

# Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA

Marie Zgarbová,<sup>†</sup> Michal Otyepka,<sup>†,‡</sup> Jiří Šponer,<sup>†,‡</sup> Filip Lankáš,<sup>§</sup> and Petr Jurečka<sup>\*,†</sup>

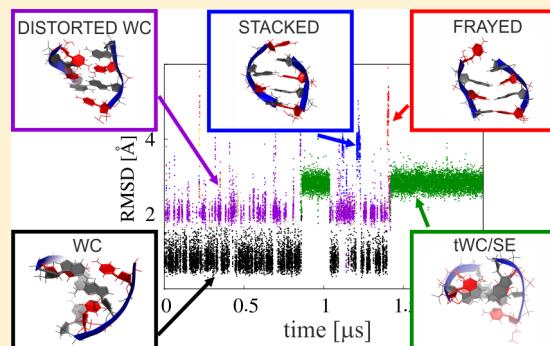
<sup>†</sup>Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University, 17. listopadu 12, 77146 Olomouc, Czech Republic

<sup>‡</sup>Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

<sup>§</sup>Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 166 10 Prague, Czech Republic

## Supporting Information

**ABSTRACT:** Terminal base pairs of DNA and RNA molecules in solution are known to undergo frequent transient opening events (fraying). Accurate modeling of this process is important because of its involvement in nucleic acid end recognition and enzymatic catalysis. In this article, we describe fraying in molecular dynamics simulations with the ff99bsc0, ff99bsc0 $\chi_{OL3}$ , and ff99bsc0 $\chi_{OL4}$  force fields, both for DNA and RNA molecules. Comparison with the experiment showed that while some features of fraying are consistent with the available data, others indicate potential problems with the force field description. In particular, multiple noncanonical structures are formed at the ends of the DNA and RNA duplexes. Among them are tWC/sugar edge pair, C–H edge/Watson–Crick pair, and stacked geometries, in which the terminal bases are stacked above each other. These structures usually appear within the first tens to hundreds of nanoseconds and substantially limit the usefulness of the remaining part of the simulation due to geometry distortions that are transferred to several neighboring base pairs (“end effects”). We show that stability of the noncanonical structures in ff99bsc0 may be partly linked to inaccurate glycosidic ( $\chi$ ) torsion potentials that overstabilize the *syn* region and allow for rapid *anti* to *syn* transitions. The RNA refined glycosidic torsion potential  $\chi_{OL3}$  provides an improved description and substantially more stable MD simulations of RNA molecules. In the case of DNA, the  $\chi_{OL4}$  correction gives only partial improvement. None of the tested force fields provide a satisfactory description of the terminal regions, indicating that further improvement is needed to achieve realistic modeling of fraying in DNA and RNA molecules.



## 1. INTRODUCTION

Fraying of nucleic acids is a reversible process in which the terminal base pairs of the duplex lose their base pairing and unstack from their neighbors, allowing frequent refolding of the terminal base pairs. The importance of these base pair opening events stems from their effects on the stability of DNA and RNA duplexes in solution as fraying has been proposed to be the first step in DNA duplex melting via the so-called “fraying–peeling” mechanism.<sup>1,2</sup> Conversely, formation of the base pairs at the end of a propagating helix may involve certain features of base fraying. Fraying also has biological importance as it influences the action of nucleic acid processing enzymes. The frayed ends may be attacked by nucleases and shortened,<sup>3</sup> or selectively bound by helicases.<sup>4</sup> Fraying of DNA ends also contributes to end recognition and enzymatic catalysis by retroviral integrases.<sup>5</sup> A deeper understanding of the fraying process could thus provide us with useful information for describing many important phenomena.

Fraying has been studied experimentally for a long time. In particular, NMR experiments have provided valuable information on the thermodynamics of this process.<sup>6,7</sup> For instance,

dissociation constants of the terminal base pair were estimated to be around  $K_d = 0.14$  for the terminal GC pair in d(CGCGATCGCG)<sub>2</sub> and  $K_d \geq 2.2$  for the AT terminal base pair in d(TAGCGCTA)<sub>2</sub> (values taken from Table 1 of ref 6 estimated from NMR chemical shift vs temperature measurements, for 25 °C). In the same duplexes, the enthalpy and entropy of fraying were measured to be around 9.1 kcal/mol and 26.3 cal/(mol.K), respectively, for the terminal GC pair and ≤5.5 kcal/mol and ≤18 cal/(mol.K), respectively, for the terminal AT base pair.<sup>6</sup> Thus, the frayed states are enthalpically unfavorable, probably due to the loss of stacking stabilization, but they are stabilized entropically. Unfortunately, no definite information as to the structure of the frayed state could be extracted from the NMR experiments due to its flexible nature.

Regarding the kinetics of fraying, experimental techniques can only provide a relatively wide interval for its rate. From NMR imino proton experiments on DNA, the process is known to be faster than 1 ms.<sup>8</sup> Time resolved Stokes shift

Received: February 13, 2014

optical experiments have suggested that opening of the terminal GC pairs in DNA takes more than 40 ns as it could not be observed on a shorter time scale (although unspecified conformational sampling other than base pair opening was reported in that experiment with a time constant of 5 ps).<sup>9</sup> However, those results were deduced for the terminal CG pair capped with a fluorescent probe, which might affect the terminal base pair dynamics.<sup>10</sup> Fluorescence correlation spectroscopy experiments on a viral DNA initiation sequence gave a fraying time for the terminal GC pair of around 110  $\mu$ s and opening and closing rates of 2300  $s^{-1}$  and 7000  $s^{-1}$ , respectively, suggesting a dissociation constant  $K_d = 0.32$ .<sup>11</sup> However, it should be noted that these values do not correspond to fraying of a blunt terminal pair, but there was a 4 nt long 3' overhang and the DNA was modified by molecules of dye. Because overhangs can thermodynamically stabilize duplexes and the stacking interaction of the dye may affect the stability of the closed conformation, the real fraying time may be much shorter. Perhaps closest to the fraying rates observed in our simulations (see below) are the NMR spin relaxation measurements on GC terminated DNA double helices, which reported end-fraying motions on nanosecond<sup>12</sup> and picosecond to nanosecond<sup>13</sup> time scales. Regarding fraying of AT DNA pairs, faster opening can be expected due to the known lower stability of these base pairs. The time scale of fraying in RNA is yet more uncertain.

Another source of experimental information that may be relevant to fraying is melting experiments on double helices. The stabilization contributed by a terminal base pair step that is lost upon pair breaking may be estimated from a nearest neighbor model derived from melting data (see, e.g., refs 14 and 15). However, if the pair is not broken completely and one of the terminal bases remains stacked to the duplex even during fraying, studies of so-called dangling ends could be relevant. Melting experiments on DNA and RNA duplexes have revealed that the dangling ends often provide additional stabilization to DNA and RNA duplexes, with notable differences between the 3' and 5' ends and between DNA and RNA molecules.<sup>16–18</sup> These differences were tentatively explained by the extent of stacking and by shielding of the hydrogen bonds of the terminal base pair.<sup>19</sup> Stacking is a function of the overlap, which in turn depends on the double helix morphology (A versus B form). It has been shown that the thermodynamics of the dangling ends in RNA correlates with stacking/unstacking data in X-ray databases,<sup>20</sup> where stacking of the 3' ends is much more common than that of its 5' counterparts. The same observation was confirmed in another study focusing on RNA duplex propagation.<sup>21</sup> This implies that the 5' and 3' ends could have different tendencies to unstack during fraying, which may correlate with the dangling-end stabilization measured in the dangling-end experiments.

Whereas frayed motifs are probably common in solution and make a significant contribution to the thermodynamic ensemble,<sup>6</sup> they are relatively scarce in X-ray structures. In RNA, motifs similar to half-unstacked pairs may be found at single strand/double strand junctions.<sup>21</sup> On the other hand, double helical DNA structures are terminated almost exclusively by Watson–Crick (WC) pairs. Rare occurrence of frayed structures in crystal structures may be due to the fact that the opened forms are favored by entropy (see the above discussion on dangling ends), whereas a crystal environment favors low entropy and low enthalpy arrangements. It is also likely that duplex ends are stabilized by contacts with the

surrounding molecules in a crystal. It should be noted that the majority of paired terminal bases are found with *anti* orientation of the glycosidic torsion angle, as in the inner nucleosides.

