JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Molecular View of Ligands Specificity for CAG Repeats in Anti-Huntington Therapy

Anna Bochicchio,[†,‡] Giulia Rossetti,[*,‡,§,#] Oriana Tabarrini,[∥] Sybille Krauβ,[⊥] and Paolo Carloni[†,‡]

[†]Computational Biophysics, German Research School for Simulation Sciences (GRS), [‡]Computational Biomedicine, Institute for Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9, and [#]Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, Wilhelm-Johnen-Straße, D-52428 Jülich, North Rhine-Westphalia, Germany

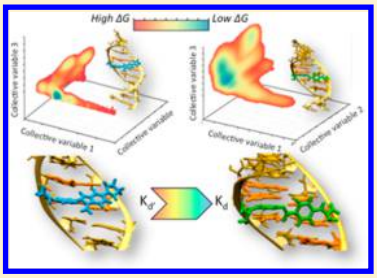
[§]Department of Oncology, Hematology and Stem Cell Transplantation, RWTH Aachen University, D-52074 Aachen, North Rhine-Westphalia, Germany

[∥]Department of Pharmaceutical Sciences, Università di Perugia, Via del Liceo 1, I-06123 Perugia, Perugia, Italy

[⊥]German Center for Neurodegenerative Diseases (DZNE), Sigmund-Freud-Strasse 25, D-53127 Bonn, North Rhine-Westphalia, Germany

Supporting Information

ABSTRACT: Huntington's disease is a fatal and devastating neurodegenerative genetic disorder for which there is currently no cure. It is characterized by Huntingtin protein's mRNA transcripts with 36 or more CAG repeats. Inhibiting the formation of pathological complexes between these expanded transcripts and target proteins may be a valuable strategy against the disease. Yet, the rational design of molecules specifically targeting the expanded CAG repeats is limited by the lack of structural information. Here, we use well-tempered metadynamics-based free energy calculations to investigate pose and affinity of two ligands targeting CAG repeats for which affinities have been previously measured. The first consists of two 4-guanidinophenyl rings linked by an ester group. It is the most potent ligand identified so far, with K_d = 60(30) nM. The second consists of a 4-phenyl dihydroimidazole and 4−1H-indole dihydroimidazole connected by a C−C bond (K_d = 700(80) nM). Our calculations reproduce the experimental affinities and uncover the recognition pattern between ligands' and their RNA target. They also provide a molecular basis for the markedly different affinity of the two ligands for CAG repeats as observed experimentally. These findings may pave the way for a structure-based hit-to-lead optimization to further improve ligand selectivity toward CAG repeat-containing mRNAs.



INTRODUCTION

Huntington's disease (HD) is a devastating autosomal dominant neurodegenerative human disorder,[1] for which there is currently no cure.[2] HD affects 4−10 individuals *per* 100,000 in the Western world.[2] It is characterized by motor, cognitive, and psychiatric disturbances.[1,3] The gene responsible for the disease (HTT) encodes the ubiquitously expressed Huntingtin protein,[4] which is essential for brain development.[5] The disease is caused by an expanded CAG repeat in the 5′-end of the HTT mRNA. This encodes an abnormally long amyloidogenic polyglutamine (polyQ) tract in the Huntingtin protein.[1,4]

HD penetrance correlates with the number n of CAG repeats.[6] Healthy individuals feature $n \leq 35$ (most commonly $17 \leq n \leq 20$).[7] The disease-threshold is $n = 36$.[7] Most HD cases in adults feature $40 \leq n \leq 50$, whereas in young patients $n \geq 50$.[6b] Similar trademarks are also present in a variety of other rarer CAG-expansion-based diseases.[8] One pathogenic mechanisms explaining disease-development is a toxic gain-of-function of excessively long strands of polyQ in the encoded protein.[9] However, other factors can contribute to the pathogenesis of CAG disorders,[10] including the mutant HTT mRNA (mRNA) transcripts with $n \geq 36$.[11] These aberrantly regulate several cellular mechanisms.[11b] Indeed, they bind to proteins and protein complexes in CAG repeat size-dependent manner.[8,12] Among these pathogenic protein-mRNA complexes, the adduct of pathogenic HTT mRNA ($n \geq 36$) with the Midline-1 protein (MID1)/protein phosphatase 2A (PP2A) complex leads to an overproduction of aberrant Huntingtin protein.[12e] Therefore, discovering ligands with high affinity for mRNA transcripts of pathogenic length constitutes a potential strategy against HD and other CAG related disorders.[11b,12a,13]

In a recent and elegant study, Disney and co-workers have succeeded in identifying 10 ligands targeting CAG repeats in RNA oligonucleotides.[12b] The ligands feature aromatic ring(s) with protonable groups. For these compounds, either IC_{50} and/or the dissociation constant (K_d) values were measured.[12b] For two of them, 4-((diaminomethylene)amino)phenyl 4-((diaminomethylene)amino)benzoate (ligand **1**, Figure 1) and 6-(4,5-dihydro-1*H*-imidazol-2-yl)-2-(4-(4,5-dihydro-1*H*-imidazol-2-yl)phenyl)-1*H*-indol-3-amine (ligand **2**, Figure 1),

Received: March 4, 2015
Published: August 31, 2015

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

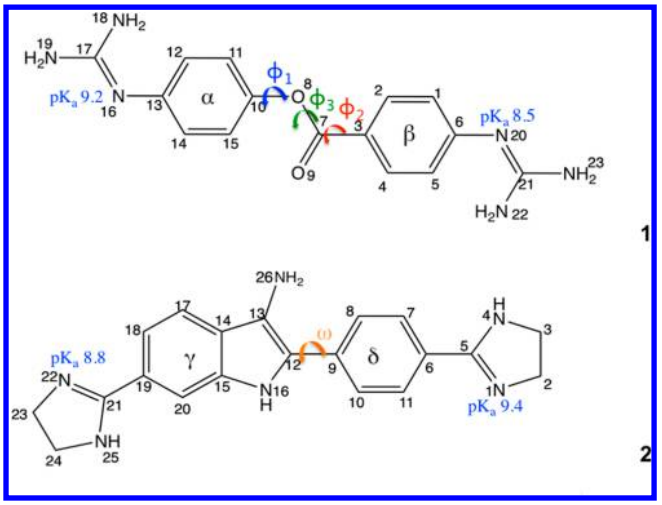


Figure 1. Chemical structure and atomic numbering of ligands **1** and **2**. In **1**, the dihedral angles C7−O8−C10−C11 and O8−C7−C3−C2 (φ_1 and φ_2, respectively), are indicated with a blue and red arrow, respectively. In **2**, the C8−C9−C12−C13 dihedral angle (ω) is indicated with an orange arrow.

both quantities were measured (Table 1). **1** is the most potent ligand for CAG repeats discovered so far,[14] with a K_d of 60(30) nM. None of the ligands identified is selective for the pathogenic CAG repeats versus the nonpathogenic ones.

Table 1. IC_{50} Values and Dissociation Constants (K_d) for Ligands 1 and 2 As Reported in Ref 12b and the Experimental ($\Delta G^0_{(exp)}$) Unbinding Free Energy, As Derived from the K_d, along Calculated ($\Delta G^0_{(calc)}$) Ones, As Obtained by WT-Meta Calculations, Are Also Reported

ligand	K_d (nM)	IC_{50} (μM)	ΔG^{exp} (kcal/mol)	$\Delta G^0_{(calc)}$ (kcal/mol)
1	60 (30)	3	9.9 (0.7)	≈10
2	700 (80)	>125	8.4 (0.1)	≈8

The binding poses of any of these ligands are unknown. This limits the possibility to rationally design new compounds featuring similar affinity for CAG repeats and possibly selectivity toward pathogenic CAG repeats. Computational studies aimed at predicting the binding poses of small ligands targeting r(CUG) and r(CCUG) repeats by combining molecular docking and molecular dynamic simulations in implicit solvent have been reported.[14] [The protocol used consisted of molecular docking based on a series of back and forth pulling of the ligands, followed by molecular dynamics based on the amber99 force field[15] with Yildirim-revised χ[16] and α/γ[17] torsional parameter sets in implicit solvent.] These calculations could in principle be used for structural insights on ligand/CAG repeats complexes. Unfortunately, however, the absence of explicit solvent in the calculations of RNA-containing systems is known to severely affect the accuracy of predictions,[18] which limit greatly the predictive power of the calculations.

Molecular simulations in explicit solvent, based on the latest version of the AMBER force field,[17,19] are revealing themselves as an increasingly powerful tool to investigate RNA in the free form and in complex with ligands,[20] reproducing rather accurately structure, (sub-) microsecond conformational fluctuations and energetics of RNA oligonucleotides in the free state,[21] including one containing CAG repeats,[22] and in complex with ligands.[17,23] Here, we have used molecular dynamics and well-tempered metadynamics (WT-meta)[24] in explicit solvent, with the above-mentioned AMBER force field, to predict binding pose and affinities of the best characterized ligands, **1** and **2**. The WT-meta approach provides the free energy of the process as a function of apt reaction coordinates (or collective variables, CVs). Such calculation does not depend on the choice of the initial structure,[25] and it is particularly suitable to target biomolecular flexibility.[12b]

Our model RNA system is a double-stranded decamer containing two CAG repeats (r(5′-G_1GCAGCAGCC$_{10}$)$_2$.[12a] It involves A-A mismatches between canonical C-G base pairs, as in CAG repeats-containing RNAs.[8,26]

Our calculations uncovered the molecular bases of the different ligands' specificity toward CAG repeats. The calculated free energies are not too dissimilar from those derived by the experimental K_d (Table 1). The simulations provide a molecular basis for the increase of affinity on passing from **2** to **1**. This work sets the basis for structure-assisted design of ligands targeting CAG repeats, which might lead to improved potency.

