

In silico direct folding of thrombin-binding aptamer G-quadruplex at all-atom level

Changwon Yang[†], Mandar Kulkarni[†], Manho Lim and Youngshang Pak^{*}

Department of Chemistry and Institute of Functional Materials, Pusan National University, Busan 609-735, South Korea

Received August 30, 2017; Revised October 13, 2017; Editorial Decision October 16, 2017; Accepted October 20, 2017

ABSTRACT

The reversible folding of the thrombin-binding DNA aptamer G-quadruplexes (GQs) (TBA-15) starting from fully unfolded states was demonstrated using a prolonged time scale (10–12 μ s) parallel tempering metadynamics (PTMetaD) simulation method in conjunction with a modified version of the AMBER bsc1 force field. For unbiased descriptions of the folding free energy landscape of TBA-15, this force field was minimally modified. From this direct folding simulation using the modified bsc1 force field, reasonably converged free energy landscapes were obtained in K^+ -rich aqueous solution (150 mM), providing detailed atomistic pictures of GQ folding mechanisms for TBA-15. This study found that the TBA folding occurred via multiple folding pathways with two major free energy barriers of 13 and 15 kcal/mol in the presence of several intermediate states of G-triplex variants. The early formation of these intermediates was associated with a single K^+ ion capturing. Interestingly, these intermediate states appear to undergo facile transitions among themselves through relatively small energy barriers.

INTRODUCTION

DNA G-quadruplexes (GQs), which are found in human telomeres, appear to play a vital role in the genome stability and cellular processes. In the presence of metal cations with appropriate sizes, GQ structures are formed by the non-canonical base pairing of guanine-rich DNA sequences. Therefore, just like protein folding, it is fundamentally important to understand how certain guanine-rich sequences spontaneously fold into GQs. Molecular dynamics (MD) simulations offer a powerful computational tool for tackling such GQ folding problems. In particular, MD enables the computation of folding free energy landscape at atomistic resolutions. This free energy landscape can serve as a

theoretical basis to assess the detailed folding mechanisms and structural and kinetic aspects of folding. As indicated in previous studies (1–9), however, GQ folding is a slow (order of seconds to hours) and very complex process. Therefore, for this folding event, free energy simulations using all-atom MD sampling are computationally challenging. Until now, using normal or enhanced all-atom MD schemes, several interesting results have been reported for GQ folding (10–18). Despite these computational efforts, two issues of force field imbalances and limited MD sampling hinder the construction of reliable GQ folding free energy landscapes. In the present work, properly addressing these issues, we endeavored to compute folding free energy landscapes of a thrombin-binding aptamer (TBA-15), which is a simplest possible GQ model system. The TBA-15, a 15-mer oligonucleotide with a sequence of 5'-GGTTGGTGTGGTTGG-3', is an anti-coagulant and also has been used as a potential target for developing aptamer-based drugs (19–23).

Earlier nuclear magnetic resonance (NMR) studies (24,25) reported that TBA-15 displayed a two-planar tetrad GQ (GQ_2) with cations captured in its central binding site or around the TGT loop region. Despite this simple GQ structure, the folding time of TBA-15 (order of seconds) (1,2) poses a severe computational difficulty with fully atomistic simulations. Notably, in a previous MD study (11) with the AMBER bsc0 force field (26), a free energy landscape of TBA folding was computed using a well-tempered metadynamics (WTMetaD) (27). From this free energy surface, interestingly a two-triad G-triplex structure (GT_2) (11,28) was located as a favorable intermediate of TBA-15 folding and supported by an NMR experiment on a 3'-terminal-truncated TBA (TBA-11). Nevertheless, this simulation run was too short (80 ns) to ensure sufficient sampling for unbiased free energy calculations. (Also, it is of note that the free energy surface from an earlier replica exchange MD study (10) was not converged.)

Enhanced MD sampling protocols and accurate all-atom force fields are essential for producing a reliable folding free energy profile of the GQ model system. For this purpose, a state-of-art parallel tempering metadynamics (PT-

^{*}To whom correspondence should be addressed. Tel: +82 51 510 2235; Fax: +82 51 516 7421; Email: ypak@pusan.ac.kr

[†]These authors contributed equally to the paper as first authors.

Present address: Prof. Youngshang Pak, Department of Chemistry and Institute of Functional Materials, Pusan National University, Busan 609-735, South Korea

MetaD) simulation method (29) was employed to accelerate MD sampling in conjunction with the recently developed all-atom AMBER bsc1 force field (30). Unfortunately even with this simulation (PTMetaD/bsc1) extended to a time scale of 10 μ s, the native structure of TBA-15 was not located as the lowest free energy minimum state. For resolving this problem, we attempted to tune the bsc1 force field via a simple correction of van der Waals (vdW) parameters using an unequal combination rule (31,32) (See ‘Materials and Methods’ section below for details). We found that this modified version of the bsc1 (bsc1_vdW) produced substantially improved free energy profiles of the TBA-15 folding with such biased energetics toward unfolded states removed. Interestingly, the resulting folding free energy landscape of TBA-15 indicated multiple folding pathways with several interesting intermediates states of G-triplex variants.

MATERIALS AND METHODS

System setup

The initial coordinates of the thrombin binding aptamer (TBA-15) d(G₂T₂G₂TGTG₂T₂G₂) were taken from the 1:1 potassium bound structure (PDB ID: 1C35)(24). In this study, the standard AMBER bsc1 force field (30) and its modified version (bsc1_vdW) were used. The detailed parameters of the bsc1_vdW force field are presented in Table 1. The NMR native structure of TBA-15 was solvated with a total number of 3292 OPC (33) water molecules in an octahedron box of 51 Å with a minimal distance of 10 Å between the DNA and box edges. A total of 14 K⁺ ions were inserted by replacing the water molecules for charge-neutralization and then additional 9 KCl ion pairs were then inserted to achieve a salt concentration of 150 mM. The Joung-Cheatham ion parameters (34) suggested for the TIP4P_{EW} water model (35) were used for K⁺ and Cl⁻ ions. These ion parameters ($\sigma = 2.83306$ Å and $\epsilon = 0.2794651$ kcal/mol for K⁺ and $\sigma = 4.91776$ Å and $\epsilon = 0.0116615$ kcal/mol for Cl⁻) appeared to be reasonable in describing the interactions between O6 atoms of G-quartets and K⁺ ions (36,37).

The final system consisted of a total of 13688 atoms. The starting structures for PTMetaD simulations were fully unfolded structures, which were obtained from 10 ns ordinary MD simulation at high temperature of 800 K initiated from the NMR structure (PDB ID: 1C35) (24) under the NVT condition. Supplementary Figure S1 shows the initial structures used for all 24 replicas.

Equilibration

The system described above was minimized using the standard protocol of the position restraints and energy minimizations and followed by 1 ns MD equilibration at 300 K and 1 atm. The pressure was controlled by using Berendsen barostat (38) with a coupling constant of 2 ps and the temperature was fixed using the modified Berendsen thermostat (39) with a coupling constant of 0.1 ps. A time step of 4 fs was employed using the hydrogen mass-repartitioning scheme (40). The electrostatic interactions were calculated using the particle-mesh Ewald (PME) method (41) with a cutoff 10 Å.