Theoretical modeling may provide an atomistic insight into the base pair opening process. A single opening event on a short time scale has been described by Cheatham and Kollman<sup>22</sup> and unpairing pathways for a terminal 5' end cytosine have been characterized by Hagan et al.<sup>23</sup> Although base pair fraying was not the prime focus of these studies, fraying has been observed in DNA unfolding computer experiments<sup>1,2</sup> and other DNA studies.<sup>24</sup> Based on their 100 ns temperature-jump simulation, Wong and Pettitt suggested a three-step mechanism for DNA melting, involving fraying and peeling in the first step,<sup>1</sup> which they suggested was consistent with recent experimental evidence on DNA folding and melting.<sup>25,26</sup> The complexity of DNA unfolding has also been demonstrated in the study by Perez and Orozco, who performed a microsecond free molecular dynamics (MD) simulation in which they observed unfolding of Dickerson–Drew's dodecamer (DD, d(CGCGAATTCGCG)<sub>2</sub>).<sup>27</sup> There have been many replica exchange studies of nucleic acids with Cornell et al.<sup>28</sup> based force fields (see, e.g., refs 29–31), and also studies of internal base pair opening<sup>32</sup> in the CHARMM force field,<sup>33</sup> but terminal base pair fraying is rarely discussed in detail in these works. Formation of noncanonical structures, which is another issue closely related to fraying behavior, has recently been described by Dršata et al.<sup>34</sup> However, as we will show below, the insight obtained from MD simulations may be limited by the short time scales available with the current computer resources and also the quality of the empirical force field used.

In force field simulations, torsion angle parametrization is a crucial factor influencing the description of noncanonical structures.<sup>35,36</sup> Here, we will show that the terminal base pair behavior is substantially influenced by the glycosidic torsion angle parametrization. The glycosidic torsion angle  $\chi$  determines the relative orientation of the nucleic acid base and sugar ring. In B-DNA,  $\chi$  is structurally fixed in the high-*anti* minimum ( $\chi \sim 300^\circ$ ) and cannot easily switch to the alternate *syn* orientation ( $\chi \sim 60^\circ$ ) because this would require significant duplex reformation. However, the terminal bases destack more easily than the inner bases and consequently may rotate along the glycosidic bond and interconvert between the *anti* and *syn* orientations. The currently available force fields provide very different descriptions of the glycosidic torsion and therefore differ in modeling the *anti*–*syn* interconversion process. Thus, fraying is tightly connected with the choice of  $\chi$  torsion parametrization, as we will demonstrate in this article.

Because base pair opening is clearly an important process, in the present work, we focused on its description in MD simulations of DNA and RNA and its effect on the structure of nucleic acid fragments. We aimed to investigate whether base fraying may influence the structure and dynamics of some of the most frequently studied DNA and RNA duplexes. Note that the typical length of simulated sequences is rather small, on the order of tens of base pairs. We further discuss how well fraying is described by current force fields and point out some possible imbalances in ff99bsc0 and other variants of the Cornell et al. force field family. In particular, we focus on an erroneous description of the *syn*/*anti* balance relating to the glycosidic torsion angle  $\chi$  and the effect of the  $\chi$  torsion barrier on base fraying.

## 2. METHODS

**Force Fields.** The tested Cornell et al.<sup>28</sup> based force field variants are shown in Table 1. The first variant used for both

**Table 1. Overview of the Cornell et al. Based Force Fields Tested in This Work**

force field description	abbreviation	recommended systems	availability in AMBER
ff99 <sup>41</sup> + bsc0 <sup>38</sup>	ff99bsc0	DNA and RNA	ff10, ff12SB, ff14SB default for DNA
ff99 <sup>41</sup> + bsc0 <sup>38</sup> + $\chi_{OL3}^{39}$ or $\chi_{OL3}$	ff99bsc0 $\chi_{OL3}$	RNA	ff10, ff12SB, ff14SB default for RNA
ff99 <sup>41</sup> + bsc0 <sup>38</sup> + $\chi_{OL4}^{36}$ or $\chi_{OL4}$	ff99bsc0 $\chi_{OL4}$	DNA	optionally in AMBER 14

the DNA and RNA simulations was the well established ff99bsc0 force field, consisting of the ff99<sup>37</sup> force field and the bsc0 correction, which prevents undesirable  $\alpha/\gamma$  transitions in DNA.<sup>38</sup> Further, we tested ff99bsc0 force field variants with two different corrections for the glycosidic angle,  $\chi_{OL3}^{39,35}$  and  $\chi_{OL4}^{36}$ . The  $\chi_{OL3}$  correction by Zgarbova et al.<sup>39,35</sup> was designed to prevent undesirable ladder-like transitions in RNA<sup>40</sup> and was first published in a study testing the performance of various force fields for RNA hairpin simulations.<sup>35</sup> The  $\chi_{OL4}$  correction was developed to improve description of the *syn* region in DNA quadruplexes while not impairing the good performance of ff99bsc0 for the canonical *anti* region.<sup>36</sup>

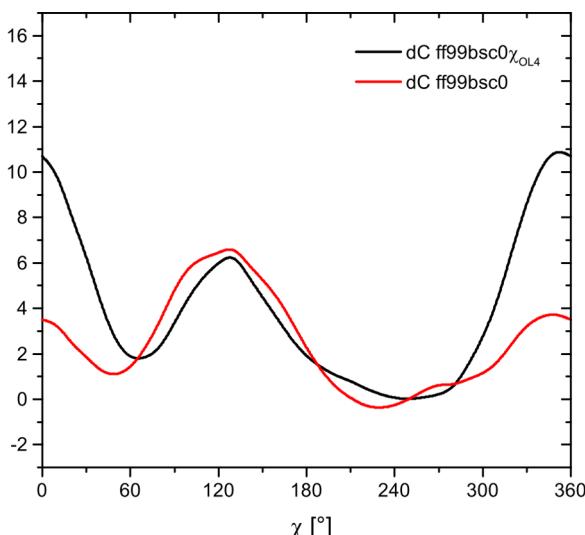
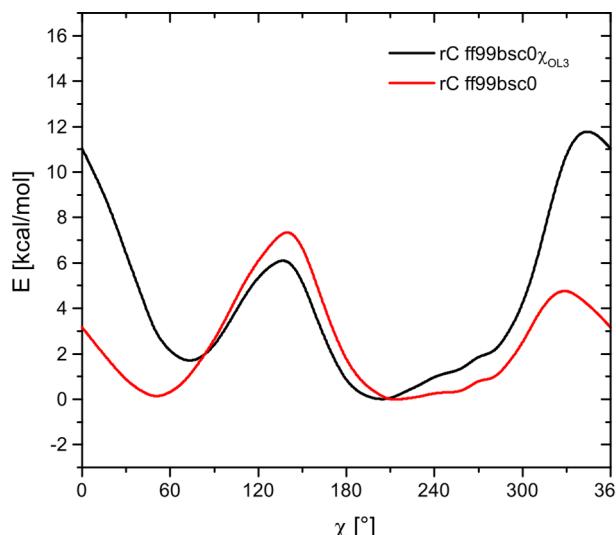
Figure 1 compares the glycosidic torsion profiles for the original ff99  $\chi$  (ff99),  $\chi_{OL3}$  for RNA and  $\chi_{OL4}$  for DNA (the OL profiles correspond to torsions optimized in implicit solvent with partially restrained sugar pucker, see refs 36 and 39). The OL torsions differ from the original force field mainly in the height of the barrier around 0° separating the *anti* and *syn* minima, which should result in a lower rate of *syn/anti* interconversions through this region. Both refinements also provide substantial corrections to the *syn* region and affect the relative energy of the *syn* and *anti* minima. The most obvious is the increased slope of  $\chi_{OL3}$  in the high-anti region.

**MD Simulations.** MD simulations described in this work are listed in Table 2. RNA tetradecamer r(U(AU)<sub>6</sub>A)<sub>2</sub><sup>42</sup> (PDB

