^{-\beta(W_i(t) - \Delta F_{\tau-t})}}{\sum_{t'} e^{-\beta(V(z_t,t') - \Delta F_{t'})}} \times \frac{e^{-\beta \Delta F}}{n_B + n_F e^{-\beta(W_i(\tau) + \Delta F)}} \quad (11)$$

Equations 10 and 11 are the combinations of the equations derived in refs 55 and 58 and provide a weight to each of the sampled configurations. On the basis of these weights, the PMF can be estimated as a function of any *a posteriori* chosen CV $\bar{z}$ as

$$G(\bar{z}) = -k_B T \log \sum_t \sum_t (w_i^F(t) + w_i^B(t)) \delta(\bar{z} - \bar{z}_t) \quad (12)$$

Our reweighting scheme (eq 12) can be rearranged to give an identical expression to the Minh–Adib bidirectional PMF estimator (eq 4) when $\bar{z} \equiv z$. However, it is more general as it allows estimating the PMF as a function of a different, *a*

dx.doi.org/10.1021/ct3009914 | J. Chem. Theory Comput. 2013, 9, 1720–1730

posteriori chosen CV. When applying this algorithm, one can be flexible in choosing the appropriate CVs for different purposes.

2.2. Computational Details. All standard MD simulations were performed using GROMACS 4.5.5.[61] A truncated octahedral box of explicit water at large size was used to ensure that TAR and L22 can reach the center-to-center distance of at least 40 Å. The box turned out to contain 11 780 water molecules. An excess ion concentration of 10 mM was set in all simulations to reproduce the experimental conditions,[50] which resulted in 23 K^+ cations and two Cl^- anions. We employed TIP3P model[62] for water, the AMBER ff99SB-ILDN force field[63] for L22, and ff99SB-ILDN with parmbsc0 reparametrization[64] for TAR. When using the standard ff99SB-ILDN force field for K^+ and Cl^- ions at an ion concentration of 150 mM, we experienced the growing of ion crystallization after the first 5 ns of MD simulation (data not shown). In fact, the AMBER ff9X force fields, i.e., ff94,[65] ff98,[66] and ff99[67,68] and their variants, have been reported to facilitate the ion crystallization due to the incorrect parametrizations which cause the imbalance between cation−anion interactions.[69−71] Therefore, the ff99SB-ILDN force field corrected with new ions' reparametrization by Joung and Cheatham[72] was used for the K^+ and Cl^- ions in our simulations. In all simulations, temperature was kept constant at 300 K using the velocity rescaling algorithm.[73] When indicated, pressure was kept constant at 1 atm using a Parrinello−Rahman barostat.[74]

A total of 4.2 μs of simulations were performed in a bidirectional steering scheme. Hereafter, we refer to the unbinding direction as *forward* SMD and the binding direction as *backward* SMD.

Our forward SMD scheme included the following: (1f) 20 ns of constant-pressure MD simulation starting from the NMR structure of L22-TAR complex—an average DHEN (−140 kJ/mol) and its standard deviation (σ = 3 kJ/mol) were calculated from this equilibrium simulation; (2f) 64 ns of CV-restrained simulation in which the value of the DHEN was restrained to the value of −140 kJ/mol—a spring constant k = 0.2 kJ mol^{-1} nm^{-2} $\approx k_BT/\sigma^2$ was used here and in the following SMD simulations; (3f) configurations were then extracted every 1 ns and used as initial structures for 64 forward (unbinding) SMD simulations (25 ns each), in which the DHEN was pulled at a constant velocity from the value of −140 to −30 kJ/mol. This target value has been chosen large enough so that the two molecules are completely separated. Indeed, among 64 final structures of unbinding SMD simulations, the smallest center-to-center distance between L22 and TAR is about 32 Å while the smallest distance between an L22 atom and a TAR atom is about 7 Å.

Our backward SMD scheme consisted of the following steps: (1b) Each of the structures obtained at the end of the forward SMD was equilibrated for 1 ns, restraining the DHEN at −30 kJ/mol. (2b) A random configuration was then extracted from each CV-restrained simulation and used as the starting structure for another set of 64 backward (binding) SMD (25 ns each), in which the DHEN CV was pulled from −30 kJ/mol back to −140 kJ/mol with the same velocity and spring constant as in the forward SMD simulations.

Among 64 bound configurations at the end of the backward SMD simulations, we found two dominant classes in which L22 binds to the TAR's upper major groove, i.e., the same binding pocket observed in NMR studies. To further quantify the difference between these two classes, for each class, we selected the structure associated with the lowest steering work. We then repeated the same forward SMD procedure as described above, which, for each selected structure, included (*i*) 30 ns of MD equilibration for step 1f, (*ii*) 12 ns of CV-restraint MD for step 2f at the same DHEN value (−140 kJ/mol) and spring constant (0.2 kJ mol^{-1} nm^{-2}), and (*iii*) 12 forward SMD simulations (25 ns each) for step 3f.