MATERIALS AND METHODS

Ligands 1 and 2. Their structures were constructed from scratch using the MOLDEN 4.0.5 software.[27] The compounds' protonation state in aqueous solution and at physiological pH values (c.a. 7.4) was determined by predicting their aqueous ionization constants (pK_a), using the Marvin (ChemAxon) code.[28]

The conformational energy of the molecules depends critically on specific dihedral angles. These are the angles around the O8−C10 and C7−C3 bonds in **1** (φ_1, φ_2 in Figure 1) [In **1**, in principle one should consider also the torsion angle around the O8−C7 bond φ_3. However, this angle has been shown to be fairly rigid at a fixed value of 180°[29] with an estimated barrier for rotation of 12 kcal·mol^{-1}.[30] Hence, the conformational energy as a function of this angle was not calculated here.] and the angle around the C9−C12 bond in **2** (ω in Figure 1).[29d] The ligands underwent full geometry optimization for values of these dihedral angles ranging from −180° to 180° at intervals of 10°. The calculations were carried out at the B3LYP/6-31G** level of theory with the PCM implicit solvent model[31] as implemented in the Gaussian 09 code[32] (see Section S1 in the SI for details).

The first principle optimized ligands structures were inserted in a dodecahedral water box of 41 × 41 × 25 Å^3 edges. They underwent 50 ns-long MD simulations using the GROMACS 4.5.5 code.[33] The force field parameters for **1** and **2** were calculated following the standard AMBER procedure, using the *gaff* force field,[34] except for the charges, which were evaluated according to the restricted electrostatic potential fit method.[35] The ligands' conformations determined in the previous step were first optimized at the HF/6-31G(d) level up to a convergence of 10^{-5} atomic units (a.u.) using the Gaussian 09 package. The electrostatic potential maps generated by the molecules were then calculated with the same level of theory. This protocol ensures to extract atomic charges consistent with the AMBER force field.[15] The atomic RESP charges were calculated from the electrostatic potential, using the ante-chamber module of the AMBER12 package[36] (see Table S2.2 in the SI). The overall charge of the ligands was neutralized by adding Na^+ ions. Further Na^+ and Cl^- ions were added so as to reach the ionic strength used in the measurements of the K_d's

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

(0.157 M).[12b] The same procedure of the ions was applied to all simulations in this study. Hence, all systems featured an ionic strength of 0.157 M, and they were electrically neutral.

The TIP3P model[37] and Smith and Dang's[38] force fields were used for water and ions, respectively. Periodic boundary conditions were applied. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method,[39] using a grid with a spacing of 1.2 Å combined with a fourth-order B-spline interpolation to compute the potential and forces in between grid points. The cutoff radius for the Lennard-Jones interactions as well as for the real part of PME calculations was set to 12 Å. The pair list was updated every 2 steps, and the LINCS algorithm[40] was used to constrain all bond lengths involving hydrogen atoms, allowing the use of a time step of 2 fs. MD simulations in the canonical ensemble (at T = 298 K, P = 1 bar) were achieved through the Nosé–Hoover[41] and Parrinello–Rahman[42] coupling schemes. The systems were energy-minimized imposing harmonic position restraints of 2.4 kcal·mol^{-1}·Å^{-2} on solute atoms. After an energy minimization of the solvent and the solute without harmonic restraints, the temperature was gradually increased from 0 to 298 K, in 40 steps of 25 K each, in 1 ns of MD. Finally, 50 ns-long MD simulations were carried out. Further details are reported in the SI (see Sections S1 and S2 in the SI).

CAG-Repeats Containing RNA. Two different crystal structures containing RNA CAG repeats have been reported in the literature: The first is the duplex r(5′-G_1GCAGCA-GCC$_{10}$)$_2$, RNA_{10} hereafter (PDB code 3NJ6,[12a]), resolved at atomic resolution (0.95 Å). It involves A-A mismatches flanking C-G base pairs, as in predicted secondary structure elongated CAG repeats.[8,26]

The second is the duplex r(5′-U_1UGGGC(CAG)3GUC-C_{19})$_2$, (pdb code: 4J50,[22]). Here, three CAG repeats are embedded by 5′-UU dangling ends.[22] This structure presents several features that might make it less amenable for modeling studies than the first structure. First, a recent analysis of the electron density,[43] calculated on the basis of the structure factors deposited in the PDB, indicates structural disorder in correspondence of all the A-A pairs of the structure, affecting the reliability of any modeling based on this structure. [In this structure, one A-A pair is in the *anti-anti* conformation, yet it does not form the C2H2···N1 H-bond interaction as in RNA_{10}. The others exhibit an *anti-syn* conformations.[22,43]] Second, the 5′-UU ends interfere with the A-A pairs, and these are not present in the sequence and in the secondary structure of HTT mRNA.[8,26] Third, the resolution of the structure (1.65 Å) is lower than that of RNA_{10}.

Hence, we performed all calculations using the RNA_{10} structure.

RNA_{10} was inserted at the center of a dodecahedral water box of 62 × 62 × 43 Å^3. This ensures that the solvent shell extends for at least 10 Å around the system. Na^+ and Cl^- ions were added so as to reach electroneutrality and ionic strength of 0.157 M, as done for ligands **1** and **2.**

For the biomolecule, we used the parmbsc0 AMBER force field[17] with the XOL[19] correction. This revised force field version eliminates the formation of ladder-like RNA structures.[19,44] Simulations based on the χ_{OL}+ff99bsc0 force field[17,19] lead to an improvement of the description of A-RNA duplexes, along with UNCG and GNRA tetraloops.[44c] For water and counterions, the same setup as in the previous section was adopted.

300 ns-long MD simulation was carried out using the same computational protocol adopted for the ligands (see Section S2 in the SI). The trajectory was clustered using the gromos method[45] implemented in GROMACS 4.5.5[33] package, and the representative structure of the most populated cluster was used for docking procedure.

The chosen ions parameter set was recently used in other nucleic acids-related MD simulations.[46] None of these simulations describes the formation of ions clusters, as reported in ref 47. This was also the case in our simulation: We did not observe ions-crystallization events. However, to further test the robustness of the chosen parameters, we performed an additional simulation of the system with exactly the same computational setup but using Joung and Cheatham parameters[48] for Na^+ and Cl^-. A section about this is offered in the SI (Section S4 in the SI).

1·RNA_{10} and 2·RNA_{10} Complexes. An educated first guess of **1** and **2**'s binding poses onto RNA_{10} was obtained using the flexible, information-driven docking HADDOCK 2.1[49] code. This procedure was not meant to provide a structural prediction but rather an initial structure for the subsequent molecular simulations.

The *ab initio* geometry optimized structures of the two ligands – very similar to that of the MD conformers in water (Section S1 in the SI) – and the representative MD structure of RNA_{10}, identified by a clustering analysis,[50] were used as input structures for the docking procedure. The HADDOCK 2.1 code defines "active" and "passive" residues[49a,b] (see section S5 in the SI).

Following the idea to target specifically the A-A mismatch of ref 12b, the active residues include the adenine nucleobases (A_4, A_7, A_{14}, and A_{17}) of RNA_{10} along with the ligands. The passive residues include the cytosine C_3, C_6, C_{13} and C_{16}, along with guanine G_5, G_8, G_{15}, and G_{18}. The parameters for the HADDOCK scoring function of RNA_{10} were generated automatically by the program.[49a,b] Those of the ligands were derived using the PRODRG Web server,[51] following the HADDOCK recommended procedure.[49a,b]

The **1**·RNA_{10} and **2**·RNA_{10} complexes with the best HADDOCK scores were immersed in 62 × 62 × 43 Å^3 water boxes for subsequent MD simulations refinement. Na^+ and Cl^- ions were added so as to reach the ionic strength of 0.157 M. The two systems were equilibrated for 200 ns, using the same computational setup described above (details in Sections S2 and S5 in the SI).

The free energy associated with the unbinding of **1** and **2** to RNA_{10} was calculated using the metadynamics[52] method in its well-tempered (WT-meta) variant[24] (see section S6 in SI for details on the method). The choice of the CVs describing the unbinding processes for both complexes was based on previous ligand–target free energy studies[53] as well as on the analysis of the interactions established between the ligands and RNA_{10} in the MD simulations. These are the distance d_{CM} between the centers of mass of the ligands and of the RNA_{10} tract r(5′-C_3AGC AG$_8$)$_2$; the number of hydrophobic contacts n_{HC} between the ligands and the bases which have shown π-stacking interactions in the previous MD simulations modeled as a coordination number:

$$n_{HC} = \sum_{ij} \frac{1 - (r_{ij}/r_0)^a}{1 - (r_{ij}/r_0)^b} \quad (1)$$

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911–4922

The parameters a and b have values of 6 and 12, respectively, while r_0 = 6 Å accounts for the typical carbon−carbon distance (4/4.5 Å) and the thermal motions amplitude (1.5/2 Å). The number of H-bonds n_{HB} between the ligands and the nucleotides that they cover in the starting structure is described also by eq 1 with a = 8, b = 12, and r_0 = 2.5 Å.

In the WT-meta calculation, the widths of Gaussians were chosen to be ∼1/3 of the typical fluctuations of the CVs during the MD simulation, following ref 54. The height of the Gaussian functions was set to ∼0.02 kcal·mol^{-1}, their deposition time to 1 ps, and an initial ΔT to 4,200 K. 700 ns WT-meta simulations were carried out employing the PLUMED 1.3 plugin[55] in combination with GROMACS 4.5.5.[33] We used, as starting structures, the last snapshot of the ligands/RNA$_{10}$ MD simulations.

For both the systems the convergence of the WT-meta run was achieved after 650 ns of simulation (see Figure S6.1 in the SI).

The simulation trajectories were clustered in the space of the three CVs in order to assign configurations to each free energy minimum. We used the *gromos* method,[45] as in ref 53a. The clustering radius of each collective variable was set to 0.1, 0.2, and 0.2 nm for d_{CM}, n_{HB}, and n_{HC} CVs, respectively.