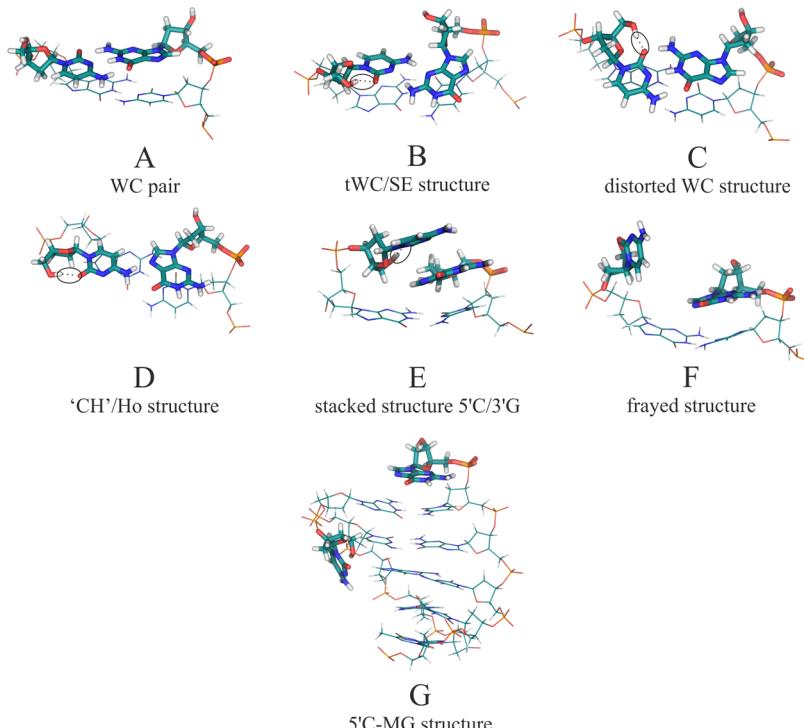
**Table 2. List of MD Simulations Described in This Work**

simulation/FF	solvent and salt	length	terminal steps
B-DNA			
d(CCGGAATTCCGCG) <sub>2</sub> (DD dodecamer)			
ff99bsc0	TIP3P, Na+	10 $\mu$ s	CG//CG
ff99bsc0 $\chi_{OL4}$	TIP3P, Na+	5 $\mu$ s	CG//CG
ff99bsc0	SPC/E, Na+	0.5 $\mu$ s	CG//CG
ff99bsc0	SPC/E, 150 mM KCl	2 $\mu$ s	CG//CG
ff99bsc0 $\chi_{OL4}$	SPC/E, 150 mM KCl	1.5 $\mu$ s	CG//CG
d(TAGCGCTA) <sub>2</sub> (TA//TA end)			
ff99bsc0	TIP3P, Na+	1 $\mu$ s	TA//TA
ff99bsc0 $\chi_{OL4}$	TIP3P, Na+	1 $\mu$ s	TA//TA
d(CCGGA <sub>6</sub> CG) <sub>2</sub> (A-tract, A6)			
ff99bsc0	TIP3P, Na+	1 $\mu$ s	CG//GC
d(CGCATATATGCG) <sub>2</sub> (3AT)			
ff99bsc0	TIP3P, Na+	1 $\mu$ s	CG//GC
d(CCCCCGGGGG) <sub>2</sub> (CSG5)			
ff99bsc0	TIP3P, Na+	1 $\mu$ s	CC//GG
d(GCCAAGAACGAGGAGTGGC) <sub>2</sub> (GC//GC end)			
ff99bsc0	SPC/E, 150 mM KCl	1 $\mu$ s	GC//GC
A-RNA			
r(U(AU) <sub>6</sub> A) <sub>2</sub> (1RNA)			
ff99bsc0	TIP3P, Na+	1 $\mu$ s	UA//UA
ff99bsc0 $\chi_{OL3}$	TIP3P, Na+	1 $\mu$ s	UA//UA
r(GCACCGUUGG) <sub>2</sub> (1QC0')			
ff99bsc0	TIP3P, Na+	1 $\mu$ s	GC//GC, GG//CC
ff99bsc0 $\chi_{OL3}$	TIP3P, Na+	1 $\mu$ s	GC//GC, GG//CC

ID: 1RNA) and DNA dodecamer d(CCGGAATTCCGCG)<sub>2</sub><sup>27</sup> (DD dodecamer, PDB ID: 1BNA) duplexes were taken from X-ray data without any further modifications. The decamer r(GCACCGUUGG)<sub>2</sub> was excised from the 1QC0<sup>43</sup> structure and is hereafter denoted as 1QC0'. Ions and water molecules were removed from the original PDB files. The d-(TAGCGCTA)<sub>2</sub> and d(GCCAAGAACGAGGAGTGGC)<sub>2</sub> duplexes were built using the NAB module of AMBER. Simulations of d(CCCCCGGGGG), d(CCGGA<sub>6</sub>CG)<sub>2</sub>, and d(CGCATATATGCG)<sub>2</sub> have been described elsewhere.<sup>44</sup> In all simulations, the total charge was neutralized by Na<sup>+</sup> ions,<sup>45,46</sup> except for two simulations of the DD dodecamer



**Figure 1.** Comparison of the ff99bsc0 glycosidic torsion profiles calculated in a Poisson–Boltzmann continuum solvent for rC (left): ff99bsc0 (red), ff99bsc0 $\chi_{OL3}$  (or ff10, black) and dC (right): ff99bsc0 (red), ff99bsc0 $\chi_{OL4}$  (black).



**Figure 2.** Snapshots of selected typical terminal base pair conformations from MD simulation of the DD dodecamer. WC terminal pair (A), tWC/SE hydrogen bonded structure (B), distorted WC structure (C), “CH”/Ho structure (D), stacked structure 5’C/3’G (E), frayed structure with C exposed to solvent (F), and structure in which the 5’C adheres to the major groove, 5’C-MG (G). Hydrogen bonds from the 5’-OH group to the O2 atom of the 5’ terminal cytosine are indicated by gray circles.

and simulation of the GC//GC terminated duplex, which were performed with 150 mM KCl using the parameters of Joung and Cheatham<sup>45</sup> (see Table 2). TIP3P<sup>47</sup> water was used to solvate the nucleic acid molecules in rectangular periodic boxes. In simulations with KCl salt excess, octahedral boxes were used. In four simulations, the SPC/E water model<sup>48</sup> was used for comparison (see Table 2). Simulations were carried out with the pmemd.cuda code from the AMBER 11 and AMBER 14 program suites.<sup>49,50</sup> Initial equilibration involved energy minimization of the solute hydrogen atoms, then of the counterions and water molecules. Then, the nucleic acid was frozen, and solvent molecules with counterions were allowed to move during a 10 ps MD run. Subsequently, the nucleobases were allowed to relax in several minimization runs with decreasing force constants (1000, 500, 125, 25 kcal/mol·Å<sup>2</sup>) applied to the backbone phosphate atoms. After full relaxation, the system was slowly heated to 298.15 K over 100 ps using 2 fs time steps and NpT conditions using a weak-coupling scheme with a coupling time of 1 ps. Production simulations were carried under NPT conditions with default temperature and pressure settings ( $\text{taut} = 1.0 \text{ ps}$  and  $\text{taup} = 1.0 \text{ ps}$ ), a 2 fs time step, 9 Å nonbonded cutoff, and SHAKE on bonds to hydrogen atoms with the default tolerance (0.00001). Three simulations of DD dodecamer (ff99bsc0 and ff99bsc0 $\chi_{\text{OL4}}$  simulations in SPC/E water with 150 mM KCl and last 5  $\mu\text{s}$  of the 10  $\mu\text{s}$  ff99bsc0 simulation in TIP3P water) were performed with a time step of 4 fs and hydrogen mass repartitioning.<sup>51</sup> The nonbonded pair list was updated according to the default settings ( $\text{nbflag} = 1$ ,  $\text{skinnb} = 2 \text{ \AA}$ ). PME was used with default grid settings and default tolerance ( $\text{dsum\_tol} = 0.00001$ ). Default scaling factors were used to scale nonbonded and Coulomb interactions ( $\text{scnb} = 2.0$  and  $\text{scee} = 1.2$ , respectively). All analyses were performed using the X3DNA code.<sup>52</sup> Mass-

weighted RMSD values were calculated with respect to the initial structure (all atoms). The MD simulations performed in this work are listed in Table 2.

### 3. RESULTS

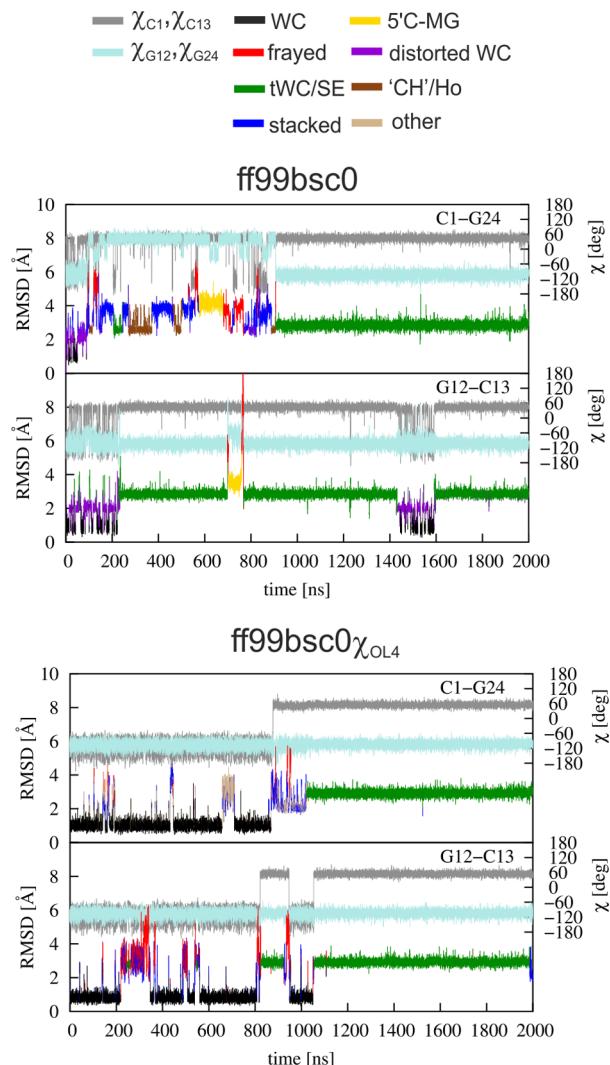
**3.1. Fraying in MD Simulations of DNA.** Owing to their stability, GC pairs are usually placed at the ends of DNA duplexes in MD simulations. Therefore, we focus on GC terminated sequences in this study, starting with the DD dodecamer, which is analyzed in detail, and then continuing with a brief survey of other GC terminated sequences. We also briefly discuss the behavior of one AT terminated DNA sequence. Simulations using the ff99bsc0 force field are compared with ff99bsc0 $\chi_{\text{OL4}}$  simulations.