SMD and CV-restrained simulations were performed using our modified version of PLUMED 1.3 integrated with GROMACS 4.5.5. In all of these simulations, eq 3 was used to determine the DHEN CV. All atoms of L22 and TAR are involved in the sets A and B, respectively. Atomic charges were extracted from the ff99SB-ILDN force field. The value of 80 was chosen for the dielectric constant of water ε_w, which, together with the ionic strength I = 10 mM, resulted in the Debye−Hückel parameter $\kappa \simeq 0.033$ Å^{-1}.

3. RESULTS

3.1. PMF Reconstruction from Bidirectional SMD Simulations. The PMF as a function of DHEN was reconstructed from bidirectional pullings using the Minh-Adib PMF estimator as in eq 4 (see Figure 4). However, within this framework, it is difficult to quantify the entropic contribution to the free energy arising from the larger conformational space available as the intermolecular distance increases, which is expected to diverge for the unbound state. We then apply the reweighting scheme (eq 12) to compute the PMF as a function of the distance between the centers of mass of the two molecules (see Figure 5). This entropic contribution can be easily evaluated and added to the PMF and hence allows estimating the free-energy difference between the bound and unbound states, which turns out to be approximately 85 ± 5 kJ/mol. This value is larger than that obtained from the experiment, which was approximately 52 kJ/mol.[50] In the Discussions section, we prove that this discrepancy does not come from sampling problems, and it is presumably due to the inaccuracy of the force fields.

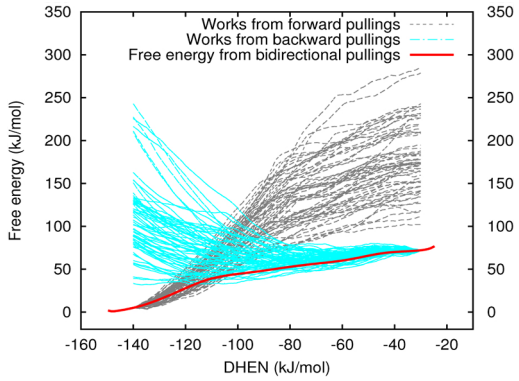


Figure 4. Reconstruction of PMF as a function of DHEN CV (solid red line). The works from forward and backward pullings are also shown (dotted gray and cyan lines, respectively). Backward works are shifted by ΔF as estimated from eq 6.

3.2. Binding Poses from Backward SMD Simulations. A total of 64 binding (backward) SMD simulations, in which the DHEN CV was pulled from −30 to −140 kJ/mol, all ended up with L22−TAR complexes. The binding poses of L22 to TAR can be classified as follows (see Table 1): (i) L22 binds to TAR at the major groove in 51 complexes (80%), among which 33 complexes (52%) can be classified as upper-major-groove

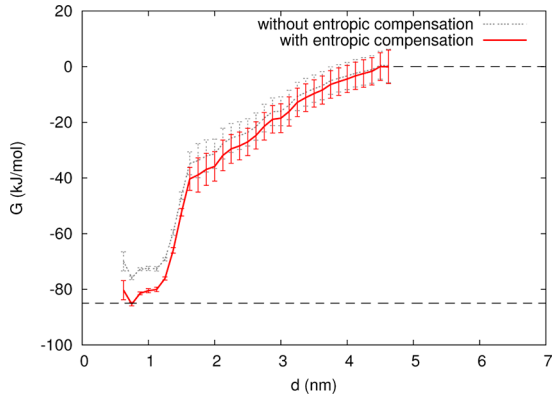


Figure 5. PMF as a function of the geometric center-to-center distance (d) obtained by our proposed reweighting scheme (eq 12) without any entropic compensation (dotted gray line) and with an entropic compensation of $2k_BT$ log d (red solid line). The plots have been shifted so that PMF profiles are aligned at d = 4.5 nm. The projection of PMF onto distance allows defining both bound and unbound states. The free-energy difference between these states is approximately 85 ± 5 kJ/mol. Errors are calculated using the bootstrapping algorithm.

binding (i.e., the same binding pocket as of the Tat protein and as observed in the NMR experiment[50]), 12 complexes (19%) feature lower-major-groove binding, and six complexes (9%) have L22 bind to TAR at the region lying between the upper and lower major groove (referred to as middle major groove hereafter). (ii) L22 binds to TAR at the minor groove in 10 complexes (16%), among which there are four complexes (7%) classified as upper-minor-groove binding and six complexes (9%) classified as lower-minor-groove binding. (iii) Three complexes (4%) do not belong to any of the above categories. In the upper-major-groove binding poses, we found 20 complexes (31%) in which the ^{L}Pro−^{D}Pro template of L22 points outward from TAR (i.e., pose **1** in Table 1 and Figure 6a) and eight complexes (13%) in which the ^{L}Pro−^{D}Pro template points inward to TAR (pose **2** in Table 1 and Figure 6b). Similarly, in the lower-major-groove binding, we also found six complexes (9%) with the ^{L}Pro−^{D}Pro template pointing outward (pose **4** in Table 1 and Figure 6c) and four complexes (7%) with the ^{L}Pro−^{D}Pro template pointing inward (pose **5** in Table 1 and Figure 6d). Pose **2** represents the same binding pose as observed in the NMR experiment.[50] However, it only appears as the second-most probable pose in our simulations. The most probable pose has the same binding pocket but with ^{L}Pro−^{D}Pro pointing to the opposite direction.