Table 1. Modified vdW radii (Gromacs format) employed in the current study (bsc1_vdW)

Atom type (atom names)	σ (Å)	
	bsc1	bsc1_vdW
Intramolecular interactions		
Nucleobase oxygen*	2.95992	2.88592
Nucleobase nitrogen*	3.25000	3.16875
Intermolecular interactions		
Nucleobase oxygen/OW _{opc}	3.06323	3.06323
Nucleobase oxygen/K ⁺	2.89649	2.89649
Nucleobase oxygen/Cl ⁻	3.93884	3.93884
Nucleobase nitrogen/OW _{opc}	3.20828	3.20828
Nucleobase nitrogen/K ⁺	3.04153	3.04153
Nucleobase nitrogen/Cl ⁻	4.08388	4.08388

*Nucleobase oxygen considers atomtype O (atom names: O2(C,T), O4(T), O6(G)). Nucleobase nitrogen considers atom type NA(atom names: N1(G),N3(T)); atom type NB (atom names: N7(A,G)); atom type NC (atom names: N1(A), N3(A,G,C)); atom type N2 (atom names: N6(A), N4(C), N2(G)). The letters, A, C, G and T, in parentheses indicate the adenine, cytosine, guanine and thymine bases, respectively. The Lorentz–Berthelot combination rule was applied differently for the intra- and intermolecular interactions.

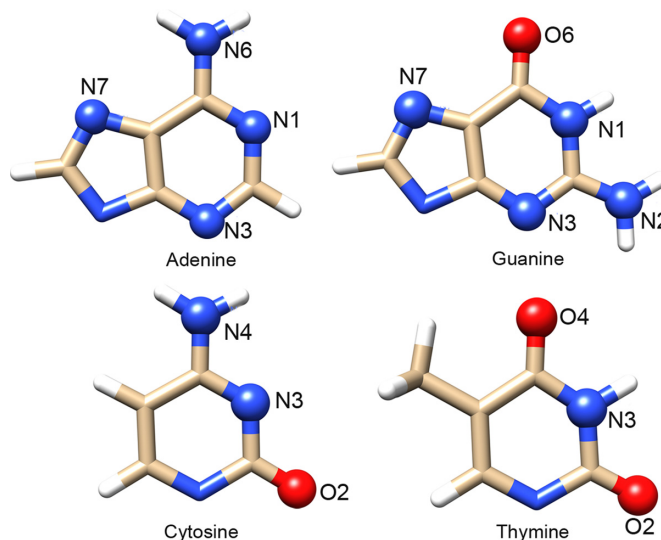


Figure 1. The oxygen and nitrogen atom types of which vdW radii were modified in the present study.

Preparation of the modified bsc1 force field parameters

For the preparation of the bsc1_vdW force field, the original bsc1 force field was modified, such that only the vdW σ values for the DNA base atom types: O, NA, NB, NC and N2 were properly scaled down. These atom types of which σ values were modified are shown in Figure 1. A scale factor of 0.975 appeared to perform well for reproducing the experimental thermodynamic stability of the TBA-15 native structure. Using an unequal combination rule (31,32), these scaled σ parameters were applied only to the intramolecular interactions (within DNA), whereas the original σ values of bsc1 were restored for the intermolecular interactions between DNA-water molecules and DNA-ions. All related files were compressed as ‘bsc1_vdW.ff.zip’ (see Supplementary Data for the compressed file.)

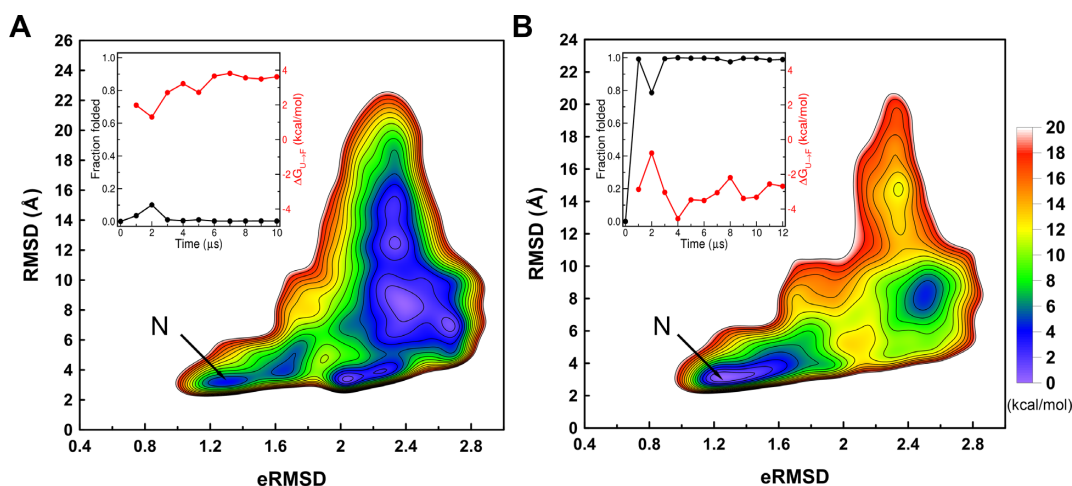


Figure 2. Free energy profiles obtained from PTMetaD simulation at 300 K using (A) bsc1 (10 μ s run per replica) and (B) bsc1_vdW force field (12 μ s run per replica). All simulations were initiated from unfolded states. The native basin is denoted as 'N' (see Figure 3 below for the predicted native structure). The contours were drawn every 1.0 kcal/mol. The cumulative time series of the folded fraction (black line) and free energy difference between the unfolded and folded state (red line) are shown in the inset for each force field tested (See Supplementary Data Section I for calculations of folding fractions).

PTMetaD simulations

As an enhanced MD sampling method for the computation of free energy landscapes, a well-tempered metadynamics (WTMetaD) (27) with parallel tempering scheme (PTMetaD) (29) was employed. Herein, 2D collective variables (CVs) were chosen judiciously to facilitate explorations of the conformational space spanned by these CVs. For the WTMetaD simulation, two CVs of RMSD (heavy atom part) and eRMSD (42,43) were employed for the definition of a 2D biasing potential. The eRMSD metric was designed to compute relative orientations of DNA bases with the backbone part excluded, whereas the RMSD metric depicts global conformational changes including DNA backbones. Therefore, a combination of these two CVs in the present PTMetaD protocol would be beneficial for an effective search of detailed base orientations and backbone conformations.

In the current PTMetaD simulation, a total of 24 replicas were used with each assigned to geometrically distributed temperatures (300.0, 306.7, 313.6, 320.7, 327.9, 335.2, 342.8, 350.5, 358.3, 366.4, 374.6, 383.0, 391.6, 400.4, 409.4, 418.6, 428.0, 437.6, 447.4, 457.5, 467.8, 478.3, 489.0 and 500.0 K). The Gaussian hill height was set initially to $0.014285k_B\Delta T$ and bias factor was set to 15. The Gaussian hill width of 0.1 units was set for both eRMSD and RMSD. The Gaussian bias potential was deposited every 1 ps (250 steps) and a replica exchange was attempted at intervals of 0.5 ps. All PTMetaD simulations were run for 10–12 μ s (per replica) starting from completely unfolded states (Supplementary Figure S1). All simulations were performed using the GROMACS program (5.0.7 version) (44) patched with the PLUMED 2.3 software (45).