**GC Terminated DNA Duplexes.** The DD dodecamer is a frequently studied system terminated with the GC base pairs in a CG//CG base pair step. During the 10  $\mu\text{s}$  of ff99bsc0 simulation, the DD dodecamer exhibited multiple opening events and several noncanonical structures formed at both ends of the duplex. Selected important structures (snapshots from the MD simulation) are shown in Figure 2. A canonical Watson–Crick pair is shown in Figure 2A for reference. We denote structures in which at least one terminal base was fully exposed to water as frayed (note, however, that also other structures may be considered as frayed—see Discussion below). For instance, the terminal cytosine in Figure 2F is fully exposed, whereas guanine remains stacked to the duplex. Simultaneous unstacking of both terminal bases was observed rarely and lasted only for a very short time. The very stable hydrogen bonded trans WC/sugar edge structure (tWC/SE, Figure 2B) was first reported by Dršata et al.<sup>34</sup> Another noncanonical hydrogen bonded structure is the C–H edge/

Hoogsteen pair ("CH"/Ho), shown in Figure 2D. As a distinct structure, we also considered the distorted WC pair, in which  $\chi$  of the 5' cytosine is in the *syn* region and the cytosine plane is almost perpendicular to the guanine plane (Figure 2C). Another frequently observed motive had the terminal 5' cytosine stacked above its former WC partner, 3' guanine, which remained approximately in place (Figure 2E). We refer to this and similar structures as "stacked" in the following discussion. Finally, the 5' cytosine may move away from the duplex end and stick to the major groove, where it interacts with the preceding bases at the third and fourth positions from the end (5'C-MG structure, Figure 2G). This structure was significantly populated in some of our other simulations and is important because it significantly influences the helical parameters of base pair steps that are relatively far from the helix end. The geometrical motifs shown in Figure 2 are typical and have often been observed in other MD simulations of GC terminated DNA duplexes. Note that several of the above-described motives (tWC/SE, "CH"/Ho, stacked, distorted WC) are stabilized by a hydrogen bond from the 5'-OH group to the O2 atom of the terminal 5' terminal cytosine.

Figure 3 shows the time evolution of the RMSD of the terminal pairs and also  $\chi$  angles of the terminal G and C bases with the ff99bsc0 force field (top) and with the ff99bsc0 $\chi_{OL4}$  (bottom). Only the first 2  $\mu$ s of the trajectories are shown in Figure 3, and the whole 10  $\mu$ s ff99bsc0 and 5  $\mu$ s ff99bsc0 $\chi_{OL4}$  simulations are given in the Supporting Information, Figures S1 and S2. In the ff99bsc0 simulation, the original WC bonding pattern was preserved for only a short time at the beginning of the simulation, for about 90 ns (C1–G24 end) and 240 ns (G12–C13 end). During this period, we observed frequent visits to the distorted WC structure. However, the terminal pair was able to restore its original WC pairing. After this period, the WC pair was lost, and several relatively stable noncanonical (non-WC bonded) structures emerged, which interconverted between each other until the end of the simulation. By far the most prevalent was the tWC/SE structure (Figure 2B).

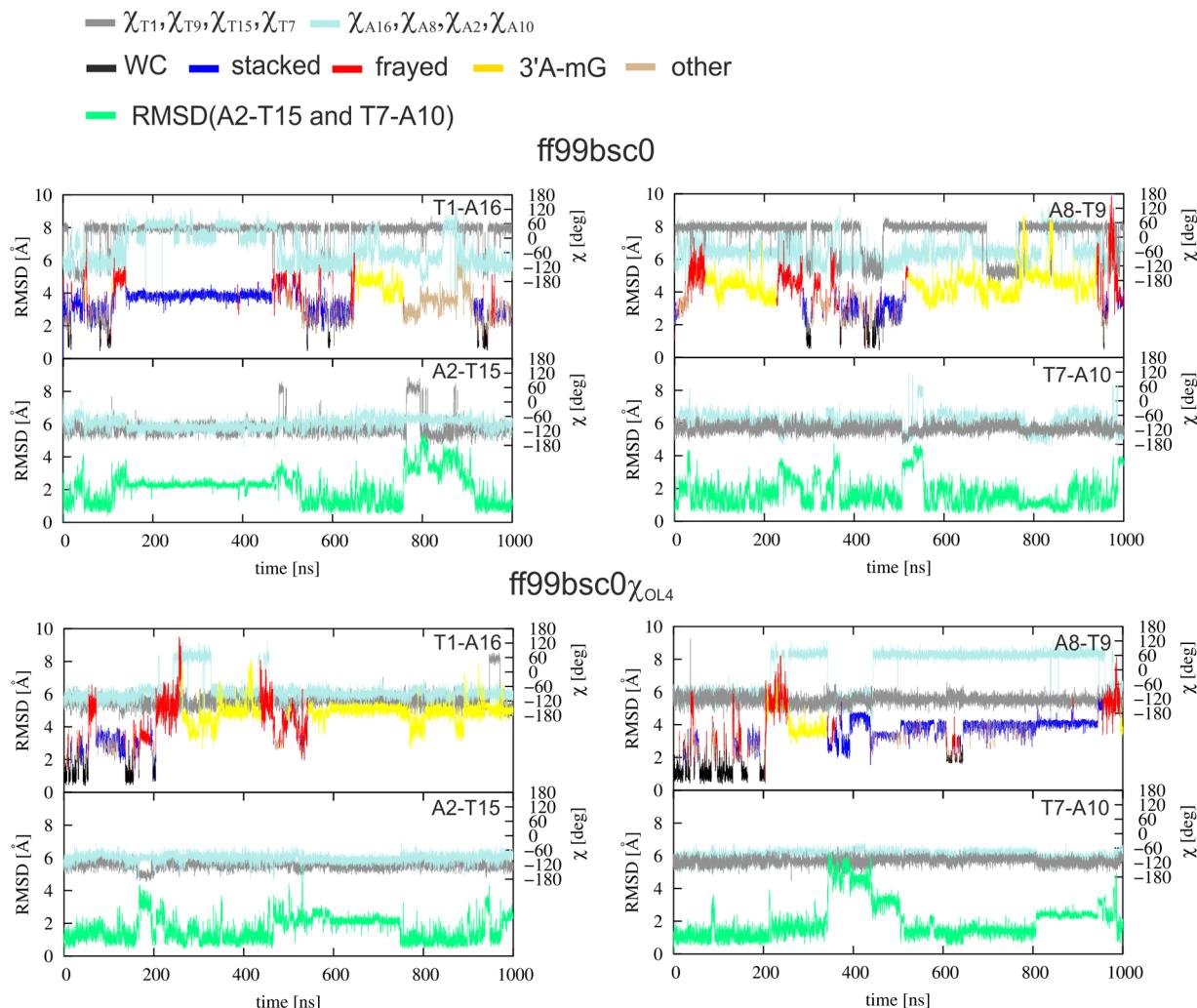
In the ff99bsc0 $\chi_{OL4}$  simulation, there were two notable differences compared to the ff99bsc0 simulation. First, the distorted WC structure was almost not present, and second, the frequency of interconversions between different structures was notably lower. Both these changes can be explained by the changes in the glycosidic dihedral term,  $\chi_{OL4}$  (Figure 1). The lower population of the distorted WC structure can be ascribed to its thermodynamic destabilization due to the  $\chi_{OL4}$  correction. In this structure, the glycosidic angle was typically between 20° and 40°, which is relatively far from the typical *syn* range of around 60°. As discussed in our earlier work,<sup>36</sup> the glycosidic dihedral term in ff99bsc0 incorrectly overstabilized the broad region around  $\chi = 0^\circ$ , which was corrected by the  $\chi_{OL4}$  modification. The slower interconversion rate between different structures can be explained by a significant increase of the barrier height at around  $\chi = 0^\circ$  in the  $\chi_{OL4}$  force field (this barrier is underestimated by ff99bsc0).<sup>39,36</sup> The increased barrier height may lead to a slower interconversion rate between the *syn* and *anti* states, depending on the interconversion path. Indeed, only a few crossings between *syn* and *anti* states can be seen in Figure 3 (bottom). As a consequence, at the start of the simulation, the DD dodecamer remained considerably longer close to the native WC state. However, when switching to the tWC/SE state, the simulation remained "locked" in it, similar to the simulation with ff99bsc0.