A more appropriate assessment of relative pose stability can be done by considering the works performed in the pulling simulations. Among the five simulations with lowest pulling works, only the one with the second-lowest work ended up in pose **3** (see Table 1). The rest of them are classified as pose **1**. The sixth-lowest-work process results in a complex of pose **2**.

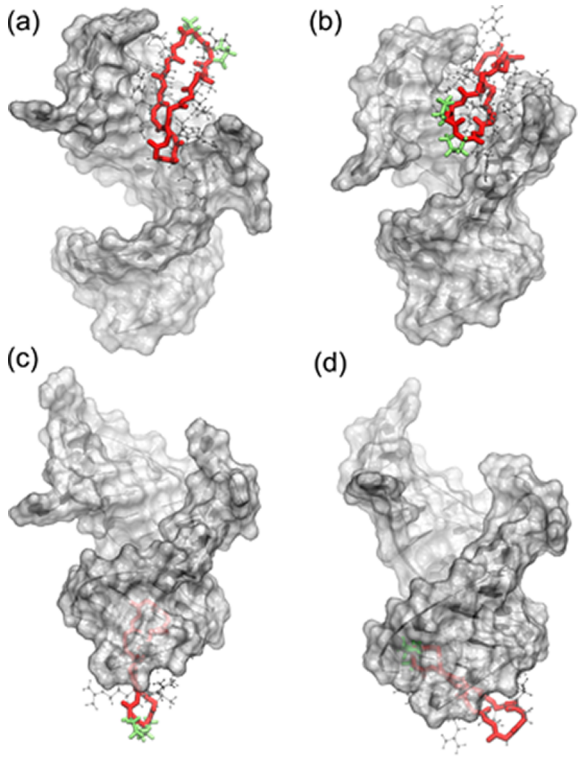


Figure 6. Representatives of four dominant binding poses obtained from 64 binding SMD simulations including (a) upper-major-groove binding pose with the ^{L}Pro−^{D}Pro template, colored in green, pointing outward from TAR (i.e., pose **1** as classified in Table 1), (b) upper-major-groove binding pose with ^{L}Pro−^{D}Pro pointing outward from TAR (pose **2**), (c) lower-major-groove binding pose with ^{L}Pro−^{D}Pro pointing outward from TAR (pose **4**), and (d) lower-major-groove binding pose with ^{L}Pro−^{D}Pro pointing inward to TAR (pose **5**). Pose **2** is close the NMR structure (see Figure 1c).

3.3. Quantitative Comparison in the Stability of the Two Dominant Upper-Major-Groove Binding Poses. As discussed in the previous section, pose **1** in Figure 6 is not only the most probable pose but also the pose for which the least work is performed. In pose **2**, the ligand occupies the same binding pocket in a different orientation, which is consistent with that obtained from NMR data. For a quantitative comparison between pose **1** and pose **2**, we performed 12 unbinding SMD simulations starting from the complex obtained by the least work in each pose (i.e., the globally least work in case of pose **1** and the sixth-least work in case of pose **2**). We then combined these 12 unbinding simulations of

Table 1. Occurrence of L22-TAR Binding Poses Obtained from 64 Binding SMD Simulations[a]

major-groove binding (80%)						minor-groove binding (16%)			others
upper (52%)			lower (19%)						
1	**2**	**3**	**4**	**5**	**6**	middle	upper	lower	
31%	13%	8%	9%	7%	3%	9%	7%	9%	4%

[a]To better classify the binding poses in the major groove, we subdivide the upper-groove and lower-groove binding poses into smaller groups. For the upper groove, **1** represents the pose in which the ^{L}Pro−^{D}Pro template of L22 points outward TAR, **2** denotes the pose with the ^{L}Pro−^{D}Pro template pointing inward TAR, and **3** contains the rest of the upper-major-groove-binding complexes. Similarly, poses **4**, **5**, and **6** respectively represent the same classification criteria for the lower-major-groove binding. It is noteworthy that pose **2** is the binding pose observed in the NMR experiment.