RESULTS

Figure 2A shows the calculated free energy profile using the standard bsc1 force field with respect to (eRMSD, RMSD) at $T = 300$ K. Also the free energy difference between

the folded and unfolded states ($\Delta G_{U\rightarrow F} \equiv G_F - G_U$) was monitored cumulatively with time (Figure 2A, Inset). As expected, the time profile of $\Delta G_{U\rightarrow F}$ converged asymptotically (46) to a near steady-state value ($\Delta G_{U\rightarrow F} = 3.8$ kcal/mol). Unfortunately, as this positive $\Delta G_{U\rightarrow F}$ implied, the native state (a two-tetrad GQ) was less stable than the unfolded state. Mostly, this unfolded state displayed diverse hairpin-like base-collapsed (HC) structures. This biased energetics toward such non-native conformations raised an issue in the interpretation of this simulation result. This discrepancy may be in part attributed to the weak H-bonds between bases, as reported in many recent studies (31,47–49). Herein, this paper proposes to increase the H-bond strength of base pairs. This can be achieved by properly decreasing the vdW radii (σ -values) of the H-bond sites of bases (polar N and O atom types, Figure 1). In a similar manner to earlier simulation studies of RNA folding (31,32), using an unequal combination rule, properly scaled σ values of bsc1 were applied to only the intramolecular interactions (DNA part only) and the original σ values were used for the intermolecular interactions between DNA and water molecules as well as between DNA and ions (See Table 1). This scheme can modulate the H-bond strength of bases selectively without much affecting the integrity of the bsc1 force field. After extensive preliminary PTMetaD simulations with those vdW σ -values systematically varied, a scaling factor of 0.975 was found to be satisfactory. (The results from other test values are shown in Supplementary Figure S2).

Initiating from completely unfolded states, a large-scale PTMetaD simulation was carried out for 12 μ s per replica (a total of 288 μ s) using these new σ parameters scaled with this factor (bsc1_vdW, see Table 1 for detailed parameters). This simulation showed that the native-like GQ fold was observed within several μ s and that reversible folding occurred via multiple folding/unfolding events (See Supplementary Figure S3A and B). The lowest free energy minimum structure (GQ₂) is well comparable to the NMR na-

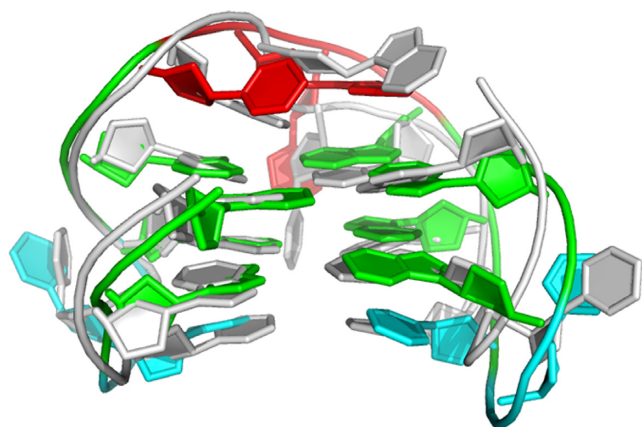


Figure 3. Comparison of the NMR experimental (gray) (PDB ID:1C35) (24) and the lowest free energy (colored) predicted structures of TBA-15. This lowest free energy minimum structure was obtained from the cluster analysis (See Supplementary Data Section IV). The TGT and TT loops are colored red and cyan, respectively. The G-stem is colored green. The predicted native structure is located at (eRMSD, RMSD) = (1.3, 2.4 Å). The predicted G-quartets match well with the experiment (eRMSD.Q = 0.5; See the main text for eRMSD.Q), but the simulated loop structures somewhat deviate from the experiment.

tive structure (Figure 3). Compared to the NMR experiment, the simulation revealed more idealized planar G-quartets but subtle differences in the loop structures. Although the strict convergence in this simulation may not be achieved due to the oscillatory nature of WTMetaD (46), the $\Delta G_{U \rightarrow F}$ value at 300 K reached a stationary value of approximately $\Delta G_{U \rightarrow F} = -2.8$ kcal/mol (See Figure 2B, Inset). Compared to the case of bsc1 (Figure 2A), the modified version of bsc1 (bsc1_vdW) improved the free energy landscape markedly in that the $\Delta G_{U \rightarrow F}$ value conformed to the experimental values of $-1.1 \sim -3.3$ kcal/mol (50–52) (Figure 2B).

The aforementioned CVs (eRMSD and RMSD) were designed to probe the overall conformational changes relative to the reference state, e.g. the TBA native structure. As indicated in Figure 2, however, the local conformational changes specific to the G-tetrads were not well resolved with this representation. To provide better-resolved atomistic pictures of the formation of the central GQ core, a subset of eRMSD (eRMSD.Q) was introduced by considering only the eight-guanine bases participating in the formation of the GQ core. With this new CV, the free energy surface, as shown in Figure 2B, was reconstructed in terms of (eRMSD.Q, RMSD) (Figure 4A) using the reweighting scheme proposed by Tiwary *et al.* (53) (See Supplementary Data Section II and Figure S4). In this new free energy representation, a total of seven minima were identified and properly labeled as shown in Figure 4A. To gain insight into which stage of folding K^+ ion is being captured, another free energy surface was computed as a function of (eRMSD.Q, CN) (Figure 4B), where CN is the ion coordination number (Supplementary Data Section III for the definition of CN). As shown in Figure 4A, TBA-15 folding occurs via multiple pathways in the presence of the four local intermediate states positioned between the unfolded (HC) and folded (GQ₂) basins. All the possible folding routes and the repre-