**Figure 3.** Fraying of the C1–G24 and G12–C13 terminal base pairs in MD simulation of the DD dodecamer with the ff99bsc0 (top) and ff99bsc0 $\chi_{OL4}$  (bottom) force fields (2  $\mu$ s, TIP3P) characterized by the RMSD of the terminal pairs and torsion angles  $\chi$  for the terminal residues (gray: C1, C13; light blue: G12, G24). The colors in the RMSD plot correspond to different end structures: WC (black), tWC/SE (green), distorted WC (violet), "CH"/Ho (brown), stacked (blue), frayed (red) and 5'C adhered to the major groove, 5'C-MG (yellow), and other (tan).

To extend the time scale of our simulations and improve overall sampling, we performed 1  $\mu$ s ff99bsc0 simulations of two additional DNA sequences terminated with the CG//CG step: d(C G C G (A)<sub>6</sub> C G)<sub>2</sub> (or A<sub>6</sub>) and d-(CGCGATATCGCG)<sub>2</sub> (or 3AT; Figures S3 and S4 in the Supporting Information). In these simulations, the tWC/SE structure was again found to be the most common non-canonical structure, and also other features of fraying were similar to those observed for the DD dodecamer.

**Influence of Neighboring Base Pairs.** All the above-described duplexes were terminated with the CG//CG base pair steps. In order to see whether character of the terminal base pair step has any influence on fraying, we simulated two additional sequences terminated with CC//GG and GC//GC steps, d(CCCCCGGGGG)<sub>2</sub> and d(GCCAAGAACAGGG-AGTGGC)<sub>2</sub>, respectively (1  $\mu$ s in ff99bsc0). Although the observed fraying behavior was in many respects similar to that

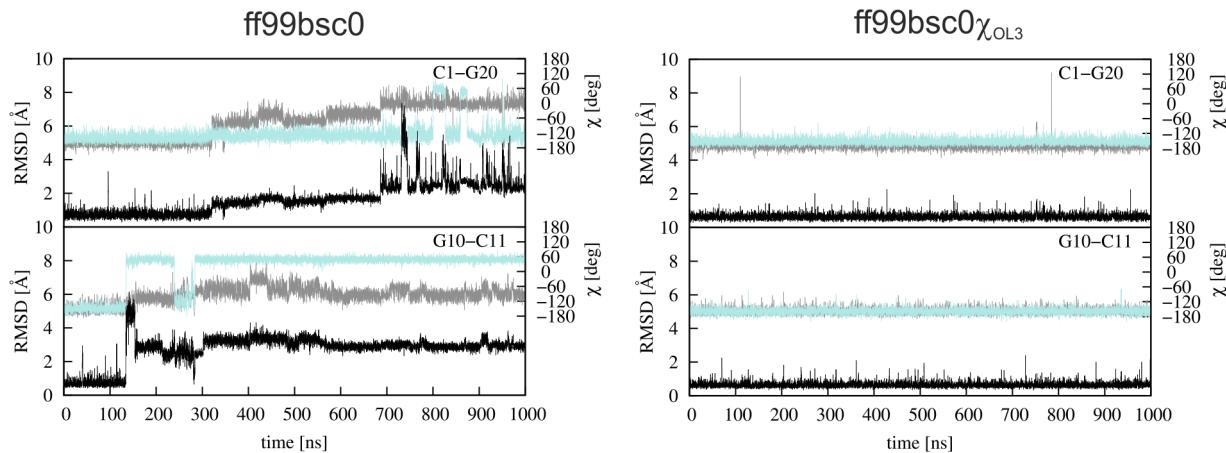


**Figure 4.** Fraying of the last two terminal pairs at each end of the  $d(\text{TAGCGCTA})_2$  duplex with the ff99bsc0 (top) and ff99bsc0 $\chi_{\text{OL4}}$  (bottom) force fields ( $1 \mu\text{s}$ ) characterized by the RMSD of the terminal pairs and torsion angles  $\chi$  of the terminal residues (gray, thymines; light blue, adenines). The colors in the RMSD plot in the upper panels correspond to different end structures: WC (black), stacked (blue), 3' adenine adhered to the minor groove (yellow), frayed (red), and other (tan). The structures in the RMSD plot of the second from the end base pair are not distinguished by color (lower panel, light green).

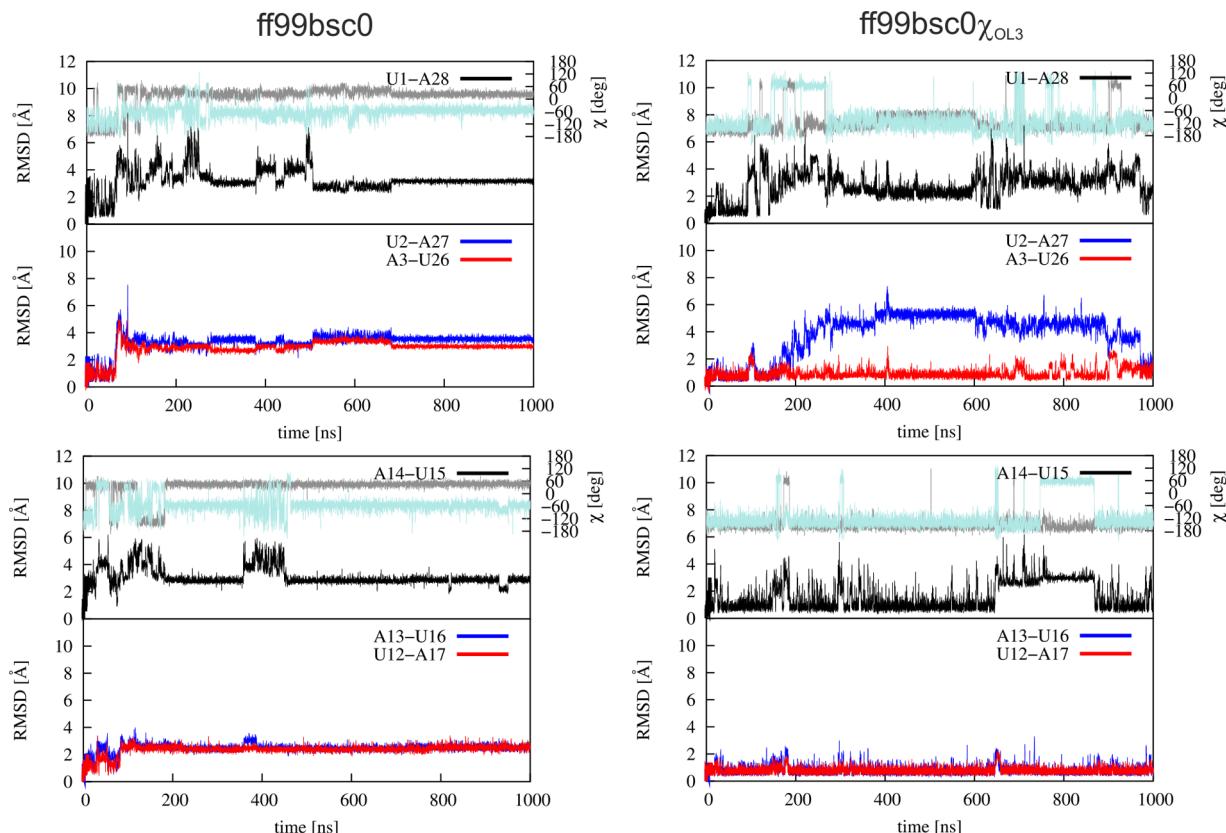
of the CG//CG terminated duplexes, there were also important differences. First, in the CC//GG terminated duplex the distorted WC structure was almost not present, and the terminal GC pair spent most of its time in the canonical WC pairing before it flipped to the very stable tWC/SE structure (Figure S5 in the Supporting Information). It should also be noted that this sequence was in the B-DNA form during the whole simulation. Second, in the GC//GC terminated duplex, we did not see either the distorted WC or the tWC/SE structure, although the groove-bound structures were still present (Figure S6 in the Supporting Information). As a result, the terminal GC pair spent overall more time in the canonical WC pairing, at least in the beginning of the simulation. We saw this favorable behavior also for other GC//GC terminated sequences simulated in our laboratory (results not shown). Note that the GC//GC base pair steps were chosen as termini for 39 Ascona B-DNA sequences containing all 136 unique tetranucleotides.<sup>53</sup> Our observations indicate that not only the terminal base pair but also the sequence context has an important influence on the fraying behavior.