dx.doi.org/10.1021/ct3009914 | J. Chem. Theory Comput. 2013, 9, 1720−1730

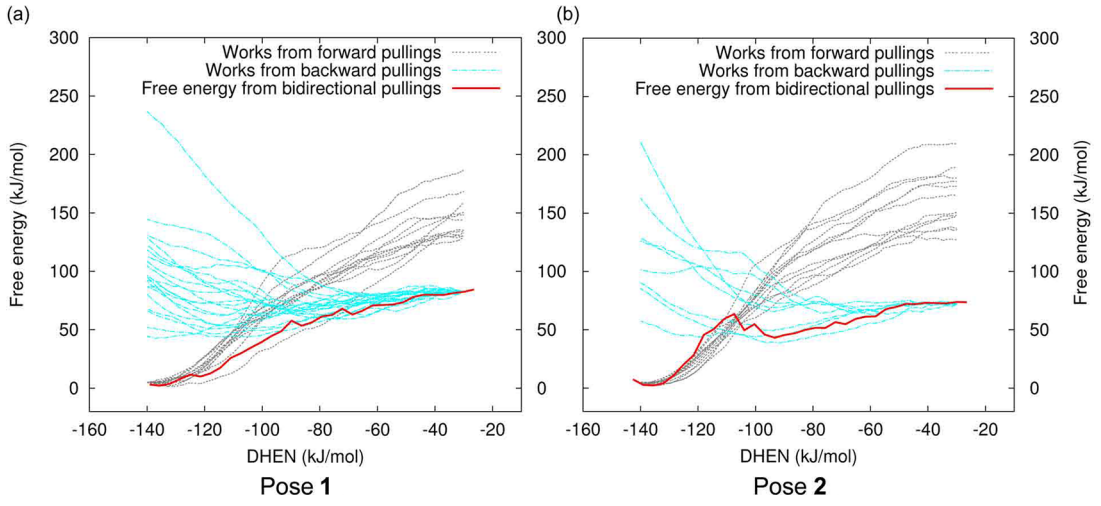


Figure 7. Reconstruction of PMF in pose **1** (a) and pose **2** (b) as a function of DHEN CV (solid red line). The works from forward and backward pullings are also shown (gray lines and cyan lines respectively).

each pose with a selected set of the previous binding simulations which ended on the same pose (i.e., 20 binding simulations resulting in pose **1** and eight simulations resulting in pose **2**). Equation 4 was then used to calculate the PMF as a function of the restrained DHEN CV based on such a combination of each pose (see Figure 7). This equation is applicable even to different numbers of forward and backward trajectories. The free-energy difference between the two end states associated with pose **1** is larger than that of pose **2** (79 versus 69 kJ/mol respectively); thus, pose **1** can be considered more stable than pose **2**.

To verify the robustness of the comparison, we performed a bootstrapping algorithm by repeating the solution of the BAR equation (eq 6; which is a special case of (eq 5) when one considers only the end states) 500 times on the randomly chosen half-set, i.e., six unbinding and 10 binding simulations for pose **1** and six unbinding and four binding simulations for pose **2**. The resulting free energy difference from 500 BAR calculations for each pose can be found in the Supporting Information. The average values of free-energy difference are 75 ± 6 and 66 ± 6 kJ/mol for pose **1** and pose **2**, respectively. Despite the random choice of the half-set of simulations to be involved in BAR calculations, pose **1** consistently shows a higher stability than pose **2**.

Applying the proposed reweighting scheme on the center-to-center distance CV, we again found that the free-energy differences as functions of distance of both poses **1** is larger than that of pose **2** and the NMR structure (see Figure 8).

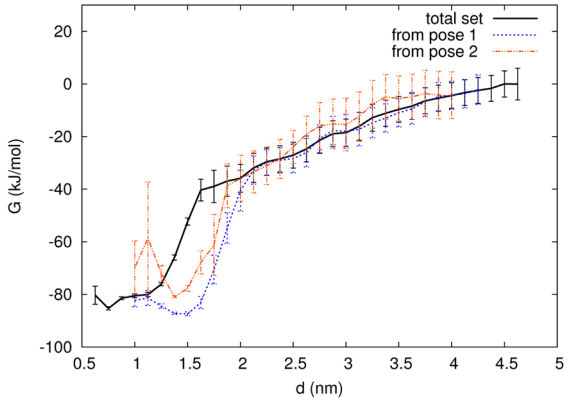


Figure 8. PMF as a function of the center-to-center distance obtained by the reweighting scheme performed on the whole set of simulations starting from the NMR structure (black solid line) and the set of simulations coming from and to pose **1** (blue dotted line) and pose **2** (red dashed line). The free-energy difference between the bound and unbound states of pose **1** is larger than those of the NMR structure and pose **2**.

4. DISCUSSIONS

Effectiveness, Computational Efficiency, and Generality of the Proposed Electrostatic-Based Collective Variable. Our proposed DHEN CV (eq 3) includes only the intermolecular electrostatic interactions and thus does not contain the noise coming from intramolecular interactions or interactions with the ionic solvent. Our CV is a function of not only atomic coordinates but also ionic strength, temperature, solvent dielectric constant, and atomic charges accessible from the force fields. Therefore, it is expected to be more "selective" and effective than the conventional center-to-center distance CV in describing RNA/peptide binding/unbinding processes. Indeed, the distance CV may disfavor the right complex formation by not taking into account the charge−charge interactions, an important driving force in most RNA/peptide recognition events.