We identified the free energy minima associated with **1**·RNA$_{10}$ and **2**·RNA$_{10}$ bound states (**B** and **B′** in Figure 3a and Figure 3b, respectively) and that relative to partly dissociated states. The latter are the only minima in which ligand/RNA direct contacts are absent (**U** and **U′** in Figure 3). The dissociation free energy value emerging from the WT-meta calculations ($\Delta G^{\text{WT-meta}}$) was then defined as the difference between **B** and **U** (or between **B′** and **U′**) free energy minima. We estimated the enthalpy contributions for each complex to the $\Delta G^{\text{WT-meta}}$ using the relation:

$$\Delta H = \Delta U + p\Delta V \tag{2}$$

The quantities in eq 2 were calculated using the g_energy code of the GROMACS 4.5.5 package.[33] The entropy contributions, $-T\Delta S$, were then determined from the relationship given in eq 3:

$$-T\Delta S = \Delta G^{\text{WT-meta}} - \Delta H \tag{3}$$

The residual dissociation free energy to take the ligands from **U** (or **U′**) to infinite distance (ΔG^{PB}) was approximately estimated by the nonlinear Poisson−Boltzmann equation, using the APBS 1.3 program[56] (see Section S6 in the SI for details). Therefore, the total free energy of unbinding

$$\Delta G^{\text{TOT}} = \Delta G^{\text{WT-meta}} + \Delta G^{\text{PB}} \tag{4}$$

was used to calculate the standard-state free energy of dissociation ($\Delta G^0_{(\text{calc})}$) for each complex:

$$\Delta G^0_{(\text{calc})} = \Delta G^{\text{TOT}} - RT \ln\left(\frac{[L]}{[L]^0}\right) \tag{5}$$

in which R is the ideal gas constant, $[L]$ is the concentration of the ligand in our simulation box (4.7 mM in 190,000 Å^3), and $[L]^0$ is the standard state concentration of 1 mol/L.[53a]

The experimental free energy values ($\Delta G^0_{(\text{exp})}$) were calculated from the K_d in Table 1 as

$$\Delta G^0_{(\text{exp})} = -RT \ln(K_d/[L]^0) \tag{6}$$

Calculated Properties. The average ions populations around the CAG repeats were calculated as in ref 57. Structural features and helical parameters, including base pair (buckle, propeller, opening, shear, stretch and stagger) and base pair step (twist, roll, tilt, rise, shift, and slide) parameters, of the RNA$_{10}$ free and in complex with the ligand, were calculated using the Curves+ code.[58] Intermolecular interactions such as H-bonds and salt-bridges between RNA$_{10}$ and the ligands were analyzed using the g_hbond code of the GROMACS 4.5.5 package.[33]

More details about the Methods used are reported in the SI.

RESULTS AND DISCUSSION

The main goal of this work is to predict binding poses and potency of compounds **1** and **2** toward RNA$_{10}$. We achieved this using a multistep strategy. First, we investigated ligands' conformational properties in aqueous solution after assessing their protonation state at neutral pH. Indeed, the spatial geometry of the two compounds in aqueous solution might affect their binding mode toward the RNA target molecule. Next, we investigated RNA$_{10}$ in water by MD simulations. Then, we docked the ligands onto the RNA$_{10}$ duplex by flexible docking and MD simulations in water. Afterward, we performed well-tempered metadynamics simulations[24] (WT-meta) to investigate the energetics of the unbinding process for each complex, along with electrostatic calculations. The free energies were calculated as a function of collective variables (CVs) already used to study ligand/target interactions. These include the following: the distance between the centers of mass of the two interacting moieties (RNA$_{10}$ and **1** or **2**) (d_{CM}), the number of intermolecular H-bonds (n_{HB}), and the number of intermolecular hydrophobic contacts (n_{HC}). Finally, we compared our results with the free energies as derived by the measured K_d's (Table 1).

Compounds 1 and 2: Protonation States. The K_d's have been measured at pH = 7.4. Our pK_a calculations, based here on the Marvin (ChemAxon) code,[28] suggest that, in water at the experimental pH, both the compounds are double protonated in correspondence of their terminal groups, with a total charge of +2. Indeed, the calculated pK_a values turned out to be equal to 8.5 and 9.5 on N20 and N16 imine nitrogen of **1** and of 8.8 and 9.4 on N22 and N1 nitrogen of **2** (see Figure 1), respectively.

Ligands 1 and 2: Molecular Dynamics Simulations in Water. The conformational properties of **1** depend critically on the dihedral angles φ_1 and φ_2 around the O8−C10 and C7−C3 bonds, respectively (Figure 1).[29d] These dictate the degree of coplanarity of the aromatic rings (α and β in Figure 1). To identify the most populated conformers, we calculated their potential energies as a function of these dihedral angles. Quantum chemistry calculations (see Methods) suggest that in the energy minimum the ligand phenyl rings α and β (Figure 1) are in a quasi-perpendicular orientation. In this configuration, φ_1 = 64° and φ_2 = 0° (Section S1 SI). The conformation is basically maintained during 50 ns long MD simulation in explicit water (Section S1 in the SI for details).

The conformational energy of **2** depends crucially on the ω dihedral angle around the C9−C12 bond (Figure 1). Quantum chemical calculations, as those carried out for **1**, show that in the energy minimum the ligand aromatic rings γ and δ are not coplanar (ω = 28° see Figure 1 and Section S1 in the SI). Such configuration is maintained during the 50 ns long MD simulation in explicit solvent (see section S1 in the SI).

Molecular Dynamics of RNA$_{10}$ in Water. In the RNA$_{10}$ X-ray structure,[12a] solved at high resolution (0.95 Å), the

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

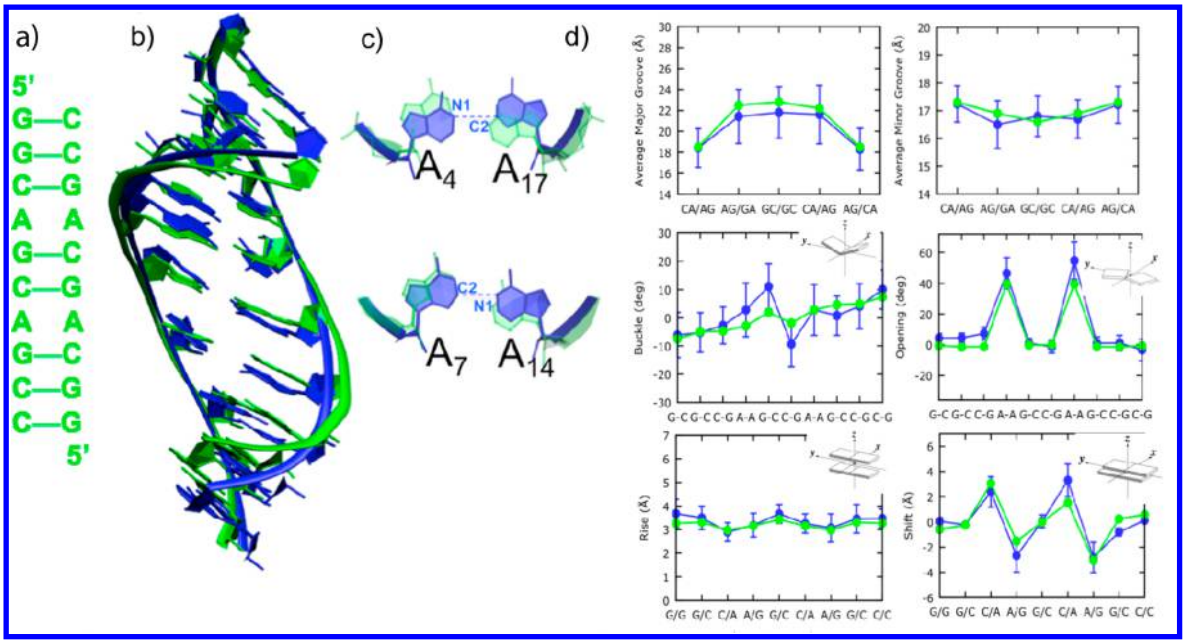


Figure 2. RNA_{10}: Secondary (a) and (b) X-ray (b) and MD (c, blue) three-dimensional structures[12a] c) Close up on the A_4-A_{17} (top) and A_7-A_{14} (bottom) pairs. The adenosines are all in the *anti* conformations with a single C2−H2···N1 H-bond. d) Base pair (opening, shear and stretch; left) and base pair step helical parameters (shift, rise and twist; right) are reported as a function of base pair sequence. Other helical parameters are reported in the SI (Figure S4.2 b) and c), respectively). In c) and d), color-coding is as in b).