**AT Terminated DNA Duplex.** AT pairs at the helix termini are much more prone to fraying than GC pairs. Therefore, they are only rarely placed at the ends of the helix in MD simulations. Here, we investigated fraying in the  $d(\text{TAGCGCTA})_2$  duplex, whose ends have been suggested by NMR measurements to spend more time in the frayed state than in the WC structure at  $25^\circ\text{C}$  ( $K_d \geq 2.2$ ).<sup>6</sup> Figure 4 shows the RMSD of the last two terminal pairs at each end of the  $d(\text{TAGCGCTA})_2$  duplex and the corresponding  $\chi$  dihedral angles during  $1 \mu\text{s}$  ff99bsc0 and ff99bsc0 $\chi_{\text{OL4}}$  force field simulations.

As expected, the terminal bases frayed very significantly in our ff99bsc0 MD simulations. In addition, the second from the end base pairs exhibited frequent opening events. Because the number of structural motifs that emerged when the two terminal base pairs were broken was quite large, we did not analyze them in detail. In Figure 4, only the canonical WC pairing, frayed, stacked, and groove-bound structures of the terminal pair are distinguished by colors, and all other structural motifs are shown in tan on the RMSD plot. Structures on the RMSD plot of the second from the end AT base pair are not



**Figure 5.** Opening of the G1–C20 and G10–C11 terminal base pairs in MD simulation of 1QC0' with the ff99bsc0 and ff99bsc0 $\chi_{OL3}$  force fields (1  $\mu$ s, TIP3P water model, neutralizing  $\text{Na}^+$ ). Opening is characterized by the RMSD of the terminal pairs (black) and torsion angle  $\chi$  for the terminal residues (gray: terminal guanines G1, G10; light blue: terminal cytosines C11, C20). Different structures on the RMSD plot are not distinguished by color.



**Figure 6.** Opening of the terminal base pairs in MD simulation of 1RNA with the ff99bsc0 (left) and ff99bsc0 $\chi_{OL3}$  (right) force fields (1  $\mu$ s). Opening was characterized by the RMSD of three terminal pairs. Top: U1–A28 (black), U2–A27 (blue), A3–U26 (red). Bottom: A14–U15 (black), A13–U16 (blue), U12–A17 (red), and torsion angle  $\chi$  for the terminal residues (gray: U1, U15; light blue: A28, A15).

distinguished by color (lower panel, light green). During fraying, the 3' adenine often entered the minor groove, where it remained bound for tens of nanoseconds (the most stable noncanonical structure). Stacked structures were the second most populated motif. At least one base was fully exposed to solvent for slightly more than 10% of the simulation time. Occasionally, both of the terminal bases were fully exposed to solvent at the same time. Both terminal bases switched between *syn* and *anti*  $\chi$  regions and for most of the simulation time, at

least one base was in *syn*. Note that although the terminal pair spent most of its time far from the WC structure, it was able to repeatedly re-establish the WC bonding, albeit only for very short periods of time (typically a few nanoseconds).

The second from the end AT pair of the d(TAGCGCTA)<sub>2</sub> duplex oscillated between the canonical WC pairing (between half and one-third of the simulation time) and noncanonical structures, which either were hydrogen bonded (for instance, Hoogsteen/sugar edge AT) or the pair was completely broken.

Nevertheless, when the pair was broken, its bases were usually not fully exposed to water but rather formed quite compact stacked structures involving the terminal bases. Both bases spent the majority of time in the  $\chi = \text{anti}$  region.

The ff99bsc0 $\chi_{\text{OL4}}$  simulation of the d(TAGCGCTA)<sub>2</sub> duplex (Figure 4, bottom) was in many respects similar to the ff99bsc0 simulation. Both ends preferred the minor-groove bound or stacked structures to the native WC pairs. Further, the second AT base pair occasionally broke, but its bases were never fully exposed to solvent. Perhaps the most important difference was that the  $\chi_{\text{OL4}}$  correction led to a smaller number of *syn/anti* interconversions of the glycosidic angle. The population of the  $\chi = \text{syn}$  state also seemed to be smaller than with ff99bsc0. However, this may have been a consequence of slower kinetics of interconversion rather than a thermodynamic result.

**3.2. Fraying in MD Simulations of RNA.** Two RNA duplexes were investigated: a 14 nt r(UU(AU)<sub>5</sub>AA)<sub>2</sub> (PDB code 1RNA) terminated with AU pairs (AA/AU steps), which was expected to fray extensively, and a 10-nt sequence r(GCACCGUUGG)<sub>2</sub>, with two different ends, one with a G<sub>1</sub>C<sub>20</sub> pair in the GC/GC step and the other with a C<sub>11</sub>G<sub>10</sub> pair in the CC/GG step, both of which were expected to be rather stable. Two force field variants were compared for the RNA simulations: ff99bsc0 and ff99bsc0 $\chi_{\text{OL3}}$  with modified glycosidic torsion  $\chi_{\text{OL3}}$ .<sup>39</sup>

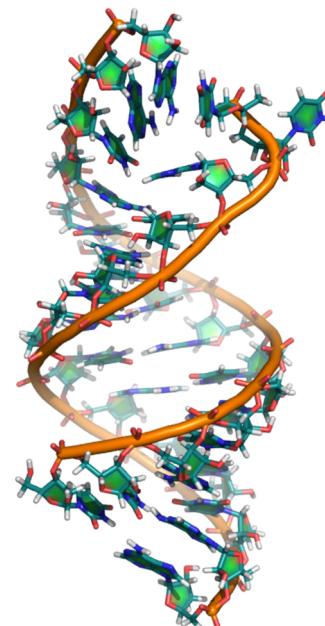
**GC Terminated RNA Duplex.** In the ff99bsc0 simulations (Figure 5), the original WC pairing remained stable for about 300 ns at the G<sub>1</sub>C<sub>20</sub> end and for about 100 ns at the C<sub>11</sub>G<sub>10</sub> end. After that, cytosine on the C<sub>11</sub>G<sub>10</sub> switched from the  $\chi = \text{anti}$  to  $\text{syn}$  glycosidic torsion conformation and formed a stacked structure by stacking above the terminal guanine. Interestingly, in this case, the stacked structure induced large scale distortion of the helix, which extended to four neighboring base pairs. The affected pairs partly lost their WC pairing, increased their propeller, and formed hydrogen bonds to the previous or following bases in the opposite strand. This large scale deformation was reflected by a substantial increase in the overall RMSD of the whole structure. Along with this, we also observed a substantial shift of the  $\chi$  angles of terminal guanine residues toward higher values, corresponding to formation of artificial ladder-like structures, which are a known ff99bsc0 artifact.<sup>40,39</sup> Beyond this point and up to the end of the simulation, the helix was distorted to such an extent that further analysis was meaningless.

In the ff99bsc0 $\chi_{\text{OL3}}$  simulation, the 1QC0' structure was stable for 1  $\mu$ s and did not show any terminal base pair opening (Figure 5). A low RMSD with respect to the 1QC0' X-ray structure and helical parameters measured during simulation showed that the simulated duplex kept close to the initial A-RNA geometry. In this case, the  $\chi_{\text{OL3}}$  correction gave a significant improvement to the ff99bsc0 parameters.

**AU Terminated RNA Duplex.** In the simulation of the r(UU(AU)<sub>5</sub>AA)<sub>2</sub> duplex with the ff99bsc0 force field, fraying and formation of noncanonical motifs occurred within the first tens of nanoseconds of the MD simulation. After about 50 ns, the whole duplex underwent a transition to a ladder-like structure, which remained stable for the rest of the simulation. Although the ladder-like state was considered unphysical, we monitored properties of three terminal base pairs throughout the whole 1  $\mu$ s simulation (Figure 6). The very high RMSD of all three terminal bases at both ends indicated significant structural distortions. A wide variety of noncanonical motifs emerged, which were characterized by significant non-native

stacking interactions and also non-native hydrogen bonds and involved bases of all three terminal base pairs. RMSD values of around 5 Å and higher in part correspond to frayed structures. A representative ladder-like structure with noncanonical terminal arrangement is shown in Figure S7.

The simulation with the ff99bsc0 $\chi_{\text{OL3}}$  force field was clearly more stable than that with the ff99bsc0 force field. Most importantly, we did not see any ladder-like artifacts in this trajectory. While all three terminal base pairs were broken in the ff99bsc0 simulation, only the last two base pairs fluctuated significantly in the corrected  $\chi_{\text{OL3}}$  force field (see Figure 6). Nevertheless, we observed the formation of multiple non-canonical structures by these two base pairs, mainly stacked motifs. For an example of a stacked noncanonical structure, see Figure 7. Full exposure of at least one base to solvent accounted for only a few percent of the simulation time.