sentative local minimum structures are summarized in Figure 5A. As can be seen from Figures 4A and 5A, the global minimum at (eRMSD.Q, RMSD) = (0.5, 2.4 Å) is unambiguously the native-like GQ₂. The GQ₂ contains the upper plane of a tetrad (G1, G6, G10, G15) and the lower plane of a tetrad (G2, G5, G11, G14) with a single K^+ captured between these planes, which is consistent with the NMR result of TBA-15 (PDB ID: 1C35) (24). The second most stable basin centered at (1.6, 8.0 Å) is the aforementioned HC structure. As the most abundant unfolded state, the HC state exhibited the diverse structural features of non-native base stacking and H-bonds with some loop formations but without K^+ ion capturing. In addition to these two major populated states of GQ₂ and HC, other shallow local minimum states need to be addressed. As the third most stable state, i.e. the major local intermediate, the energy minimum at (1.2, 6.0 Å) represents the GT₂ structure. This intermediate displays two triads of (G1, G6, G10) and (G2, G5, G11) with a single K^+ ion captured in between, matching well with the earlier NMR structure of TBA-11. In GT₂, the 3'-terminal G14 and G15, which were truncated in the TBA-11, are dangling from the lower G-triad. The GT₂ shows auxiliary H-bonding features, so that the TGT loop interacts with the upper triad and the two TT-loops form a tight H-bond bridge between T4 and T13. Interestingly, this TT-loop bridge tends to prevent the dangling of G14 and G15 from approaching the G-stem for a final assembly of the native G-tetrads. The local minimum state at (1.2, 3.0 Å) exhibited a mixed structure of triad/tetrad (GTQ'); its upper triad (G1, G6, G10) matches with that of the NMR structure of TBA-11, but its lower tetrad (G2, G5, G11, G15) deviates from the native arrangement: G2, G5, G11, G14. G14 is displaced from the lower tetrad and forms a non-native H-bond with the T4 (TT-loop) because G15 replaces the G14 incorrectly in the lower tetrad. The local minimum at (1.7, 4.0 Å) corresponds to a loosely formed two-triad structure (GT₂''). Its upper triad displays native-like base ordering (G1, G6, G10), but its lower triad indicates non-native base ordering (G5, G11, G15). In contrast to GT₂ and GT₂', the triads of GT₂'' were somewhat distorted from their planar geometry through non-native H-bonding. Another local minimum at (1.4, 5.0 Å) exhibits a misfolded two-triad structure (GT₂') at variance with the GT₂ structure, such that the G1 (upper triad) and G2 (lower triad) are swapped. Because of this swapping, the T3T4 loop was severely distorted (Figure 5A). Note that the free energy levels of GT₂', GT₂'' and GTQ' are comparable, but slightly higher than that of GT₂. Interestingly, such G-triplex-like structures were also predicted by the simulation study of a human telomere sequence by Stadlbauer *et al.* (15).

As shown in Figure 4B, the free energy map on (eRMSD.Q, CN) enabled K^+ -binding states to be traced in detail during the course of TBA-15 folding. In the unfolded state (HC state), any specific ion binding pocket was not created, thereby no sign of the ion binding. In fact, the K^+ capturing was commenced upon the transition from the HC state to the following local triplex variants: GT₂', GT₂ and GTQ'. Hence, the ion binding appears to play a prominent role in the formation of such triplex intermediates by stabilizing these structures. For example, this ion coordination ($I + K^+ \rightarrow I \cdots K^+$) stabilized the GT₂ or GTQ'

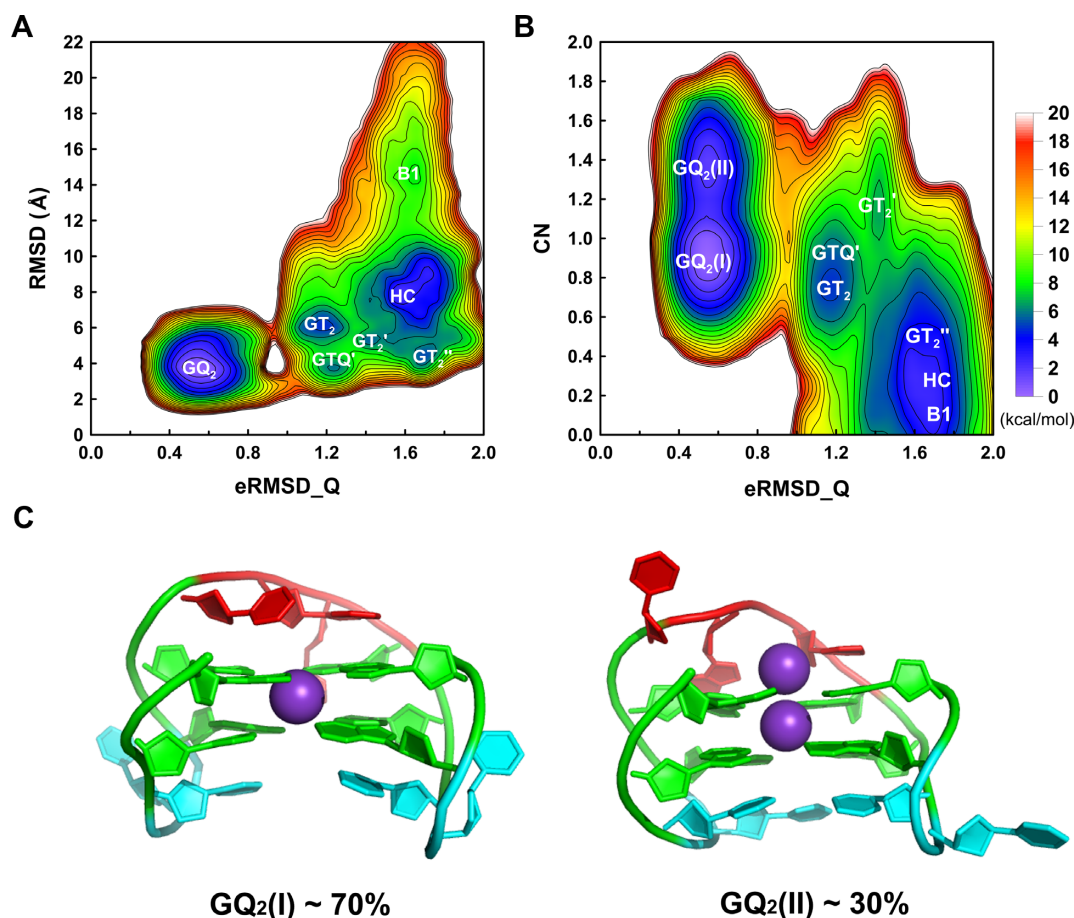


Figure 4. Reweighted free energy profiles of the bsc1_vdW at 300 K in terms of (A) (eRMSD_Q, RMSD) and (B) (eRMSD_Q, CN). The local minima are labeled as GQ₂ (Native), GTQ', GT₂, GT₂', GT₂'', HC and B1 (See the main text for details). (C) The single and double coordinated K⁺ ion structures of GQ₂(I) (left) and GQ₂(II) (right). The TGT and TT loops are colored red and cyan, respectively. The G-stem is colored green and the K⁺ ion is shown as a violet sphere.

state by almost 6 kcal/mol. This free energy gain was obtained with reference to the state without the ion binding at (eRMSD_Q, CN) = (1.20, 0.10) (Figure 4B).

The native basin (GQ₂) was resolved into two distinct minima (Figure 4B), which corresponded to the single and double K⁺-binding GQ₂ structures (Figure 4C), which were denoted as GQ₂(I) and GQ₂(II), respectively. In terms of the population, the single K⁺-binding GQ₂ was more abundant than the double K⁺-binding GQ₂. The presence of GQ₂(II) was previously suggested by Marathias and Bolton (24,54). For GQ₂(II), they proposed that the first binding site be located between the stem-loop junction of the lower quartet (G2, G5, G11, G14) and the TT loop. The second binding site was proposed to be located between the TGT loop and the upper quartet (G1, G6, G10, G15). Unfortunately, the first binding site proposed was not supported by the earlier (18) and present simulation studies. Instead, the central core of the G-quartets is the binding site for a single K⁺ ion. The second binding site proposed matched well with the present simulation result (See Figure 4C). In the present GQ₂(II) structure, the second K⁺ ion was bound primarily in the TGT loop region, but this ion can stay in this binding site transiently since it can easily be detached from the binding pocket by crossing a small barrier of 0.6 kcal/mol to

form more stable GQ₂(I). Compared to the GQ₂(II) structure derived from the NMR experiment (24), the TGT loop as well as the upper G-tetrad were somewhat distorted.