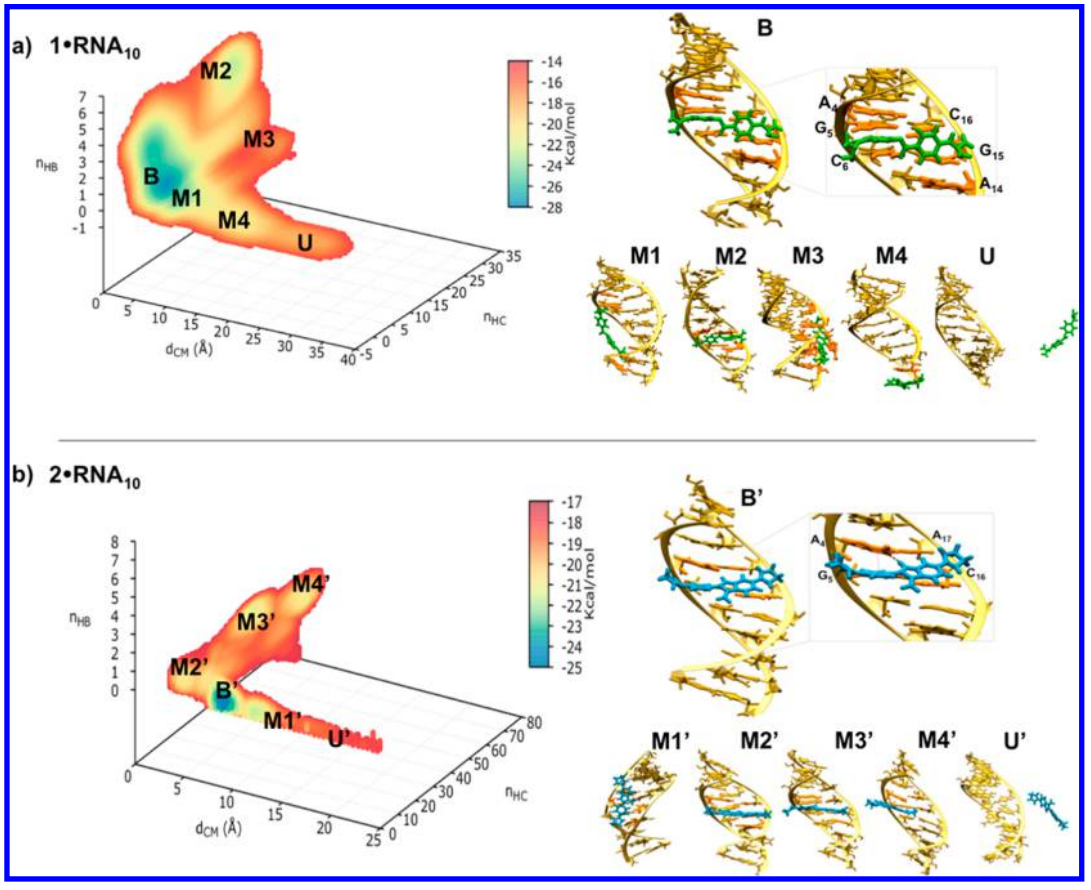


Figure 3. Unbinding free energy profiles (kcal/mol) as a function of the d_{CM}, n_{HC}, and n_{HB} CVs for **1**·RNA_{10} (a) and **2**·RNA_{10} (b) complexes. d_{CM} is the distance between the centers of mass of RNA_{10} and **1** and **2**; n_{HC} is the number of hydrophobic contacts at interface between the two moieties; n_{HB} is the number of H-bonds between the two moieties. **B** and **B′** denote the bound states, while **U** and **U′** represent unbound states' minima for **1**·RNA_{10} and **2**·RNA_{10}, respectively. **M1**-**M4** and **M1′**-**M4′** are local minima. For each minimum, representative structures in cartoon representations are provided on the right.

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

double helix conformation is right handed. The two A-A pairs, embedded between the canonical C-G and G-C pairs, are in the *anti* conformation, shifted toward the major groove. One is shifted more than the other in a conformation described as "thumbs-up". This allows the formation of a weak C2H2···N1 H-bond (Figure 2). The mismatch results in a local distortion and unwinding of the helix. However, the interstrand C1′−C1′ distances for the adenosines are only slightly larger than average for the A-form. During 300 ns MD based on this structure, the conformation remains similar to the initial one. The RMSD of the non-hydrogen atoms is 2.1(0.4) Å. The right-handed double helix conformation of RNA_{10}[12a] as well as the conformation of its A_4/A_{14} and A_{17}/A_7 wobble base pairs were preserved (Figure 2) and were in *anti* conformation. The C2−H2(A_4/A_{14})···N1(A_{17}/A_7) H-bond was maintained during the dynamics (Figure 2). The calculated helical parameters of the base pairs and of the base pair step were similar to those of the X-ray structure[12a] (Figure 2 d). The Watson−Crick (C−G and G−C) canonical pairs exhibited more stacking overlaps than those of noncanonical A-A pairing (Figure 2), as found in the X-ray structure. The AG/CA steps showed a large degree of unwinding of the duplex, as manifested by correspondent wider major widths values (Figure 2) similarly to what was observed in the X-ray structure.[12a]

The distribution of counterions around RNA is important for its structure and dynamics.[18,22] It is here analyzed in detail.

The Na^+ counterions were mainly localized in the major groove and specifically close to both CAG triplets (see Figure S3.1 in the SI). The large presence of cations in the major groove has been explained in terms of a rather negative potential in this region for A-form duplexes,[23b,59] including the r(5′-U_1UGGGC(CAG)3GUCC$_{19}$)$_2$ duplex containing three CAG repeats.[22]

The Na^+ ions in the major groove were partially hydrated, and they formed one coordination bond with adenine's N7 atoms ($d_{(Na+-N7)}$ ∼ 2.2 Å) in *anti-anti* orientation.[22] The maximum persistency of these Na^+ ions was 10 ns. All of these features were previously seen in the MD study of the (5′-U_1UGGGC(CAG)3GUCC$_{19}$)$_2$ duplex.[22]

As expected, the Cl^- ions are basically absent around the RNA_{10} (see Figure S3.1 in the SI). Their presence starts beyond 10 Å from the RNA_{10} helical axis and increases from there.

***1**·RNA$_{10}$ and **2**·RNA$_{10}$: Structural Determinants and Ligands' Potency.* An educated guess of **1**·RNA_{10} and **2**·RNA_{10} initial structural model for molecular simulations studies was obtained by using the fully flexible HADDOCK docking code.[49b] The latter takes in consideration both the intrinsic flexibility of RNA (in particular of the mismatched A-A pairs) as well as that of the ligand. We docked **1** and **2** conformers onto RNA_{10} MD structure using HADDOCK.[49a,b] The complexes corresponding to the best HADDOCK score[49a,b] were the starting point for 200 ns MD simulations in explicit solvent (see Method and section S5 in the SI for details). [The docked and MD structures should be taken only as an initial guess for the free energy calculations, and they do not represent a structural prediction.] Next, we performed free WT-meta energy calculations[24,52,60] as a function of the d_{CM}, n_{HB}, and n_{HC} CVs.

In **1**, different minima (**B**, **M1**-**M4**; Figure 3) featured significant intermolecular interactions. The absolute minimum (**B**, d_{CM} ≈ 4.8 Å, n_{HC} ≈ 5, n_{HB} ≈ 3.5) has 100% Boltzmann population. [This minimum differs markedly from the docked and the MD structure of the complex.] It represents the fully bound complex. Here, **1** binds to the r(5′-$\underline{A}_4$GC$\underline{A}_7$)$_2$ tract, into a wedge formed in the major groove by the two mismatched A_4-A_{17} and A_7-A_{14} pairs, intercepting both RNA_{10} strands. The ligand β ring establishes a π-stacking interaction with the A_4 base; the ligand α ring forms a T-shaped π-stacking interaction with A_{14} (see section S4 in the SI). The two positively charged guanidine tails form two salt-bridges with G_5 and G_{15} phosphates, the N23 ligand atom forms an H-bond with the O5′(G_{15}) atom, and the ligand ester carbonyl oxygen forms an H-bond with N4(C_6) (Table 2). The RNA_{10} structure in the complex shares many similarities with that in solution emerging from our MD simulation above (RMSD ∼ 2 Å). Not unexpectedly, the largest differences are found for the A_4-A_{17} and A_7-A_{14} pairs. Indeed, the first pair is more exposed to the major groove than in the free state (as shown by a larger value

Table 2. Salt Bridges and H-Bonds of (a) 1·RNA$_{10}$ and (b) of 2·RNA$_{10}$ Emerging from our WT-meta Calculations[a]

	donor (D)	acceptor (A)	DH···A distance (Å)	DH···A angle (deg)	persistency
(a)	**1**·RNA_{10}				
B					
	N22 (**1**)	OP2 (G_{15})	2.7(0.2)	162(1)	37%
	N23 (**1**)	OP2 (G_{15})	2.9(0.1)	141(3)	30%
	N23 (**1**)	O5′ (G_{15})	3.2(0.2)	139(3)	27%
	N18 (**1**)	OP2 (G_5)	2.8(0.2)	166(1)	29%
	N4(C_6)	(**1**) (O1)C	2.8(0.2)	126(2)	2%
M1					
	N16 (**1**)	OP2 (A_4)	2.9(0.2)	169(1)	28%
	N22 (**1**)	O6 (G_{11})	2.8(0.1)	166(1)	37%
	N23 (**1**)	O6 (G_{12})	2.9(0.1)	172(1)	35% C
	N19 (**1**)	OP2 (C_3)	2.7(0.2)	160(1)	10%
M2					
	N19 (**1**)	OP2 (A_7)	2.9(0.2)	152(1)	35%
	N18 (**1**)	O5′ (C_6)	3.1(0.1)	163(1)	17%
M3					
	N21 (**1**)	O2′ (A_7)	3.0(0.1)	148(1)	36%
	N19 (**1**)	O2 (C_{13})	2.7(0.2)	150(1)	28%
M4					
	N22 (**1**)	OP2 (C_{20})	2.8(0.2)	175(1)	67%
(b)	**2**·RNA_{10}				
B′					
	N4 (**2**)	OP2 (G_5)	2.8(0.2)	160(1)	57%
	N16 (**2**)	N1 (A_4)	2.9(0.1)	155(3)	34%
	N22^w (**2**)	N1^w (A_4)	5.6(0.3)	148(2)	15%
M1′					
	N16 (**2**)	O2′ (A_6)	3.0(0.2)	167(1)	38%
	N26 (**2**)	O2′ (C_{16})	2.9(0.1)	156(1)	17%
M2′					
	N4 (**2**)	OP2 (G_5)	2.9(0.2)	162(1)	41%
	N26 (**2**)	O6 (G_{15})	3.0(0.1)	161 1)	27%
M3′					
	N4 (**2**)	OP2 (G_5)	2.9(0.1)	158(1)	39%
	N26 (**2**)	O6 (G_5)	2.9(0.2)	151(1)	38%
M4′					
	N22 (**2**)	O6 (G_5)	2.7(0.2)	173(1)	37%
	N26 (**2**)	OP2 (G_5)	3.0(0.2)	165(1)	26%

[a]The values (within a bond cut-off of 3.2 Å and an angle cut-off of 120°) are reported for the free-energy minima **B**, **M1**-**M4** in **1**·RNA_{10} (a) and for the free-energy minima **B′**, **M1′**-**M4′** in **2**·RNA_{10} (b).