Besides the parameters that are trivially determined from the simulation setups, our proposed CV does not require any other extra parameters that need to be updated during the simulations. This is the computational advantage of this CV.

In addition, our proposed CV has a general formalism that, in theory, is applicable for any binding event that is driven by electrostatics, which is the case of most nucleic acid/ligand bindings.

Bidirectional Steering Outperforms Unidirectional Steering. Steering involves out-of-equilibrium processes. Although the unidirectional Jarzynski's equality[75] and the Hummer−Szabo PMF estimator[76] allow reconstructing the free energy from nonequilibrium works, the resulting free energy is strongly dominated by the low work values due to the behavior of the exponential average. These low work values are associated with the trajectories of the "rare" events which are poorly sampled. This poses a convergence challenge to unidirectional SMD simulations. Indeed, the more a system

dx.doi.org/10.1021/ct3009914 | J. Chem. Theory Comput. 2013, 9, 1720−1730

departs from its initial equilibrium state, the more unidirectional estimators tend to overestimate the free energy change.

The bidirectional estimators such as the one introduced by Minh−Adib[55] and our proposed reweighting scheme provide an optimal combination of the works from both forward and backward processes. When the forward and backward processes are properly combined, the overestimations of free energy in both directions are then "averaged" out. Bidirectional estimators hence outperform unidirectional ones.

Verification of the Statistical Accuracy. The bootstrapping procedure of analyzing a selected subset of the binding and unbinding trajectories allows us to assess the statistical accuracy of the results. Indeed, bidirectional pulling protocols can still exhibit large statistical errors when backward pullings do not properly bring the system to the correct starting point. However, by performing the pulling out of the actually reached bound pose, we can guarantee that their stability is fairly evaluated.

"Blind" Prediction of the Binding Pocket. By employing bidirectional SMD simulations with our proposed DHEN CV, we were able to find the binding pocket in agreement with the experimental structure, i.e., the upper major groove. Although the L22−TAR binding pose and its structural properties were well characterized in experiments, we did not use any of these experimental observations to bias or constrain the simulations. Indeed, our pullings were performed in a "blind" manner, in which L22 is not constrained to bind to the known NMR binding pocket but rather free to decide its encounter paths guided by the electrostatic interactions. In such a manner, we are still able to find two dominant binding poses, in both of which L22 binds to the correct pocket as observed by experiments. Figure S3 in the Supporting Information shows the positive correlation between the total works and the last values of DHEN in 64 backward simulations. It is clear that the DHEN CV points to the upper major groove as the best binding site. However, we also observe that the total work is a better estimator of the reliability of the final conformation reached by a steered simulation since (*i*) it takes into account the entire binding process and (*ii*) it has a rigorous foundation which, in the limit of an infinite number of pulling simulations, is independent of the way the DHEN is estimated. We recalculated the DHEN using 100 mM as the ionic strength for the final conformations resulted from the backward steered trajectories. A positive correlation between these new DHEN values and the total works (see Figure S4 in Supporting Information) suggests that this method would be able to distinguish the two grooves also at higher ion concentrations. These results not only confirm the assumption that electrostatics plays an important role in L22−TAR binding but also strongly justifies the use of DHEN, an approximation of the electrostatic interaction free energy, as a CV in accelerated simulations.

TAR is an RNA containing 29 nucleotides featuring two stems, a bulge, and an apical loop. The apical loop partially closes the access to the upper major groove associated with the upper stem, which is also the binding pocket of Tat. Any designed molecule able to bind to TAR at this pocket is a promising candidate for HIV-transcription inhibition. L22 is a ligand with a rigid β-hairpin backbone and long side chains (i.e., mostly composed of arginine side chains), which make it difficult to navigate and end up inside a partially closed pocket. Interestingly, in experiments, L22 was reported to bind and fit completely in this pocket. We were able to reproduce such a nontrivial binding mode only by using SMD simulations pulling on an electrostatic-potential-energy-based CV without any further guidance from experiments. The self-guiding feature of this methodology is very important since it is applicable even when experimental information in unavailable, which is the case of most computational drug design efforts. This feature could further help design a robust and automatic computational strategy for studying biomolecular complexes and hence help reduce experimental expenses.