Within the context of the current simulation, the overall picture of TBA-15 folding can be described briefly as follows (See Figure 5A). Starting from a single stranded B-DNA structure (B1), the folding is initiated by nonspecific and rapid collapse into the HC state. From the HC basin, the uphill transition to each neighboring basin of the three variants of the G-triplex proceeds by passing over moderate energy barriers of 4–5 kcal/mol. In this stage, mostly a single K⁺ ion is accommodated by the guanine bases. Another local state of GTQ' can be accessible from its two neighboring states (GT₂ and GT₂'). All four intermediate states, which have similar free energy levels, can be interconvertible through various energy barriers ranging from 0.6 to 3.6 kcal/mol. Among these four intermediates, however, only GTQ' and GT₂ can serve as productive folding channels for completing each transition to the native basin. Therefore, as these key intermediate states were reached, two track folding events from GTQ' and GT₂ occurred by overcoming the substantial energy barriers of 9.6 kcal/mol on a fast track and 12.3 kcal/mol on a slow track, respectively. For these two key barrier-crossing events, Figure 5B illustrates

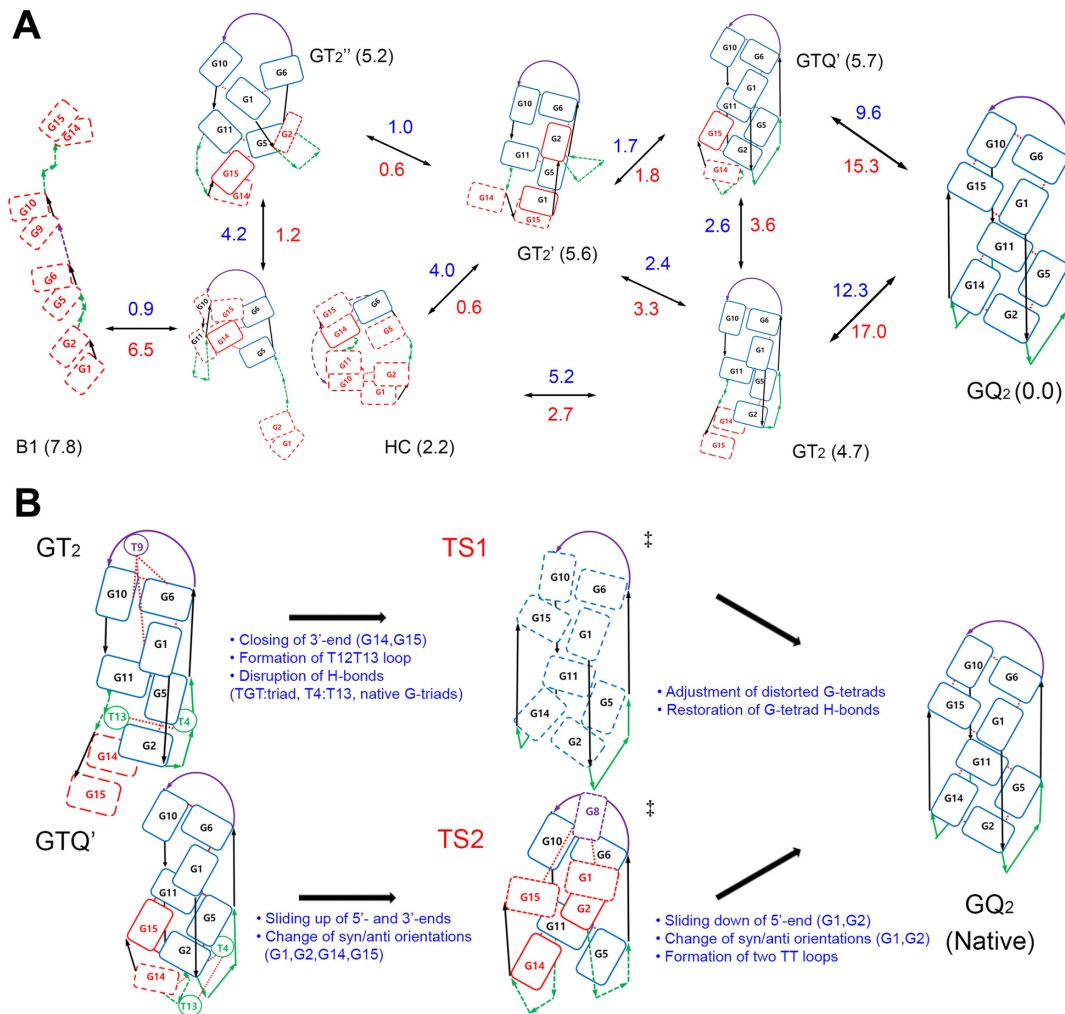


Figure 5. (A) Schematic representation of all possible folding pathways showing the progress from the fully unfolded B1 state to the folded GQ₂ state. The free energy level of each local minimum relative to the GQ₂ state (kcal/mol) is given in the parentheses. At each stage of progress, the energy barriers in the forward (folding) and backward (unfolding) directions (kcal/mol) are given alongside the arrow in blue and red numbers, respectively. The rounded rectangles with the solid and dash borderlines indicate guanine bases with native (misfolded) and non-native (misfolded) arrangements, respectively. The blue and red borderlines of the rectangles depict the guanine bases in plane and out of plane of the G-triad, respectively. The native TGT and TT loops are colored violet and green, respectively. The misfolded TT loop is represented by a green dash line. The hydrogen bond is denoted as a red dot line connecting two bases. (B) Schematic views of two large barrier-crossing events starting from the two-triad (GT₂) and the mixed triad/tetrad (GTQ') intermediates. TS1 and TS2 are the transition states for these two events.

detailed molecular pictures including transition states (TS1 and TS2). In the case of the slow track barrier crossing, the transition from GT₂ to TS1 involved the closing of 3'-end G14 and G15 residues, the formation of TT loop (T12 and T13), and the disruption of various H-bonds between the TGT loop and the upper triad, T13 and T4, and the guanine bases in the triads. In particular, the disrupted H-bonds in the G-triads caused a distortion of G-triads to some extent. The final transition from the transition state (TS1) to the native state was completed by the adjustment of distorted G-tetrads, the restoration of the native H-bonds in the G-tetrads, and anti/syn changes of G11 and G14 bases. On the other hand, in the case of the fast track barrier crossing, the barrier crossing from GTQ' encountered the sliding up of G1 and G2 at the 5'-end and G14 and G15 at the 3'-end, followed by the alteration of syn/anti orientations of these terminal guanine bases. From the transition state (TS2), the

native state was formed by a series of interesting molecular events: sliding down of G1 and G2, finding their correct syn/anti orientations, and finally creation of the native two TT loops (See Figure 5B).