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

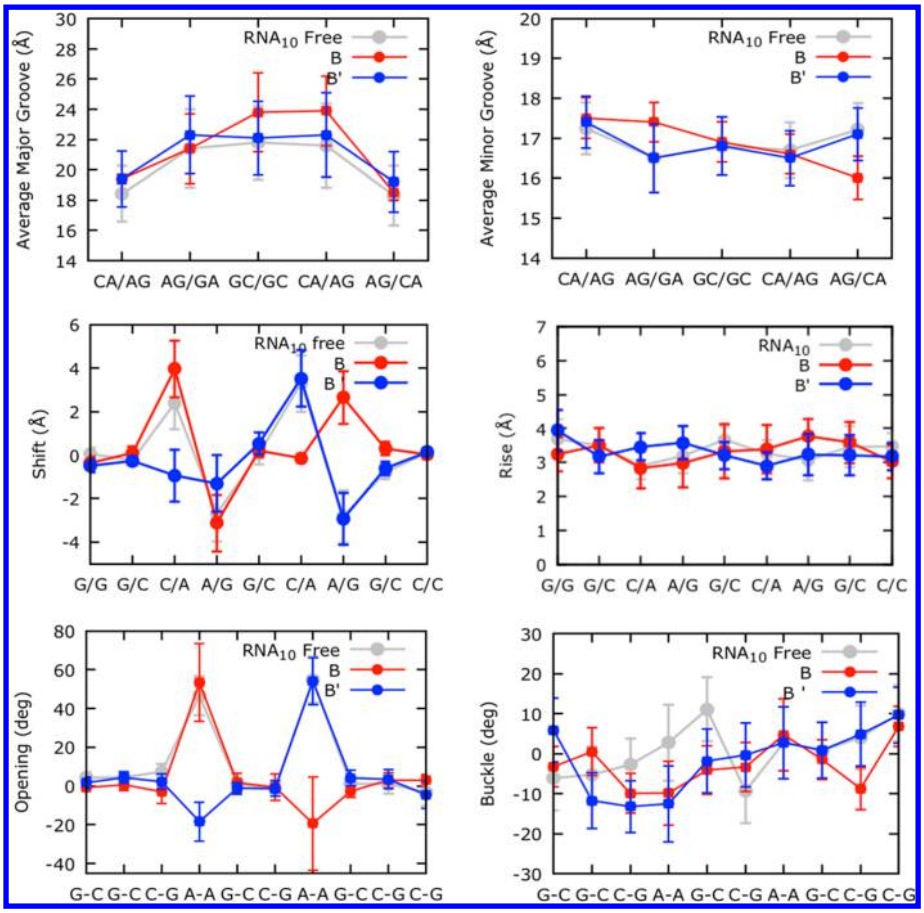


Figure 4. Base pair (A) and base pair step (B) helical parameters, as a function of base pair sequence, for the MD simulation of RNA_{10} in the free state and for the representative structures of the WT-meta free energy minima **B** of the **1**·RNA_{10} complex and **B′** of the **2**·RNA_{10} complex. Numerical values for all the free energy minima (**B**, **M1**-**M4**, **U** for **1**·RNA and **B′**, **M1′**-**M4′**, **U′** for **2**·RNA_{10}) are reported in the Supporting Information (Sections S9 and S10).

of the base pair opening,[61] Figure 4), although they preserve their *anti-anti* conformation as in the free state. A_4 forms a π-stacking interaction with C_3 but not with G_5, as it happens in the free state. However, the base pair step rise value is smaller than those of the other base pair, and the buckle value is negative (Figure 4), unlike what happens in the free state. The N1(A_7)···N6−H6(A_{14}) H-bond replaces the weak N1(A_7)··· C2−H2(A_{14}) H-bond of the free state. As for the first adenine pair, we notice that (i) A_{14} does not form π-stacking interactions with G_{15} as in the free state; (ii) the relative base pair opening of the complex is negative, while that of the free state is positive; (iii) the local winding value of the A_7-A_{14} major groove is larger than that of the free state (Figure 4). The conformation of the ligand remained the same as in water solution here and in all of the other minima in solution (see Table S10.1 in the SI).

The minimum **M1** ($d_{CM} \approx 8$ Å, $n_{HB} \approx 3.5$, $n_{HC} \approx 1.0$) is ≈3.8 kcal/mol higher in energy than **B**. Here, the ligand binds to the RNA_{10} backbone, following the curvature of the first strand (i.e., in direction 5′−3′). On one ligand end, the guanidine N20 atom forms a salt-bridge with A_4's phosphate. On the other ligand end, the N19 and N18 atoms form H-bonds with the O6(G_{11}) and O6(G_{12}) atoms, respectively (Table 2). Conversely from **B**, the **1** ester oxygen does not interact with the RNA_{10}. Instead, it forms H-bonds with the solvent. The axis bending is smaller than that of the free state. The conformations of all adenines except A_4 are similar to those in the free state. A_4 is instead more twisted toward the major groove and more solvent exposed than in the free state (Section S9 in the SI).

M2 ($d_{CM} \approx 4.8$ Å, $n_{HC} \approx 28$, $n_{HB} \approx 1.5$) is ≈4.5 kcal/mol higher in energy than **B**. The molecule exhibits, similarly to **B**, a cross-strand binding mode, intercepting the $(C_6A_7)_2$ tract in the lower major groove. The molecules' aromatic rings form π-stacking interactions with C_6 and A_{14} bases. One ligand guanidine tail forms two salt-bridges with C_6 and A_7 phosphate groups (Table 2). The other guanidine group and the ligand ester oxygen are solvent exposed. The RNA_{10} structure is, here, more bent than in the free state. In particular, the base pair opening of the A_4-A_{17} base pair is larger than that of the free state one. The A_{14} base is twisted toward the major groove (Section S9 in the SI).

Minima **M3**-**M4** are ≈9−10 kcal/mol higher in energy than **B**. In **M3** ($d_{CM} \approx 11.3$ Å, $n_{HC} \approx 28$, $n_{HB} \approx 2.5$; Figure 4) A_{14} is completely flipped out from the helical stack. It establishes a π-stacking interaction with the ligand's β ring. The α aromatic ring forms a parallel-distorted π-stacking with the sugar group of G_8 (see section S7 in the SI). On average two H-bonds involving the charged guanidine tails and O2′(A_7) and O2(C_{13}) atoms, respectively, are observed (Table 2). Although the nucleic acids undergo significant structural changes, the distortions remain quite localized to the binding region, and the duplex preserve the A-RNA form, without featuring overall bending with respect to the free state (see Section S9 in the SI).

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

The flipping of the A_{14} makes the lower major groove (r(5′-$C_6\underline{A}_7G_8C_9C_{10})_2$ tract) narrower compared to the free state, with a lower average P−P distance in that tract (see Section S9 in the SI). In **M4** ($d_{CM} \approx 18$ Å, $n_{HC} \approx 0.5$ $n_{HB} \approx 2$), the ligand is bound to the 5′-end of the RNA sequence, loosing the interactions with both the minor and the major groove floors. Its guanidine ligand tail forms a salt-bridge with C_{20} phosphate (Table 2). The RNA_{10} conformation is similar to that of the free state (see Section S9 in the SI).

In the minimum **U** ($d_{CM} \approx 29.8$ Å, $n_{HC} \approx 0$, $n_{HB} \approx 0$), about ≈10 kcal/mol higher in free energy than **B**, direct intermolecular **1**·RNA_{10} interactions are absent. The A-A pairs adopt the peculiar configurations reported for the X-ray structure;[12a] the structure shares the same local deformations from the standard RNA A-type structure reported in ref 12a.

The free energy of unbinding from the WT-meta calculations ($\Delta G^{WT\text{-}meta}$) from **B** to **U** is therefore ≈10 kcal/mol. The residual unbinding free energy on passing from minimum **U** to a complete separation of the two molecules is expected to be significant because of the long-range electrostatic interactions between these two highly charged molecules. This free energy was estimated by using the nonlinear Poisson−Boltzmann equation (ΔG^{PB}) on the structure of the minimum **U**: $\Delta G^{PB} \approx$ 2.4(0.7) kcal/mol. Therefore, the total unbinding free energy reads $\Delta G^{TOT} = \Delta G^{WT\text{-}meta} + \Delta G^{PB} \approx 12$ kcal/mol. The standard state free energy ($\Delta G^0_{(calc)}$), which takes into account the concentration of the ligand in our simulation, gives a value of $\Delta G^0_{(calc)} \approx 10$ kcal/mol. This is slightly larger than the experimental value of 9.9(0.7) kcal/mol.[12b] The small discrepancy between our evaluations of the binding affinity should be ascribed, in the first instance, to the differences in RNA structure:[12b] the experiments were performed on a short hairpin containing a single CAG internal loop.

The calculated enthalpy (ΔH = 28(1) kcal/mol) and entropy ($-T\Delta S \approx -18$ kcal/mol) changes between **B** and **U** govern the binding process. The increase in enthalpy favors the bound state, while the conformational entropy favors the unbound state. However, the absolute value of the entropic contribution is smaller than that of the enthalpy, resulting in a net decrease in free energy upon binding. We expect that the relatively small contribution from the residual unbinding free energy does not change this picture. Hence, overall, our calculations suggest that the **1**·RNA_{10} complex formation is enthalpy driven, as seen in other ligands/nucleic acids complexes.[62]

The free energy landscape associated with the unbinding of **2**·RNA_{10} in water is characterized by several minima, similarly to what is observed in **1**·RNA_{10}.