Predicted Affinity. As we discussed above, we are confident in the statistical accuracy of our calculations. Moreover, contributions such as the conformational entropy loss upon binding are implicitly taken into account in our approach. Thus, the discrepancy between our estimate of the affinity and the one reported in ref 50 should be ascribed to other causes such as the difference between the anion type we used (Cl^-) and the one used in the experiments (mixture of HPO_4^{2-} and $H_2PO_4^-$), the not so large simulation box and the resulting periodic boundary artifacts which are even more amplified in smaller simulation boxes, and finally, force field inaccuracies. These latter inaccuracies could be likely associated with (*i*) the challenging description of the multi-degree-of-freedom sugar−phosphate backbone using a constant-point-charge model,[77,78] (*ii*) the difficulties in describing the RNA noncanonical structural elements,[79−82] i.e., the bulge and hairpin loop in our case, (*iii*) subtle force-field dependence on ionic strength and types,[83,84] and most importantly (*iv*) the well-known inaccuracies of the nonpolarizable force fields in describing the electrostatic interaction between the RNA and the anions (i.e, Cl^- ions) as well as the strong electrostatic interaction between the highly polarizable phosphodiester moiety of RNA and the positively charged atoms[12,85] including the cations K^+ and those of the peptidic ligand L22 in our case.

Sufficiency of DHEN As a CV. A physics-based CV is expected to outperform a geometrical CV in guiding the transition process. However, it is not guaranteed that our DHEN CV is sufficient, since it may not distinguish some degenerate conformations. In this work, the effect of this drawback is reduced by performing multiple steered simulations to sample other possibly hidden relevant degrees of freedom with a brute-force approach. Subsequently, the obtained conformations can be interpreted with their correct statistical weights using our reweighting scheme. In other simulation frameworks, such as metadynamics or umbrella sampling with weighted histogram analysis method, orthogonal CVs could be biased simultaneously with DHEN CV to alleviate this issue.

5. CONCLUSIONS

Motivated by the idea of using the potential energy of a system as a collective variable and based on the observation that binding between an RNA and a positively charged ligand is driven by electrostatic interaction, we have proposed an electrostatic-based collective variable that is an approximation of the intermolecular electrostatic component of the free energy. Our proposed collective variable is a physics-based indicator which is computationally efficient and simple to be integrated with atomistic simulations.[86] By performing bidirectional steered molecular dynamics on this collective variable, and taking advantage of a novel reweighting scheme, we provide quantitative insight on the binding energetics of the L22−TAR complex and blindly reproduce the correct binding pocket and pose as observed in NMR experiments. This shows an efficient self-guiding feature of our simulation protocol,

which is extremely important for drug design when experimental structures are not available. A statistical validation of our results allows the overestimation of the binding affinity to be ascribed to the well-known force-field deficiencies and/or to other differences between the computational and experimental setups. We foresee the application of this technique in studying other electrostatics-driven processes, such as assembly of complexes of charged molecules including ligand/nucleic acid, protein/nucleic acid, and protein/protein interactions. The DHEN collective variable can also be used within the framework of other biasing techniques such as umbrella sampling, adaptive biasing force, and metadynamics, and with temperature-based schemes. As a possible extension, more advanced estimates of the electrostatic interaction free-energy, such as those based on the generalized Born model,[87] could be used as collective variables for biased sampling.

■ ASSOCIATED CONTENT

Supporting Information

A detailed derivation of the Debye–Hückel free-energy collective variable is available. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bussi@sissa.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the CINECA Award no. HP10AQXC5N for the availability of high performance computing resources. G.B. acknowledges the European Research Council for funding through the Starting Grant S-RNA-S (no. 306662). G.V. was supported by NSF grant MCB-1050702. This work has been supported by the BANDO AIDS Grant from the Italian Government for "Targeting HIV transcription to control infection and to purge postintegrative latency".

■ REFERENCES

(1) Froeyen, M.; Herdewijn, P. RNA as a Target for Drug Design, the Example of Tat-TAR Interaction. *Curr. Top. Med. Chem.* **2002**, *2*, 1123–1145.

(2) Zacharias, M. Perspectives of Drug Design that Targets RNA. *Curr. Med. Chem.: Anti-Infect. Agents* **2003**, *2*, 161–172.

(3) Vicens, Q. RNA's coming of age as a rug target. *J. Incl. Phenom. Macrocycl. Chem.* **2009**, *65*, 171–188.

(4) Puglisi, J.; Chen, L.; Frankelti, A.; Williamson, J. Role of RNA Structure in Arginine Recognition of TAR RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 3680–3684.

(5) Auffinger, P.; Westhof, E. Water and Ion Binding around RNA and DNA (C,G) Oligomers. *J. Mol. Biol.* **2000**, *300*, 1133–1131.

(6) Hermann, T. Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition. *Biochimie* **2002**, *84*, 869–875.

(7) Hermann, T.; Tor, Y. RNA as a Target for Small-molecule Therapeutics. *Expert Opin. Ther. Pat.* **2005**, *15*, 49–62.

(8) Auffinger, P.; Hashem, Y. Nucleic Acid Solvation: from Outside to Insight. *Curr. Opin. Struct. Biol.* **2007**, *17*, 325–333.

(9) Gallego, J.; Varani, G.; Targeting, R. N. A. with small-molecule drugs: therapeutic promise and chemical challenges. *Acc. Chem. Res.* **2001**, *34*, 836–843.