The PTMetaD allowed us to directly compute free energy profiles at different temperature of TBA-15 and its melting profile (folding fraction vs. temperature). Figure 6A and B present the free energy surfaces at a near melting temperature (358.3 K) and a higher temperature (467.8 K), respectively. (For the error estimates of these free energy profiles, see Supplementary Figure S5). As shown in Figure 6A, the GT₂ structure was still populated at the near melting point. Eventually at 467.8 K, the GT₂ population was mostly diminished and as expected, the unfolded state (HC) appeared as the global minimum (Figure 6B). The melting profile (Figure 6C) was well fitted to the standard two-state thermodynamic equations (see Supplementary Data Section V).

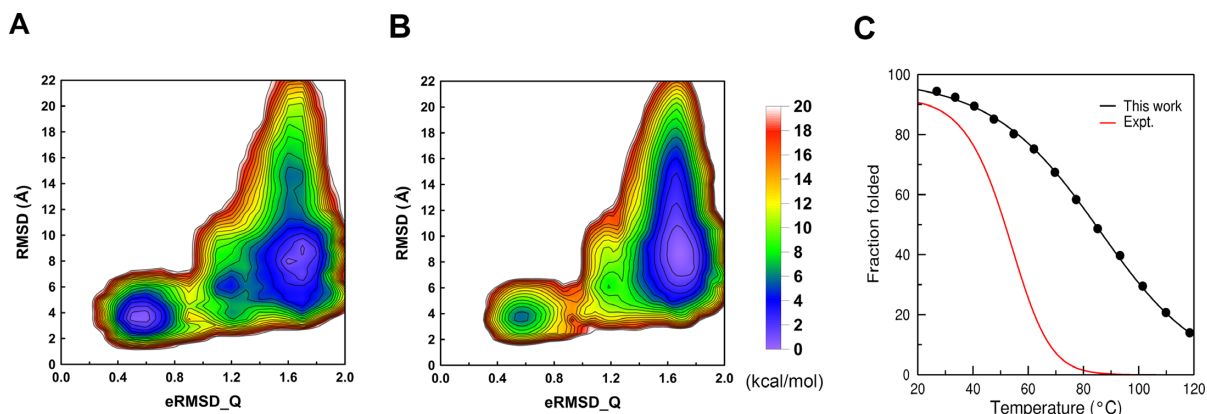


Figure 6. Free energy profiles of the bsc1_vdW in terms of eRMSD_Q and RMSD (A) at near melting temperature (358.3 K) and (B) at higher temperature of 467.8 K. (C) Comparison of the simulated melting profile (black) with the experimental melting profile (58) (red).

As was the case with most all-atom force fields (55–57), the melting temperature (T_m) was overestimated and the heat capacity change (ΔC_v) was underestimated compared to the experiment (58) (See Figure 6C and Supplementary Table S1). In terms of reproducing temperature-dependent properties, the all-atom force field needs to be further improved to close the gap between simulation and experiment.

DISCUSSION

In the present study, we performed extended time scale PT-MetaD simulations using a modified version of the AMBER bsc1 force field (bsc1_vdW) as well as the standard bsc1 force field. In contrast to the case of bsc1, the bsc1_vdW successfully produced energetically reliable free energy representations of the TBA-15 folding. The resulting free energy surface revealed the three distinct basins of the unfolded state (U), the intermediate state (I), and the folded state (F). The unfolded basin consists of the HC state. These HC conformations are somewhat similar to entrapped misfolded hairpin conformations observed in the previous simulation study of small DNA hairpin folding (59,60). The intermediate basin contains diverse G-triplex structures: GT_2 , GT_2' , GT_2'' and GTQ' . These G-triplexes can be easily transformed into one another by crossing relatively small barriers. In the transition from the U to I state, a single K^+ ion capturing facilitated the formation of the G-triplex intermediates of GT_2 , GT_2' and GTQ' . These G-triplex structures were substantially stabilized by the complexation with the ion. Energetically, a single ion binding to the triplex intermediate of GT_2 or GTQ' in its binding pocket lowered the free energy by about 6 kcal/mol. Thus, it is evident that the GQ folding cannot be possible without the ions capturing. As indicated by the relatively moderate free energy barriers (4–5 kcal/mol), this kinetic event from U to I seems relatively fast. The G-triplex conformations in the I-state appeared to be misaligned or topologically frustrated, encountering substantial folding barriers of 10–12 kcal/mol. Thus, we expect that the transition from the I- to F-state would be rather slow, accompanying diverse molecular rearrangements prior to the native GQ folding, such as the disruption of H-bond networks, the slippage of both 5'- and 3'-side strands, and the proper changes of anti/syn base

orientations (See Figure 5B). All these key observations: the ion coordination during the formation of G-triplex variants as a fast kinetic event and a subsequent GQ folding through various molecular rearrangements as a slow kinetic process are indicative of possibly a three-state folding kinetics.

Previously, a kinetic partitioning mechanism (KPM) (61) was suggested for the possible folding kinetics of GQs (48). Especially for the human telomeric GQs (three layers of G-tetrad; folding time of minutes to hours), many experimental (3–9) and theoretical studies (14,15,17,62) indicate that the KPM is an operating scenario for the folding kinetics. But at least for the TBA-15 (two layers of G-tetrad), the current free energy landscape shows that the KPM is unlikely to be the case, due to the rather facile transitions among these local intermediates states on the multiple folding tracks. However, considering the limited dimensionality of the 2D CV space employed in this work, we do not exclude the possibility that the current simulation protocol may not capture other slower folding funnel including off-folding traps relevant to KPM (48).

Although the present study was performed using the AMBER bsc1 force field and its modified version (AMBER bsc1_vdW), it is worthwhile to mention other promising DNA force field of AMBER OL15 (26,63–65). Previous benchmark studies (66,67) on DNA force fields showed that the performance of AMBER OL15 is comparable to that of the AMBER bsc1. The recent NMR and simulation study of a short G-hairpin DNA indicated that the AMBER OL15 also underestimated the strength of base-pair H-bonds (49). Hence, in the folding of TBA-15, similar improvement can be achieved when the present vdW correction is applied to the AMBER OL15.

The present vdW correction scheme proposed in this study can be a simple and comprehensive way to modulate specific non-bonding interactions in molecular mechanics force fields. Herein, this scheme was applied to only base pair H-bonding sites for enhancing their H-bond strength. Although it is straightforward to extend this correction to the interaction between base and water molecules or between bases and ions, we found it sufficient to target only these base H-bonding sites at least for the TBA-15 case. This is because the ion parameters employed in this study (an

optimized version for TIP4P_{EW}) appear to be reasonable in describing base-ion interactions (37).