The lowest minimum (**B′**, at d_{CM} = 8.5 Å, n_{HC} = 4, n_{HB} = 2, see Figure 3) is the bound state. **2** binds to the r(5′-$\underline{A}_4G_5)_2$ tract in the upper major groove. The ligand δ ring forms a distorted π-stacking interaction with the A_4 base. The ligand's dihydroimidazole on the γ moiety forms van der Waals interactions with the C_{16} base. The ligand's N4 atom forms a salt bridge with G_5's phosphate group, while N16 and N22 form a direct and a water mediated H-bond with the N1 in A_4, respectively (Table 2). The ligand aromatic rings are more rotated with respect to each other than in the free state (Figure 1): the angle ω passes from 28°(3°) in the free state to 48° (3°) (Table S11.1 in the SI).

The RNA_{10} structure is affected locally by ligand binding. Indeed, its major and minor groove widths are not dissimilar from those of free RNA_{10} (Figure 4). However, the A_4-A_{17} pair, while preserving the *anti-anti* orientation, features a negative opening value and a smaller buckle than the free state. In addition, the N3(A_4)···N6−H6(A_{17}) H-bond replaces the N1(A_7)···C2−H2(A_{14}) one present in the free state.

M1′ ($d_{CM} \approx 12$ Å, $n_{HC} \approx 5$, $n_{HB} \approx 1$; Figure 3) is ≈3.5 kcal/mol higher than **B′**. Here, the ligand binds to the r(5′-$\underline{A}_4G_5C_6\underline{A}_7G_8$) tract in the minor groove. The ligand's N16 and N26 atoms (Figure 1) form H-bonds with O2′(C_6) and O2′(C_{16}) atoms, respectively (Table 2). The RNA structure is rather similar to that of the free state (see section S10 in the SI). The ligand aromatic rings are here almost coplanar (ω is ≈9°(2°), see Table S11.1 in the SI).

M2′ ($d_{CM} \approx 4.5$ Å, $n_{HC} \approx 40$, $n_{HB} \approx 3$; Figure 3) is ≈4.7 kcal/mol higher in energy by with respect to **B′**. Here, the ligand binds to the r(5′-$A_4G_5C_6A_7$) tract in the major groove (Figure 3). The ligand's γ and δ rings (Figure 1) establish parallel-distorted π-stacking interactions with the A_{14} and A_4 rings, respectively. One of the two dihydroimidazole moieties (Figure 1) forms a salt bridge with the G_{15} phosphate group. The other one establishes a salt bridge with the G_5 phosphate group (Table 2). The ligand N26 atom forms an H-bond with O6(G_{15}) (Table 2). The overall bending of the RNA_{10} structure is similar to the one of the free state with the exception of the r(5′-$\underline{A}_{14}G_{15}$) tract. The latter is slightly more bent than the same tract in the free state (see section S10 in the SI).

M3′ ($d_{CM} \approx 3$ Å, $n_{HC} \approx 20$, $n_{HB} \approx 1.5$; Figure 3) is ≈5.4 kcal/mol higher in free energy with respect to **B′**. Here the ligand δ ring establishes a parallel-distorted π-stacking with A_4. The ligand's dihydroimidazole moiety on δ ring forms van der Waals interaction with the A_{14} base. The ligand's N4 nitrogen atom forms a salt-bridge with the G5′s phosphate (Table 2). The N26 nitrogen atom forms an H-bond with O6(G_5) atom (Table 2). The A_4 and A_{14} bases are both more twisted toward the major groove than in the free state with corresponding base-pair opening values greater than in the free state (section S10 in the SI). The ligand structure is unperturbed with respect to its conformation in water solution (Section S11 in the SI).

M4′ ($d_{CM} \approx 5.5$ Å, $n_{HC} \approx 65$, $n_{HB} \approx 2$; Figure 3) is almost isoenergetic with **M3′**. In the major groove, the γ ring forms a π-stacking with the A_4 base. The ligand's δ ring is completely solvent exposed. The ligand's N22 and N26 atoms (Figure 1) form H-bonds with O6(G_5) and the G_5 phosphate group, respectively (Table 2). The A_4 is more bent than in the free state and does not form π-stacking interactions with the upper C_3 and lower G_5 bases (see Section 10 in the SI). The ligand conformation is the same as that in water solution (see Section S11 in the SI).

U′ corresponds to an unbound state ($d_{CM} \approx 23$ Å, n_{HC} = 0, n_{HB} = 0; Figure 3): direct interactions between **2** and RNA_{10} are absent. It is about 8 kcal/mol higher than **B′**. The RNA structure is here similar to that of the free state, sharing similar overall bending, base pair and base pair-step values (see Section 10 in the SI).

The WT-meta free energy of unbinding of **2**, $\Delta G^{WT\text{-}meta}$, from **B′** to **U′** is ≈8 kcal/mol. The estimate of the residual unbinding free energy on passing from minimum **U′** to a state in which the two moieties are completely separated, calculated as for **1**, is $\Delta G^{PB} \approx 2.6(0.5)$ kcal/mol, with a resulting total unbinding free energy $\Delta G^{TOT} = \Delta G^{WT\text{-}meta} + \Delta G^{PB} \approx 10$ kcal/mol. Taking into account the concentration of the ligand in the simulation box, the standard state free energy is $\Delta G^0_{(calc)} \approx 8$ kcal/mol. As in the case of the **1**·RNA_{10} complex, the calculated $\Delta G^0_{(calc)}$ value is in fair agreement with the experimental one (differences between 0.4 kcal/mol or smaller see Table 1).

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

As for the **1**·RNA_{10} complex, the estimated enthalpy (ΔH = 23(1) kcal/mol) and entropy ($-T\Delta S \approx -13$ kcal/mol) changes between **B′** and **U′** lead us to suggest that the binding event is enthalpy driven.

Accuracy of the Free Energy Predictions. As in any molecular modeling work, several limitations may affect the prediction of energetics and binding poses. First, the choice of the collective variables (CVs), while already successful in studying biomolecular recognition process,[53b,60,63] take into account the most important interactions at short distances, becoming void and ineffective at long distances. This is even more important in the effort to describe the binding/unbinding process of two highly charged molecules like RNA_{10} and the charged ligands investigated in which long-range electrostatic interactions play a fundamental role.[23b,64] Another factor affecting the accuracy of our predictions is, for instance, the free energy penalty due to the loss of conformational entropy upon binding. However, this effect may be limited by the fact that the bound conformations of the ligands are close to their lowest energy conformation.

Moreover, nonpolarizable force fields such as AMBER,[17,19,36a] although rather accurate in the current version,[44b] may show difficulties in characterizing RNA noncanonical structural elements such as A-A mismatches[21b,44b,64,65] and in describing electrostatic interactions[20b,64,66] between the RNA and the anions as well as the one between the highly polarizable phosphodiester moiety of RNA and the positively charged atoms including the cations and those of the ligands in our case. The calculated affinities and poses would benefit from the possibility to take explicitly into account possible (yet not likely) changes in the charge state of the RNA and of the ligand upon binding.[67] These changes are directly related to binding-induced changes in pK_a values of ionizable groups in the receptor and in the ligands. These could be properly predicted by accurate but computational expensive quantum mechanics/molecular mechanics (QM/MM) simulations.[68] This is a very general approach for pK_a calculations of molecules in water and in complex with biomolecules.[69]

In addition, the calculated overall free energy, ΔG^{TOT}, contains also a significant source of inaccuracy from the highly approximate nonlinear Poisson−Boltzmann calculations as pointed out, for instance, in ref 70.

CONCLUSION

We have used well-tempered metadynamics to uncover the binding mode of the best characterized ligand targeting CAG trinucleotide repeats (Figure 1): **1**, the most potent ligand identified so far,[12b] and **2**, which binds to CAG repeats with less affinity. The calculated affinity to our model system of CAG repeats, RNA_{10} (Figure 2), is similar to that experimentally determined. The RNA_{10} structure in the complexes was not considerably altered by the ligands binding, despite distortions affecting locally the A-A base pairs, which indeed play the most important role in the molecular recognition process.

1 features a dominant binding mode in which it intercepts both RNA_{10} strands, binding in the major groove of the r(5′-$\underline{A_4}G_5C_6\underline{A_7}$)$_2$ region into a wedge formed by the two A-A pairs. It features five main "anchoring" interactions to the RNA_{10}, the salt-bridges between the charged guanidine tails and the G_{15} and G_5 phosphates, the H-bond of the N23 nitrogen with the O5′(G_5) atom, as well as the two π-stacking interactions, one parallel (β ring) one T-shaped (α ring) with the A_4 and A_{14} bases, respectively. On the other hand, ligand **2** binds in the upper major groove to the A_4-A_{17} pair only. It forms a salt-bridge with the G_5 phosphate group, H-bonds with the N1(A_4) atom, π-stacking of its δ ring with A_4, and a hydrophobic interaction with the C_6. Hence, **1** is able to establish more molecular interactions with RNA_{10} than **2**. This is likely to be a key factor affecting the binding event, which is suggested here to be enthalpy-driven.

These predictions, validated against comparison with experiment, could help the design of ligands highly specific for CAG-repeat containing RNA, able to interfere with pathological mRNA−protein complexes in cells.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00208.

Section S1, ligands conformations; Section S2, MD simulation parameters; Section S3, NaCl ions distribution along the 300 ns-long MD simulation of RNA_{10}; Section S4, some considerations on the ions' force field used; Section S5, molecular docking; Section S6, free energy calculations; Section S7, π-stacking interactions geometries in the binding poses of **1**; Section S8, comparison between the HADDOCK docked binding pose and the well-tempered metadynamics binding pose; Section S9, major, minor, base-pair, and base-pair steps parameters for RNA_{10} upon **1** binding; Section S10, major, minor, base-pair, and base-pair steps parameters for RNA_{10} upon **2** binding; Section S11, ligands conformations upon binding (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: g.rossetti@fz-juelich.de.