(10) Bosshard, H. Molecular recognition by induced fit: how fit is the concept? *News Physiol. Sci.* **2001**, *16*, 171–173.

(11) Boehr, D.; Nussinov, R.; Wright, P. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5*, 789–796.

(12) Fulle, S.; Gohlke, H. Molecular Recognition of RNA: Challenges for Modelling Interactions and Plasticity. *J. Mol. Recognit.* **2010**, *23*, 220–231.

(13) Leulliot, N.; Varani, G. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **2001**, *40*, 7947–7956.

(14) Al-Hashimi, H. Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *Chem. Biol. Chem.* **2005**, *6*, 1506–1519.

(15) Hamy, F.; Felder, E.; Heizmann, G.; Lazdins, J.; Aboul-Ela, F.; Varani, G.; Karn, J.; Klimkait, T. An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 3548–3553.

(16) Lipari, G.; Szabo, A. Model Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.

(17) Lipari, G.; Szabo, A. Model Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570.

(18) Setny, P.; Bahadur, R.; Zacharias, M. Protein-DNA docking with a coarse-grained force field. *BMC Bioinf.* **2012**, *13*, 228.

(19) Banavali, N.; Roux, B. Atomic radii for continuum electrostatics calculations on nucleic acids. *J. Phys. Chem. B* **2002**, *106*, 11026–11035.

(20) Rizzo, R.; Aynechi, T.; Case, D.; Kuntz, I. Estimation of absolute free energies of hydration using continuum methods: accuracy of partial, charge models and optimization of nonpolar contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.

(21) Draper, D.; Grilley, D.; Soto, A.; Ions, R. N. A. folding. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 221–243.

(22) Chen, A.; Marucho, M.; Baker, N.; Pappu, R. Simulations of RNA interactions with monovalent ions. *Methods Enzymol.* **2009**, *469*, 411–432.

(23) Mackerell, A.; Nilsson, L. Molecular Dynamics Simulations of Nucleic Acid-protein Complexes. *Curr. Opin. Struct. Biol.* **2008**, *18*, 194–199.

(24) Shaw, D.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.; Eastwood, M.; Bank, J.; Jumper, J.; Salmon, J.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346.

(25) Lindorff-Larsen, K.; Piana, S.; Dror, R.; Shaw, D. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.

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

(27) Mezei, M. Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias. *J. Comput. Phys* **1987**, *68*, 237–248.

(28) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **1996**, *271*, 997–999.

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

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

(31) Rosso, L.; Minary, P.; Zhu, Z.; Tuckerman, M. On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *J. Chem. Phys.* **2002**, *116*, 4389–4402.

(32) Marsili, S.; Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. Self-healing Umbrella Sampling: A Non-equilibrium Approach for Quantitative Free Energy Calculations. *J. Phys. Chem. B* **2006**, *110*, 14011–14013.