This correction scheme can also be used to improve all-atom RNA force fields. Despite much effort to develop all-atom RNA force fields, it is well-known that the standard RNA force fields exhibit some imbalances in the intramolecular interactions involving the phosphate oxygen atoms and base hydroxyl group, base pair H-bonds and base stacking interactions (31,32,47,68,69). As shown in the previous folding study of RNA tetranucleotides, the problem with the oxygen atoms in the phosphate group and hydroxyl group was successfully resolved using this vdW correction scheme (32). We anticipate that other imbalances in the RNA force field can be properly dealt with this scheme.

The present study appears to be the first successful direct folding simulation of TBA-15 to date at all-atom level. We expect that this simulation protocol in conjunction with bsc1_vdW can pave a way to unravel more challenging issues of GQ folding problems of human telomeric sequences.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Research Foundation [NRF-2014R1A4A1001690, NRF-2016R1A2B4009517]; Pusan National University ('2016 Post-Doctoral Development Program') (to C.Y.); KISTI [KSC-2017-C3-0022 program]; IREMB supercomputers at DGIST. Funding for open access charge: National Research Foundation of Korea [NRF-2014R1A4A1001690].

Conflict of interest statement. None declared.

REFERENCES

- Shim, J.W., Tan, Q.L. and Gu, L.Q. (2009) Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity. *Nucleic Acids Res.*, **37**, 972–982.
- Liu, W., Fu, Y., Zheng, B., Cheng, S., Li, W., Lau, T.C. and Liang, H.J. (2011) Kinetics and mechanism of conformational changes in a G-quadruplex of thrombin-binding aptamer induced by Pb²⁺. *J. Phys. Chem. B*, **115**, 13051–13056.
- Lee, J.Y., Okumus, B., Kim, D.S. and Ha, T.J. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 18938–18943.
- Aznauryan, M., Sondergaard, S., Noer, S.L., Schiott, B. and Birkedal, V. (2016) A direct view of the complex multi-pathway folding of telomeric G-quadruplexes. *Nucleic Acids Res.*, **44**, 11024–11032.
- Bessi, I., Jonker, H.R., Richter, C. and Schwalbe, H. (2015) Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex. *Angew. Chem. Int. Ed.*, **54**, 8444–8448.
- Gray, R.D., Trent, J.O. and Chaires, J.B. (2014) Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.*, **426**, 1629–1650.
- Gabelica, V. (2014) A pilgrim's guide to G-quadruplex nucleic acid folding. *Biochimie*, **105**, 1–3.
- Marchand, A. and Gabelica, V. (2016) Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Res.*, **44**, 10999–11012.
- Long, X. and Stone, M.D. (2013) Kinetic partitioning modulates human telomere DNA G-quadruplex structural polymorphism. *PLoS One*, **8**, e83420.
- Kim, E., Yang, C. and Pak, Y. (2012) Free-energy landscape of a thrombin-binding DNA aptamer in aqueous environment. *J. Chem. Theory Comput.*, **8**, 4845–4851.
- Limongelli, V., De Tito, S., Cerofolini, L., Fragai, M., Pagano, B., Trotta, R., Cosconati, S., Marinelli, L., Novellino, E., Bertini, I. et al. (2013) The G-Triplex DNA. *Angew. Chem. Int. Ed.*, **52**, 2269–2273.
- Zeng, X.J., Zhang, L.Y., Xiao, X.C., Jiang, Y.Y., Guo, Y.Z., Yu, X.Y., Pu, X.M. and Li, M.L. (2016) Unfolding mechanism of thrombin-binding aptamer revealed by molecular dynamics simulation and Markov State Model. *Sci. Rep.*, **6**, 24065.
- Sun, L., Jin, H., Zhao, X., Liu, Z., Guan, Y., Yang, Z. and Zhang, L. (2014) Unfolding and conformational variations of thrombin-binding DNA aptamers: synthesis, circular dichroism and molecular dynamics simulations. *Chemmedchem*, **9**, 993–1001.
- Stadlbauer, P., Kuhrova, P., Banas, P., Koca, J., Bussi, G., Trantirek, L., Otyepka, M. and Sponer, J. (2015) Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res.*, **43**, 9626–9644.
- Stadlbauer, P., Trantirek, L., Cheatham, T.E. 3rd, Koca, J. and Sponer, J. (2014) Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie*, **105**, 22–35.
- Bian, Y.Q., Tan, C., Wang, J., Sheng, Y.B., Zhang, J. and Wang, W. (2014) Atomistic picture for the folding pathway of a hybrid-I type human telomeric DNA G-quadruplex. *PLOS Comput. Biol.*, **10**, e1003562.
- Stadlbauer, P., Krepl, M., Cheatham, T.E., Koca, J. and Sponer, J. (2013) Structural dynamics of possible late-stage intermediates in folding of quadruplex DNA studied by molecular simulations. *Nucleic Acids Res.*, **41**, 7128–7143.
- Reshetnikov, R.V., Sponer, J., Rassokhina, O.I., Kopylov, A.M., Tsvetkov, P.O., Makarov, A.A. and Golovin, A.V. (2011) Cation binding to 15-TBA quadruplex DNA is a multiple-pathway cation-dependent process. *Nucleic Acids Res.*, **39**, 9789–9802.
- Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.*, **19**, 239–250.
- Jiang, Y.L. and Liu, Z.P. (2010) Metallo-Organic G-Quadruplex Ligands in Anticancer Drug Design. *Mini. Rev. Med. Chem.*, **10**, 726–736.
- Cummaro, A., Fotticchia, I., Franceschin, M., Giancola, C. and Petraccone, L. (2011) Binding properties of human telomeric quadruplex multimers: a new route for drug design. *Biochimie*, **93**, 1392–1400.
- Ralph, S.F. (2011) Quadruplex DNA: A Promising Drug Target for the Medicinal Inorganic Chemist. *Curr. Top. Med. Chem.*, **11**, 572–590.
- Xing, H., Wong, N.Y., Xiang, Y. and Lu, Y. (2012) DNA aptamer functionalized nanomaterials for intracellular analysis, cancer cell imaging and drug delivery. *Curr. Opin. Chem. Biol.*, **16**, 429–435.
- Marathias, V.M. and Bolton, P.H. (2000) Structures of the potassium-saturated, 2:1, and intermediate, 1:1, forms of a quadruplex DNA. *Nucleic Acids Res.*, **28**, 1969–1977.
- Schultze, P., Macaya, R.F. and Feigon, J. (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Biol.*, **235**, 1532–1547.
- Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E. 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
- Barducci, A., Bussi, G. and Parrinello, M. (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, **100**, 020603.
- Cerofolini, L., Amato, J., Giachetti, A., Limongelli, V., Novellino, E., Parrinello, M., Fragai, M., Randazzo, A. and Luchinat, C. (2014) G-triplex structure and formation propensity. *Nucleic Acids Res.*, **42**, 13393–13404.
- Bussi, G., Gervasio, F.L., Laio, A. and Parrinello, M. (2006) Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.*, **128**, 13435–13441.
- Ivani, I., Dans, P.D., Noy, A., Perez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goni, R., Balaceanu, A. et al. (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.