Author Contributions

All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JURECA at Jülich Supercomputing Centre (JSC).

REFERENCES

(1) Huntington, G. On Chorea. *Medical and Surgical Reporter: A Weekly Journal* **1872**, *26*, 5.

(2) Ross, C. A.; Tabrizi, S. J. Huntington's disease: from molecular pathogenesis to clinical treatment. *Lancet Neurol.* **2011**, *10* (1), 83.

(3) Walker, F. O. Huntington's disease. *Lancet* **2007**, *369* (9557), 218.

(4) MacDonald, M.; Ambrose, C.; Duyao, M.; Myers, R.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.; James, M.; Groot, N.; MacFarlane, H.; Jenkins, B.; Anderson, M.; Wexler, N.; Gusella, J.; Bates, G.; Baxendale, S.; Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott, R.; Zehetner, G.; Sedlacek, Z.; Poustka, A.; Frischauf, A.-M.; Lehrach, H.; Buckler, A.; Church, D.; Doucette-Stamm, L.; O'Donovan, M.; Riba-Ramirez, L.; Shah, M.; Stanton, V.; Strobel, S.; Draths, K.; Wales, J.; Dervan, P.; Housman, D.; Altherr, M.; Shiang, R.; Thompson, L.; Fielder, T.; Wasmuth, J.; Tagle, D.; Valdes, J.; Elmer, L.; Allard, M.; Castilla, L.; Swaroop, M.; Blanchard, K.; Collins, F.;

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

Snell, R.; Holloway, T.; Gillespie, K.; Datson, N.; Shaw, D.; Harper, P. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **1993**, *72* (6), 971.

(5) Reiner, A.; Dragatsis, I.; Zeitlin, S.; Goldowitz, D. Wild-type huntingtin plays a role in brain development and neuronal survival. *Mol. Neurobiol.* **2003**, *28* (3), 259.

(6) (a) Ross, C. A. When more is less: pathogenesis of glutamine repeat neurodegenerative diseases. *Neuron* **1995**, *15* (3), 493. (b) Gusella, J. F.; MacDonald, M. E. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nat. Rev. Neurosci.* **2000**, *1* (2), 109. (c) Rosenblatt, A.; Liang, K. Y.; Zhou, H.; Abbott, M. H.; Gourley, L. M.; Margolis, R. L.; Brandt, J.; Ross, C. A. The association of CAG repeat length with clinical progression in Huntington disease. *Neurology* **2006**, *66* (7), 1016.

(7) Myers, R. Huntington's Disease Genetics. *NeuroRx* **2004**, *1* (2), 255.

(8) de Mezer, M.; Wojciechowska, M.; Napierala, M.; Sobczak, K.; Krzyzosiak, W. J. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res.* **2011**, *39* (9), 3852.

(9) Housman, D. Gain of glutamines, gain of function? Nat. *Nat. Genet.* **1995**, *10* (1), 3.

(10) Rudnicki, D. D.; Margolis, R. L.; Pearson, C. E.; Krzyzosiak, W. J. Diced triplets expose neurons to RISC. *PLoS Genet.* **2012**, *8* (2), e1002545.

(11) (a) Wojciechowska, M.; Krzyzosiak, W. J. CAG repeat RNA as an auxiliary toxic agent in polyglutamine disorders. *RNA Biol.* **2011**, *8* (4), 565. (b) Nalavade, R.; Griesche, N.; Ryan, D. P.; Hildebrand, S.; Krauss, S. Mechanisms of RNA-induced toxicity in CAG repeat disorders. *Cell Death Dis.* **2013**, *4* (8), e752.

(12) (a) Kiliszek, A.; Kierzek, R.; Krzyzosiak, W. J.; Rypniewski, W. Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res.* **2010**, *38* (22), 8370. (b) Kumar, A.; Parkesh, R.; Sznajder, L. J.; Childs-Disney, J. L.; Sobczak, K.; Disney, M. D. Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG)-containing transcripts. *ACS Chem. Biol.* **2012**, *7* (3), 496. (c) Ranum, L. P. W.; Cooper, T. A. RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci.* **2006**, *29*, 259. (d) Fiszer, A.; Krzyzosiak, W. J. RNA toxicity in polyglutamine disorders: concepts, models, and progress of research. *J. Mol. Med.* **2013**, *91* (6), 683. (e) Krauss, S.; Griesche, N.; Jastrzebska, E.; Chen, C.; Rutschow, D.; Achmuller, C.; Dorn, S.; Boesch, S. M.; Lalowski, M.; Wanker, E.; Schneider, R.; Schweiger, S. Translation of HTT mRNA with expanded CAG repeats is regulated by the MID1-PP2A protein complex. *Nat. Commun.* **2013**, *4*, 1511.

(13) Fiszer, A.; Krzyzosiak, W. J. Oligonucleotide-based strategies to combat polyglutamine diseases. *Nucleic Acids Res.* **2014**, *42* (11), 6787.

(14) (a) Rzuczek, S. G.; Park, H.; Disney, M. D. A Toxic RNA Catalyzes the In Cellulo Synthesis of Its Own Inhibitor. *Angew. Chem., Int. Ed.* **2014**, *53* (41), 10956−10959. (b) Childs-Disney, J. L.; Yildirim, I.; Park, H.; Lohman, J. R.; Guan, L.; Tran, T.; Sarkar, P.; Schatz, G. C.; Disney, M. D. Structure of the Myotonic Dystrophy Type 2 RNA and Designed Small Molecules That Reduce Toxicity. *ACS Chem. Biol.* **2014**, *9* (2), 538. (c) Childs-Disney, J. L.; Stepniak-Konieczna, E.; Tran, T.; Yildirim, I.; Park, H.; Chen, C. Z.; Hoskins, J.; Southall, N.; Marugan, J. J.; Patnaik, S.; Zheng, W.; Austin, C. P.; Schatz, G. C.; Sobczak, K.; Thornton, C. A.; Disney, M. D. Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. *Nat. Commun.* **2013**, *4*, 2044.

(15) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−5197 .

(16) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. Reparameterization of RNA χ Torsion Parameters for the AMBER Force Field and Comparison to NMR Spectra for Cytidine and Uridine. *J. Chem. Theory Comput.* **2010**, *6* (5), 1520−1531.

(17) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92* (11), 3817.

(18) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813.

(19) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886.

(20) (a) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* **2009**, *15* (6), 1219. (b) Fulle, S.; Gohlke, H. Molecular recognition of RNA: challenges for modelling interactions and plasticity. *J. Mol. Recognit.* **2009**, *23* (2), 220. (c) Daldrop, P.; Reyes, F. E.; Robinson, D. A.; Hammond, C. M.; Lilley, D. M.; Batey, R. T.; Brenk, R. Novel ligands for a purine riboswitch discovered by RNA-ligand docking. *Chem. Biol.* **2011**, *18* (3), 324. (d) Chen, L.; Calin, G. A.; Zhang, S. Novel Insights of Structure-Based Modeling for RNA-Targeted Drug Discovery. *J. Chem. Inf. Model.* **2012**, *52* (10), 2741.

(21) (a) Musiani, F.; Rossetti, G.; Capece, L.; Gerger, T. M.; Micheletti, C.; Varani, G.; Carloni, P. MD simulations identify time scale of conformational changes responsible for conformational selection in molecular recognition of HIV-1 TAR RNA. *J. Am. Chem. Soc.* **2014**, *136* (44), 15631. (b) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Chetam, T.; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and CNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836. (c) Faustino, I.; Perez, A.; Orozco, M. Toward a consensus view of duplex RNA flexibility. *Biophys. J.* **2010**, *99* (6), 1876. (d) Deng, N. J.; Cieplak, P. Free energy profile of RNA hairpins: a molecular dynamics simulation study. *Biophys. J.* **2010**, *98* (4), 627. (e) Banáš, P.; Sklenovský, P.; Wedekind, J. E.; Šponer, J.; Otyepka, M. Molecular Mechanism of preQ1 Riboswitch Action: A Molecular Dynamics Study. *J. Phys. Chem. B* **2012**, *116*, 12721−12734.

(22) Yildirim, I.; Park, H.; Disney, M. D.; Schatz, G. C. A Dynamic Structural Model of Expanded RNA CAG Repeats: A refined X-ray Structure and Computational Onvestigations Using Molecular Dynamics and Umbrella Sampling Simulations. *J. Am. Chem. Soc.* **2013**, *135*, 3528.

(23) (a) Fulle, S.; Christ, N. A.; Kestner, E.; Gohlke, H. HIV-1 TAR RNA spontaneously undergoes relevant apo-to-holo conformational transitions in molecular dynamics and constrained geometrical simulations. *J. Chem. Inf. Model.* **2010**, *50* (8), 1489. (b) Do, T. N.; 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* (2), 688.

(24) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.

(25) Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. Exploring Complex Protein−Ligand Recognition Mechanisms with Coarse Metadynamics. *J. Phys. Chem. B* **2009**, *113* (14), 4807.

(26) Sobczak, K.; Krzyzosiak, W. J. CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J. Biol. Chem.* **2005**, *280* (5), 3898.

(27) Schaftenaar, G.; Noordik, J. H. Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 123−134.

(28) *Marvin 5.2*; ChemAxon: Budapest, 2009.

(29) (a) Tsuji, T.; Takeuchi, H.; Egawa, T.; Konaka, S. Effects of molecular structure on the stability of a thermotropic liquid crystal. Gas electron diffraction study of the molecular structure of phenyl benzoate. *J. Am. Chem. Soc.* **2001**, *123*, 6381. (b) Kon, M.; Kurokawa,

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

H.; Takeuchi, H.; Konaka, S. Conformational studies by liquid crystal NMR and ab initio calculations: methyl nicotinate and methyl isonicotinate. *J. Mol. Struct.* **1992**, *268* (1), 155. (c) Kiyono, H.; Tatsunami, R.; Kurai, T.; Takeuchi, H.; Egawa, T.; Konaka, S. Structure determination of methyl nicotinate and methyl picolinate by gas electron diffraction combined with ab initio calculations. *J. Phys. Chem. A* **1998**, *102* (8), 1405. (d) Wrzalik, R.; Merkel, K.; Kocot, A. Ab initio study of phenyl benzoate: structure, conformational analysis, dipole moment, IR and Raman vibrational spectra. *J. Mol. Model.* **2003**, *9*, 248−258.