**Figure 7.** Snapshot from the ff99bsc0 $\chi_{\text{OL3}}$  simulation of 1RNA.

**Influence of Water Model and Salt Concentration.** Because it is known that the water model can have a significant influence on the predicted structure of biomolecules,<sup>54,55</sup> we decided to also run one simulation of DD dodecamer with ff99bsc0 force field in the SPC/E water model. The results were very similar to those obtained with the TIP3P water model (Figure S8; compare with the results of the simulation of the same sequence in the TIP3P model shown in Figure 3). In the first part of the trajectories, frequent transitions between the WC and distorted WC structures were seen. After that, the tWC/SE structure was formed after some time and remained stable for the rest of the simulation. One difference was that the guanine did not visit the  $\chi = \text{syn}$  region and the corresponding noncanonical structures were not present (see Figure 3, C1–G24 end, cca 100–850 ns). However, because the guanine visits to  $\text{syn}$  seemed to be rare and stochastic, this difference was probably not significant on our time scale. Another factor that may influence structure and dynamics of nucleic acids is salt concentration. Most simulations presented in this work were performed with sodium ions neutralizing the negative charge of the nucleic acid backbone (see Methods). In order to see

whether the presence of salt excess could influence fraying behavior, we performed also two simulations of the DD dodecamer with 150 mM KCl using the parameters of Joung and Cheatham<sup>45</sup> and SPC/E water model (2  $\mu$ s with ff99bsc0 and 1.5  $\mu$ s with ff99bsc0 $\chi_{OL4}$  force field). Also, these simulations were very similar to the above-described simulations with neutralizing sodium ions and the TIP3P water model (Figures S9 and S10; compare with Figure 3). Thus, neither the different water models nor ionic conditions seemed to significantly influence the fraying behavior of the terminal bases.

#### 4. DISCUSSION

**Non-Canonical Structures at Helix Termini.** Formation of well-structured noncanonical conformations is probably the most interesting feature of current MD simulations. For instance, the tWC/SE motif that forms at GC terminated DNA duplexes seems to be even more stable than the canonical GC WC pair, but also the stacked motifs that replace hydrogen bonded pairs are very frequent, mainly in AT terminated sequences. However, such structures are not frequently found in X-ray databases. Instead, the canonical WC structure is by far the most prevalent at the ends of the GC terminated DNA duplexes, and the tWC/SE pair has not been observed. Although in more complex RNA structures, single bases stacked at the duplex ends are present, stacking of the complementary bases from different strands one above the other has not been observed.<sup>21</sup> Moreover, there is no evidence of highly stable noncanonical structures at the helix ends even in NMR experiments. In the case of stacked structures, this is not surprising because most of the motifs observed in our MD simulations are short living and flexible. Therefore, they probably could not be detected on the NMR time scale and thus may be considered as frayed from this point of view. However, the structured tWC/SE conformation is clearly inconsistent with the high resolution NMR structure of the DD dodecamer<sup>56</sup> and the structure obtained with input from large-angle X-ray scattering, which support the presence of WC pairing at DNA duplex ends in solution.<sup>57</sup> Thus, we propose that at least some of the noncanonical structures reported herein represent a force field artifact.

A further indication that the noncanonical structures observed in MD simulations may be spurious stems from the fact that they can be linked to two known problems of empirical force fields. The first is overrestabilization of the *syn* region of the glycosidic torsion potential (see below): most of the noncanonical structures assume the rare  $\chi = \text{syn}$  conformation rather than the common *anti* conformation. The second is possible overestimation of stacking stabilization,<sup>58–60</sup> which could explain the overpopulation of stacked structures. Another issue that we need to bear in mind is formation of a hydrogen bond between the cytosine O2 carbonyl and 5'-OH group of the same residue. This hydrogen bond appears in most of the noncanonical structures at the GC end of a DNA duplex (tWC/SE, "CH"/Ho, stacked, distorted WC). We assume that this bond is formed as a consequence of spatial proximity when  $\chi$  is the *syn* region. However, we still cannot rule out that its strength is overestimated by the force field. More detailed discussion of this bond and a similar hydrogen bond found in guanine quadruplexes can be found in the literature.<sup>34,61–63</sup> In the following discussion, we focus on the glycosidic torsion potential, whose influence was the main aim of this work.

**Glycosidic Torsion and Non-Canonical Structures.** As noted above, the *syn* orientation of the glycosidic torsion was involved in most of the noncanonical structures described for GC terminated DNA, i.e., the tWC/SE CG pair (C is *syn*, G is *anti*), "CH"/Ho pair (both C and G are *syn*), distorted WC pair (C in *syn*), and stacked structure (C in *syn*), and was also frequent in the AT terminated DNA duplex. In the RNA duplexes, the *syn* region was also populated when noncanonical structures were formed. Thus, occurrence of noncanonical structures can be related to the *syn/anti* equilibrium in the force field. This equilibrium is actually a widely discussed topic regarding the Cornell et al. family of force fields. Based on comparison with NMR experiments, it has been argued that the *syn* minimum is overstabilized in the ff99 force field for RNA molecules.<sup>64</sup> We arrived at the same conclusion on the basis of accurate QM calculations that included solvation effects.<sup>39</sup> Therefore, at least for RNA, we believe it is reasonable to assume that an important contribution to the stability of noncanonical structures in the ff99bsc0 (and ff99) force fields stems from inaccurate balancing of the *syn/anti* regions of the glycosidic torsion. Note, however, that the ff99bsc0 and ff99 force fields provide unstable RNA simulations on a longer time scale due to imbalance in the *anti* to high-*anti* region, which has been corrected by  $\chi_{OL3}$ .<sup>35,39</sup> Although the situation is less clear for DNA, it is possible that the stability of the *syn* region is also overestimated by the current force field. Note that the ff99bsc0 parameter set shares the same glycosidic torsion parameters for DNA and RNA molecules. Thus, if *syn* overstabilization occurs in RNA, it may also be present in DNA, especially for cytosine. A guanine *syn* conformation, on the other hand, is common in DNA guanine quadruplexes (G-DNA). *Syn* nucleotides occur in all antiparallel and hybrid G-DNA topologies.<sup>65,66</sup> *Syn* guanines are also stabilized when 5' terminal guanines of the DNA strand are present in the G-stem,<sup>67,68</sup> which has been explained by formation of a 5'-OH...N3(G) terminal H-bond.<sup>61</sup> Recent NMR studies have even suggested minor population of *syn* tetrads in parallel tetrameric quadruplexes with 5'-flanking nucleotides.<sup>69</sup> This, together with the Z-DNA architecture, indicates that the *syn* state of guanine is easily accessible in DNA. G-RNA, on the other hand, always forms all-*anti* G-stem structures.

**Comparison of Glycosidic Torsion Parameterizations.** The glycosidic torsional profile is one of the most important torsions in nucleic acid force fields. In our simulations, we compared the behavior of three glycosidic torsion parameters, the original ff99bsc0, which is identical to ff99 and is used for both DNA and RNA molecules, ff99bsc0 $\chi_{OL3}$  (used for RNA only) and ff99bsc0 $\chi_{OL4}$  (DNA only). Let us first discuss the DNA profiles. As shown in Figure 1, the original ff99 glycosidic torsion differed from the  $\chi_{OL4}$  modification mainly by the height of the barrier around  $\chi = 0^\circ$  and, to a lesser extent, by the shape and relative stability of the *syn* and *anti* minima. The significant increase of the barrier height in our simulations was reflected by a slow *syn/anti* interconversion rate: much fewer transitions were seen with ff99bsc0 $\chi_{OL4}$  (Figure 3, bottom) than with ff99bsc0 (Figure 3, top), especially at the beginning of the simulation. The relative *syn/anti* energy was quite similar in both force fields, and therefore also the populations of the *syn* states. In particular, the tWC/SE pair (with  $\chi = \text{syn}$ ) seemed to be very stable in both force fields. Note that small differences between *syn/anti* populations could not be detected on the scale of a few microseconds because of the very long lifetimes of some of the structures.

Another notable difference between the two  $\chi$  force field variants for DNA was the disappearance of the distorted WC structure in the DD dodecamer with the ff99bsc0 $\chi_{OL4}$  force field. In the distorted WC, the glycosidic angle was found to be between 20 and 40°, which deviates significantly from the typical *syn* range of around 60° and lies close to the barrier of around  $\chi = 0^\circ$ . As discussed in our earlier work,<sup>36</sup> the glycosidic dihedral term in ff99bsc0 incorrectly overstabilizes this region. Also, NMR experiments do not support the distorted structure but rather a canonical WC pair.<sup>70</sup> Therefore, the distorted WC structure should be considered a ff99bsc0 force field artifact, which is removed by the  $\chi_{OL4}$  correction.