31. Chen, A.A. and Garcia, A.E. (2013) High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 16820–16825.
32. Yang, C., Lim, M., Kim, E. and Pak, Y. (2017) Predicting RNA structures via a simple van der Waals correction to an all-atom force field. *J. Chem. Theory Comput.*, **13**, 395–399.
33. Izadi, S., Anandakrishnan, R. and Onufriev, A.V. (2014) Building water models: a different approach. *J. Phys. Chem. Lett.*, **5**, 3863–3871.
34. Joung, I.S. and Cheatham, T.E. 3rd (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020–9041.
35. Horn, H.W., Swope, W.C., Pitera, J.W., Madura, J.D., Dick, T.J., Hura, G.L. and Head-Gordon, T. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, **120**, 9665–9678.
36. Fadrna, E., Spackova, N., Sarzynska, J., Koca, J., Orozco, M., Cheatham, T.E. 3rd, Kulinski, T. and Sponer, J. (2009) Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields. *J. Chem. Theory Comput.*, **5**, 2514–2530.
37. Havrila, M., Stadlbauer, P., Islam, B., Otyepka, M. and Sponer, J. (2017) Effect of monovalent ion parameters on molecular dynamics simulations of G-quadruplexes. *J. Chem. Theory Comput.*, **13**, 3911–3926.
38. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
39. Bussi, G., Donadio, D. and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 014101.
40. Feenstra, K.A., Hess, B. and Berendsen, H.J.C. (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.*, **20**, 786–798.
41. Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
42. Bottaro, S., Di Palma, F. and Bussi, G. (2014) The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res.*, **42**, 13306–13314.
43. Bottaro, S., Banas, P., Sponer, J. and Bussi, G. (2016) Free energy landscape of GAGA and UUCG RNA tetraloops. *J. Phys. Chem. Lett.*, **7**, 4032–4038.
44. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
45. Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C. and Bussi, G. (2014) PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.*, **185**, 604–613.
46. Dama, J.F., Parrinello, M. and Voth, G.A. (2014) Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.*, **112**, 240602.
47. Kuhrova, P., Best, R.B., Bottaro, S., Bussi, G., Sponer, J., Otyepka, M. and Banas, P. (2016) Computer folding of RNA tetraloops: identification of key force field deficiencies. *J. Chem. Theory Comput.*, **12**, 4534–4548.
48. Sponer, J., Bussi, G., Stadlbauer, P., Kuhrova, P., Banas, P., Islam, B., Haider, S., Neidle, S. and Otyepka, M. (2017) Folding of guanine quadruplex molecules—funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies. *Biochim. Biophys. Acta*, **1861**, 1246–1263.
49. Gajarsky, M., Zivkovic, M.L., Stadlbauer, P., Pagano, B., Fiala, R., Amato, J., Tomaska, L., Sponer, J., Plavec, J. and Trantirek, L. (2017) Structure of a Stable G-Hairpin. *J. Am. Chem. Soc.*, **139**, 3591–3594.
50. Smirnov, I. and Shafer, R.H. (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry*, **39**, 1462–1468.
51. Olsen, C.M., Lee, H.T. and Marky, L.A. (2009) Unfolding thermodynamics of intramolecular G-quadruplexes: base sequence contributions of the loops. *J. Phys. Chem. B*, **113**, 2587–2595.
52. Avino, A., Portella, G., Ferreira, R., Gargallo, R., Mazzini, S., Gabelica, V., Orozco, M. and Eritja, R. (2014) Specific loop modifications of the thrombin-binding aptamer trigger the formation of parallel structures. *FEBS J.*, **281**, 1085–1099.
53. Tiwary, P. and Parrinello, M. (2015) A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B*, **119**, 736–742.
54. Marathias, V.M. and Bolton, P.H. (1999) Determinants of DNA quadruplex structural type: sequence and potassium binding. *Biochemistry*, **38**, 4355–4364.
55. Yang, C., Jang, S. and Pak, Y. (2014) A fully atomistic computer simulation study of cold denaturation of a beta-hairpin. *Nat. Commun.*, **5**, 5773.
56. Jiang, F. and Wu, Y.D. (2014) Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. *J. Am. Chem. Soc.*, **136**, 9536–9539.
57. Miner, J.C., Chen, A.A. and Garcia, A.E. (2016) Free-energy landscape of a hyperstable RNA tetraloop. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 6665–6670.
58. Majhi, P.R., Qi, J.Y., Tang, C.F. and Shafer, R.H. (2008) Heat capacity changes associated with guanine quadruplex formation: An isothermal titration calorimetry study. *Biopolymers*, **89**, 302–309.
59. Kannan, S. and Zacharias, M. (2007) Folding of a DNA hairpin loop structure in explicit solvent using replica-exchange molecular dynamics simulations. *Biophys. J.*, **93**, 3218–3228.
60. Portella, G. and Orozco, M. (2010) Multiple routes to characterize the folding of a small DNA hairpin. *Angew. Chem. Int. Ed.*, **49**, 7673–7676.
61. Thirumalai, D. and Woodson, S.A. (1996) Kinetics of folding of proteins and RNA. *Acc. Chem. Res.*, **29**, 433–439.
62. Stadlbauer, P., Mazzanti, L., Cragolini, T., Wales, D.J., Derreumaux, P., Pasquali, S. and Sponer, J. (2016) Coarse-grained simulations complemented by atomistic molecular dynamics provide new insights into folding and unfolding of human telomeric G-quadruplexes. *J. Chem. Theory Comput.*, **12**, 6077–6097.
63. Zgarbová, M., Šponer, J., Otyepka, M., Cheatham, T.E., Galindo-Murillo, R. and Jurečka, P. (2015) Refinement of the sugar–phosphate backbone torsion β for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.*, **11**, 5723–5736.
64. Krepl, M., Zgarbová, M., Stadlbauer, P., Otyepka, M., Banáš, P., Koča, J., Cheatham, T.E., Jurečka, P. and Šponer, J. (2012) Reference simulations of noncanonical nucleic acids with different χ variants of the AMBER force field: quadruplex DNA, quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.*, **8**, 2506–2520.
65. Zgarbová, M., Luque, F.J., Šponer, J., Cheatham, T.E., Otyepka, M. and Jurečka, P. (2013) Toward improved description of DNA backbone: revisiting ϵ and ζ torsion force field parameters. *J. Chem. Theory Comput.*, **9**, 2339–2354.
66. Galindo-Murillo, R., Robertson, J.C., Zgarbova, M., Sponer, J., Otyepka, M., Jurecka, P. and Cheatham, T.E. (2016) Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.*, **12**, 4114–4127.
67. Dans, P.D., Ivani, I., Hospital, A., Portella, G., González, C. and Orozco, M. (2017) How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.*, **45**, 4217–4230.
68. Vangaveti, S., Ranganathan, S.V. and Chen, A.A. (2017) Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *Wiley Interdiscip. Rev. RNA*, **8**, e1396.
69. Bergonzo, C. and Cheatham, T.E. (2015) Improved force field parameters lead to a better description of RNA structure. *J. Chem. Theory Comput.*, **11**, 3969–3972.