(30) Sun, H. Force field for computation of conformational energies, structures, and vibrational frequencies of aromatic polyesters. *J. Comput. Chem.* **1994**, *15* (7), 752−768.

(31) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3094.

(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. M.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachar, I. K.; Rendell, A.; Buran, T. J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; 2009.

(33) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29* (7), 845.

(34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157.

(35) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.* **1993**, *97* (40), 10269.

(36) (a) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; 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. *AMBER 12*; University of California: San Francisco, 2012. (b) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668.

(37) 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* (2), 926.

(38) (a) Smith, D. E.; Dang, L. X. Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.* **1994**, *100* (5), 3757. (b) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1−3), 43.

(39) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089.

(40) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. M. E. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463−1472.

(41) (a) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31* (3), 1695. (b) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52* (2), 255.

(42) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52* (12), 7182.

(43) Kiliszek, A.; Rypniewski, W. Structural studies of CNG repeats. *Nucleic Acids Res.* **2014**, *42* (13), 8189.

(44) (a) Ode, H.; Matsuo, Y.; Neya, S.; Hoshino, T. Force field parameters for rotation around χ torsion axis in nucleic acids. *J. Comput. Chem.* **2008**, *29* (15), 2531−2542. (b) Yildirim, I. S.; Tubbs, H.; Kennedy, J.; Turner, S.; Benchamrking, D. 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**, *6*, 9261−9270. (c) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E. I.; Šponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6* (12), 3836.

(45) Torda, A. E.; van Gunsteren, W. F. Algorithms for Clustering Molecular Dynamics Configurations. *J. Comput. Chem.* **1994**, *15*, 1331.

(46) (a) Dans, P. D.; Faustino, I.; Battistini, F.; Zakrzewska, K.; Lavery, R.; Orozco, M. Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res.* **2014**, *42* (18), 11304. (b) Pasi, M.; Maddocks, J. H.; Lavery, R. Analyzing ion distributions around DNA: sequence-dependence of potassium ion distributions from microsecond molecular dynamics. *Nucleic Acids Res.* **2015**, *43* (4), 2412−2423.

(47) Auffinger, P.; Cheatam, I. T. E.; Vaiana, A. C. Spontaneous formation of KCl aggregates in biomolecular simulations: a force field issue? *J. Chem. Theory Comput.* **2007**, *3* (5), 1851−1859.

(48) Joung, I. S.; Cheatham, T. E. d. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020−9041.

(49) (a) van Dijk, M.; van Dijk, A. D.; Hsu, V.; Boelens, R.; Bonvin, A. M. Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res.* **2006**, *34* (11), 3317−3325. (b) Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125* (7), 1731. (c) de Vries, S. J.; van Dijk, A. D.; Krzeminski, M.; van Dijk, M.; Thureau, A.; Hsu, V.; Wassenaar, T.; Bonvin, A. M. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins: Struct., Funct., Genet.* **2007**, *69* (4), 726.

(50) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide folding: When simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38* (1−2), 236−240.

(51) Schuttelkopf, A. W.; van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60* (8), 1355−1363.

(52) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562.

(53) (a) Kranjc, A.; Bongarzone, S.; Rossetti, G.; Biarnes, X.; Cavalli, A.; Bolognesi, M. L.; Roberti, M.; Legname, G.; Carloni, P. Docking Ligands on Protein Surfaces: The Case Study of Prion Protein. *J. Chem. Theory Comput.* **2009**, *5* (9), 2565. (b) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. Dissociation of minor groove binders from DNA: insights from metadynamics simulations. *Nucleic Acids Res.* **2008**, *36* (18), 5910. (c) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. Sliding of alkylating anticancer drugs along the minor groove of DNA: new insights on sequence selectivity. *Biophys. J.* **2008**, *94* (2), 550. (d) Vargiu, A. V.; Ruggerone, P.; Magistrate, A.; Carloni, P. Anthramycin-DNA binding explored by molecular simulations. *J. Phys. Chem. B* **2006**, *110* (48), 24687−24695. (e) Vargiu, A. V.; Magistrato, A. Detecting DNA mismatches with

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922

metallo-insertors: a molecular simulation study. *Inorg. Chem.* **2012**, *51* (4), 2046.
(54) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, *127* (8), 2600.
(55) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180* (10), 1961.
(56) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (18), 10037.
(57) Lavery, R.; Maddocks, J. H.; Pasi, M.; Zakrzewska, K. Analyzing ion distributions around DNA. *Nucleic Acids Res.* **2014**, *42* (12), 8138.
(58) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.* **2009**, *37*, 5917.
(59) (a) Lavery, R.; Pullmann, B. The electrostatic filed of DNA: the role of the nucleic acid conformation. *Nucleic Acids Res.* **1982**, *10* (14), 4383. (b) Pullmann, B. Electrostatics of polymorphic DNA. *J. Biomol. Struct. Dynam.* **1983**, *1* (3), 773−794. (c) Corbin, S.; Lavery, R.; Pullmann, B. The molecular Electrostatic Potential and Steric Accessibility of Double-helical A-RNA. *Biochim. Biophys. Acta, Gene Struct. Expression* **1982**, *698* (1), 86−92.
(60) Zheng, W.; Vargiu, A. V.; Rohrdanz, M. A.; Carloni, P.; Clementi, C. Molecular recognition of DNA by ligands: roughness and complexity of the free energy profile. *J. Chem. Phys.* **2013**, *139* (14), 145102.
(61) Trotta, E.; Paci, M. Solution structure of DAPI selectively bound in the minor groove of a DNA TT mismatch-containing site: NMR and molecular dynamics studies. *Nucleic Acids Res.* **1998**, *26*, 4706.
(62) (a) Joynt, S.; Morillo, V.; Leng, F. Binding the mammalian high mobility group protein AT-hook 2 to AT-rich deoxyoligonucleotides: enthalpy-entropy compensation. *Biophys. J.* **2009**, *96* (10), 4144. (b) Cui, T.; Wei, S.; Brew, K.; Leng, F. Energetics of binding the mammalian high mobility group protein HMGA2 to poly(dA-dT)2 and poly(dA)-poly(dT). *J. Mol. Biol.* **2005**, *352* (3), 629.
(63) Nguyen, T. H.; Rossetti, G.; Arnesano, F.; Ippoliti, E.; Natile, G.; Carloni, P. Molcular Recognition of Platinated DNA from Chromosomal HMGB1. *J. Chem. Theory Comput.* **2014**, *10*, 3578.
(64) Do, T. N.; Carloni, P.; Varani, G.; Bussi, G. RNA/Peptide Binding Driven by Electrostatics-Insight from Bidirectional Pulling Simulations. *J. Chem. Theory Comput.* **2013**, *9* (3), 1720.
(65) (a) Orozco, M.; Noy, N.; Perez, A. Recent advances in the study of nucleic acids flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185. (b) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Jr. 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* (9), 1929.
(66) (a) McDowell, S.; Spackova, N.; Sponer, J.; Walter, N. Molecular Dynamics simulations of RNA: An In Silico Single Moleculae Approach. *Biopolymers* **2007**, *85*, 169. (b) Reynolds, C. A.; Richards, W. G.; Essex, J. W. Atomic charges for variable molecular conformations. *J. Am. Chem. Soc.* **1992**, *114*, 9075−9079. (c) Cieplak, P.; Cornell, W.; Bayly, C.; Kollmann, P. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation fro DNA, RNA and proteins. *J. Comput. Chem.* **1995**, *16* (11), 1357−1377.
(67) (a) Sackett, K.; TerBush, A.; Weliky, D. P. HIV gp41 six-helix bundle constructs induce rapid vesicle fusion at pH 3.5 and little fusion at pH 7.0: understanding pH dependence of protein aggregation, membrane binding, and electrostatics, and implications for HIV-host cell fusion. *Eur. Biophys. J.* **2011**, *40* (4), 489. (b) Onufriev, A. V.; Alexov, E. Protonation and pK changes in protein-ligand binding. *Q. Rev. Biophys.* **2013**, *46* (2), 181.
(68) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium - ion in reaction of lysozyme. *J. Mol. Biol.* **1976**, *103* (2), 227.
(69) (a) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26* (9), 915. (b) Mangold, M.; Rolland, L.; Costanzo, F.; Sprik, M.; Sulpizi, M.; Blumberger, J. Absolute pKa Values and Solvation Structure of Amino Acids from Density Functional Based Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2011**, *7* (6), 1951−1961. (c) Li, G.; Cui, Q. pKa calculations with QM/MM free energy perturbations. *J. Phys. Chem. B* **2003**, *107* (51), 14521−14528. (d) Riccardi, D.; Schaefer, P.; Cui, Q. pKa calculations in solution and proteins with QM/MM free energy perturbation simulations: a quantitative test of QM/MM protocols. *J. Phys. Chem. B* **2005**, *109* (37), 17715−17733. (e) Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. The prediction of protein pKa's using QM/MM: the pKa of lysine 55 in turkey ovomucoid third domain. *J. Phys. Chem. B* **2002**, *106* (13), 3486.
(70) Kirmizialtin, S.; Silalahi, A. R.; Elber, R.; Fenley, M. O. The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations. *Biophys. J.* **2012**, *102* (4), 829.

DOI: 10.1021/acs.jctc.5b00208
J. Chem. Theory Comput. 2015, 11, 4911−4922