(33) Maragliano, L.; Vanden-Eijnden, E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* **2006**, *426*, 168−175.
(34) Sotomayor, M.; Schulten, K. Single-Molecule Experiments in Vitro and in Silico. *Science* **2007**, *316*, 1144−1148.
(35) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603−020606.
(36) Bartels, C.; Karplus, M. Probability Distributions for Complex Systems: Adaptive Umbrella Sampling of the Potential Energy. *J. Phys. Chem. B* **1998**, *102*, 865−880.
(37) Wang, F.; Landau, D. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.
(38) Micheletti, C.; Laio, A.; Parrinello, M. Reconstructing the Density of States by History-Dependent Metadynamics. *Phys. Rev. Lett.* **2004**, *92*, 170601−170604.
(39) Michel, C.; Laio, A.; Milet, A. Tracing the Entropy along a Reactive Pathway: The Energy As a Generalized Reaction Coordinate. *J. Chem. Theory Comput.* **2009**, *5*, 2193−2196.
(40) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104*, 190601−190604.
(41) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(42) Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett.* **1992**, *19*, 451−458.
(43) Berg, B.; Neuhaus, T. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* **1992**, *68*, 9−12.
(44) Dingwall, C.; Ernberg, I.; Gait, M.; Green, S.; Heaphy, S.; Karn, J.; Lowe, A.; Singh, M.; Skinner, M.; Valerio, R. Human immunodeilciency virus I Tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 6925−6929.
(45) Dingwall, C.; Ernberg, I.; Gait, M.; Green, S.; Heaphy, S.; Karn, J.; Lowe, A.; Singh, M.; Skinner, M. HIV-1 Tat Protein Stimulates Transcription by Binding to a U-rich Bulge in the Stem of the TAR RNA Structure. *EMBO J.* **1990**, *9*, 4145−4153.
(46) Weeks, K.; Ampe, C.; Schultz, S.; Steitz, T.; Crothers, D. Fragments of the HIV-1 Tat Protein Specifically Bind TAR RNA. *Science* **1990**, *249*, 1281−1285.
(47) Calnan, B.; Tidor, B.; Biancalana, S.; Hudson, D.; Frankel, A.; Arginine-mediated, R. N. A. Recognition: the Arginine Fork. *Science* **1991**, *252*, 1167−1171.
(48) Delling, U.; Reid, L.; Barnett, R.; Ma, M.; Climie, S.; Sumner-Smith, M.; Sonenberg, N. Conserved nucleotides in the TAR RNA stem of human immunodeficiency virus type 1 are critical for Tat binding and trans activationmodel for TAR RNA tertiary structure. *J. Virol.* **1992**, *66*, 3018−3025.
(49) Huthoff, H.; Berkhout, B. Mutations in the TAR hairpin affect the equilibrium between alternative conformations of the HIV-1 leader RNA. *Nucleic Acids Res.* **2001**, *29*, 2594−2600.
(50) Davidson, A.; Leeper, T.; Athanassiou, Z.; Patora-Komisarska, K.; Karn, J.; Robinson, J.; Varani, G. Simultaneous Recognition of HIV-1 TAR RNA Bulge and Loop Sequences by Cyclic Peptide Mimics of Tat Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11931−11936.
(51) Do, T.; Ippoliti, E.; Carloni, P.; Varani, G.; Parrinello, M. Counterion Redistribution upon Binding of a Tat-Protein Mimic to HIV-1 TAR RNA. *J. Chem. Theory Comput.* **2012**, *8*, 688−694.
(52) Leach, A. *Molecular Modelling: Principles and Applications*, 2nd ed.; Prentice-Hall: New York, 2001; p 604.
(53) Baker, N.; Sept, D.; Joseph, S.; Holst, M.; McCammon, J. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(54) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.; Parrinello, M. PLUMED: a portable plugin for free energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(55) Minh, D.; Adib, A. Optimized Free Energies from Bidirectional Single-Molecule Force Spectroscopy. *Phys. Rev. Lett.* **2008**, *100*, 180602−180606.
(56) Bennett, C. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(57) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **2009**, *30*, 1615−1621.
(58) Colizzi, F.; Bussi, G. RNA unwinding from reweighted pulling simulations. *J. Am. Chem. Soc.* **2012**, *134*, 5173−5179.
(59) Kumar, S.; Rosenberg, J.; Bouzida, D.; Swendsen, R.; Kollman, P. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(60) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(61) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(62) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(63) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.; Dror, R.; Shaw, D. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.
(64) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T.; Laughton, C.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(65) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(66) Cheatham, T. I.; Cieplak, P.; Kollman, P. A Modified Version of the Cornell *et al.* Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(67) Wang, J.; Cieplak, P.; Kollman, P. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(68) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712−725.
(69) Savelyev, A.; Papoian, G. Inter-DNA Electrostatics from Explicit Solvent Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2007**, *129*, 6060−6061.
(70) Auffinger, P.; Cheatham, T.; Vaiana, A. Spontaneous formation of KCl aggregates in biomolecular simulations: a force field issue? *J. Chem. Theory Comput.* **2007**, *3*, 1851−1859.
(71) Chen, A.; Pappu, R. Parameters of monovalent ions in the AMBER-99 forcefield: Assessment of inaccuracies and proposed improvement. *J. Phys. Chem. B* **2007**, *111*, 11884−11887.
(72) Joung, I.; Cheatham, T., III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(73) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101−014107.
(74) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(75) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.

(76) Hummer, G.; Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3658−3661.
(77) Reynolds, C.; Essex, J.; Richards, W. Atomic charges for variable molecular conformations. *J. Am. Chem. Soc.* **1992**, *114*, 9075−9079.
(78) Cieplak, P.; Cornell, W.; Bayly, C.; Kollman, P. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(79) Orozco, M.; Noy, A.; Perez, A. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185−193.
(80) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T.; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.
(81) Yildirim, I.; Stern, H.; Tubbs, J.; Kennedy, S.; Turner, D. Benchmarking AMBER force fields for RNA: comparisons to NMR spectra for single-stranded r(GACC) are improved by revised χ torsions. *J. Phys. Chem. B* **2011**, *115*, 9261−9270.
(82) Denning, E.; Priyakumar, U.; Nilsson, L.; Mackerell, A. Impact of 2′-hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(83) Besseova, I.; Otyepka, M.; Reblova, K.; Sponer, J. Dependence of A-RNA simulations on the choice of the force field and salt strength. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10701−10711.
(84) Noy, A.; Soteras, I.; Luque, F.; Orozco, M. The impact of monovalent ion force field model in nucleic acids simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10596−10607.
(85) McDowell, S.; Spackova, N.; Sponer, J.; Walter, N. Molecular Dynamics Simulations of RNA: An *In Silico* Single Molecule Approach. *Biopolymers* **2007**, *85*, 169−184.
(86) Our modifications to the PLUMED software are available on request.
(87) Bashford, D.; Case, D. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.