Next, we compare ff99bsc0 and ff99bsc0 $\chi_{OL3}$  variants for RNA simulations. Regarding the shape of the profile, ff99bsc0 $\chi_{OL3}$  again showed a much larger barrier at  $\chi = 0^\circ$  (Figure 1). In addition, the *syn* state was significantly destabilized in the ff99bsc0 $\chi_{OL3}$  force field. Noteworthy also is the change in the shape of the *anti* minimum, which has been shown to prevent degradation of RNA structures in ff99bsc0 simulations.<sup>39,35</sup> In the MD simulations with ff99bsc0 $\chi_{OL3}$ , these changes resulted in more stable trajectories, both with the GC and AT terminated RNA duplexes (Figures 5 and 6). In the AT terminated 1RNA sequence, the destabilization of the *syn* region led to lower populations of the *syn* states as compared to the ff99bsc0 simulation. In the GC terminated 1QC0' duplex, the end pairs were so stable in the ff99bsc0 $\chi_{OL3}$  simulation that no opening event was observed during the 1  $\mu$ s time, whereas multiple opening events were observed with the original force field. Thus, we can conclude that that the  $\chi_{OL3}$  parametrization for RNA provides an overall better description of the dynamics of the terminal pairs than the original ff99bsc0 force field.

**Impact of Non-Canonical Structures on the Structure and Dynamics of DNA.** The conformation of the terminal bases in MD simulations is known to affect the geometries of neighboring base pairs. For this reason, the last two terminal base pairs are usually excluded when MD results are compared with experimental data. However, the effects of end structures may extend even further into the helix. A recent detailed study by Dršata et al. has shown that formation of the noncanonical tWC/SE pair may strongly affect helical parameters of not only the neighboring pair but also the more distant second and third neighbors.<sup>34</sup> In addition, populations of the BI/BII states are significantly influenced by the presence of the noncanonical end structures (Figure S7 in the Supporting Information of ref 44). The effect may be even more dramatic in the case of formation of the 5'C-MG structure, in which the 5'C makes contact in the major groove next to the fourth and fifth base pair from the end. Thus, when noncanonical structures with a terminal base bound to the groove appear in the trajectory, it may not be sufficient to just exclude the last two terminal pairs. If the noncanonical structures were real (not artificial), this would represent a significant problem because structural parameters of the duplex ends would converge only very slowly in MD simulations due to the long time necessary to equilibrate the populations of all the long-lived noncanonical structures. From current simulations, it is clear that time scales of a few nanoseconds may not suffice, and therefore much longer simulations are needed, perhaps on the order of hundreds of microseconds. However, if the noncanonical structures are indeed a force field artifact, their influence on the geometry and dynamics of the helices would also be artificial, and consequently even converged time averages would

not represent the real behavior of the duplex. As discussed above, we believe that this is the case in current force fields.

**Kinetics of Fraying in MD Simulations.** Regarding the kinetics of fraying, little is known from experiments about DNA fraying and yet less about RNA. In DNA, one experimental measurement has suggested that fraying occurs on a time scale of hundreds of microseconds for DNA GC terminal pairs.<sup>11</sup> However, this result may have been affected by approximations in the model molecule, i.e., the presence of nucleotide overhangs and attached probe molecules, which have been discussed in the Introduction. If we consider only the more reliable upper and lower limits for fraying obtained from other experiments, we arrive at a range of 40 ns to 1 ms for GC terminated DNA (see Introduction). In our simulations on GC terminated DNA duplexes, we observed several fraying events on the microsecond time scale, corresponding to lifetimes of the end structures at the lower end of the experimental range. Thus, it is possible that fraying kinetics is modeled relatively well by the current force field simulations, at least for GC terminated DNA sequences, where relevant experimental data are available.

**Comparison with NMR and Melting Experiments.** Structural details of the frayed state are not known from the experiment. It is not even clear whether both terminal bases fray at the same time or whether only one frays and the other remains stacked to the duplex. Nevertheless, melting experiments have shown that dangling ends (one nucleotide overhangs) may stabilize the double helix. Furthermore, it has been hypothesized that this stabilization stems from stacking of the dangling base on the double helix.<sup>19,20</sup> Therefore, it is reasonable to assume that one of the terminal bases may remain stacked to the helix during fraying, while the other is exposed to solvent and samples a wide range of conformations. This scenario is consistent with the results of our MD simulations, in which one base frayed frequently but simultaneous fraying of both terminal bases was rare.

Also relative stabilities of the GC and AT (AU) terminal pairs in our MD simulations may be compared with the experiment. However, first, we would like to emphasize that this comparison is only qualitative because our MD populations were far from being converged on the available MD time scales. Another problem is that it is not clear which structures should be considered as frayed. The structures in which at least one base is fully exposed to water would clearly appear as frayed in NMR experiments. Also flexible and short living structures (e.g., stacked and groove-bound structures) would likely appear as frayed in NMR experiments, because they could not be observed on the NMR time scale. However, the very stable tWC/SE structure is likely an artifact, and as such, it should be excluded from the trajectory analysis. If we exclude the tWC/SE parts of the trajectories and consider all other noncanonical structures as frayed, we can see that the GC terminated DNA duplexes fray markedly less than the AT terminated duplexes in both ff99bsc0 and ff99bsc0 $\chi_{OL4}$  force fields. This is in qualitative agreement with NMR experiments, which reported about 12% and 69% of frayed structures in GC and AT terminated duplexes, respectively ( $K_d(GC) = 0.14$  and  $K_d(AT) \geq 2.2^6$ ). A similar trend can be seen also in RNA sequences, although in this case accurate experimental data are missing. Note, however, that behavior of the terminal pairs depends also on their sequence context (see above).

**Convergence of Simulations.** It should be emphasized that populations of noncanonical end structures obtained on

the time scales of about 1  $\mu$ s are far from being converged. Even the longest simulation discussed in this article (10  $\mu$ s of DD dodecamer in ff99bsc0) can provide only a semiquantitative estimate of the relative stability of the most prevalent motifs. Nevertheless, although accurate relative stabilities of the observed noncanonical structures could not be determined, the main conclusions discussed in this article can be drawn from the present data with reasonably good certainty. It should be noted that the objective of this work was to describe artifacts seen in the standard unbiased molecular dynamics simulations on the nowadays typical 1  $\mu$ s time scale. Evaluation of accurate populations of various noncanonical end structures is beyond the scope of this work.

## CONCLUSIONS

Fraying of the terminal base pairs in MD simulations of DNA and RNA duplexes was examined using several Cornell et al. based force field variants: ff99bsc0, ff99bsc0 $\chi_{OL3}$  (ff10), and ff99bsc0 $\chi_{OL4}$ . Simulations on a microsecond time scale showed frequent disruption of the terminal base pair, exposure of the bases to solvent (fraying), and formation of stable noncanonical structures. While some features of fraying seemed to be consistent with the available experimental results, others pointed toward potential problems with the force field description of DNA and RNA molecules.

Only one of the terminal bases was usually exposed to solvent during fraying in our MD simulations; the other one remained stacked to the preceding base pair. This is consistent with thermodynamic data from dangling-end experiments, which have shown that a base stacked to one end of the helix (dangling end) may provide additional thermodynamic stabilization to the duplex. Therefore, fraying of one base is expected to be thermodynamically more favorable than simultaneous fraying of both terminal bases. Consistent with the experiment, we observed that GC terminal base pairs were significantly more stable than the AT/AU terminal pairs, both in DNA and RNA simulations.

Multiple noncanonical structures were formed at duplex termini, usually within the first tens to hundreds of nanoseconds of the MD simulations. Among these structures were the tWC/sugar edge pair, CH edge/Watson–Crick pair, and stacked geometries, in which the terminal bases were stacked above each other. Some of these structures seemed to be more stable than the canonical WC pairing, namely the tWC/SE and stacked structures, whose lifetimes were on the order of hundreds of nanoseconds. Because X-ray and NMR experimental data support WC pairing at the helix termini, we consider the significant population of the noncanonical structures to be a force field artifact.

The noncanonical structures represent a considerable problem for the modeling of duplex termini. First, their formation affects the structure of base pair steps in their vicinity (end effects). The two neighboring base pair steps are affected substantially and even steps further away may be deformed if the frayed terminal base forms a contact in the major groove. Second, the relatively slow equilibration associated with the long life times of noncanonical structures complicates the convergence of the structural and helical parameters. For this reason, convergence of the structural parameters is not achievable within the currently available time scales of several microseconds.

The simulations with different force field variants showed that the appearance of noncanonical structures may at least

partly be ascribed to inaccurate glycosidic ( $\chi$ ) torsion parametrization. The ff99bsc0 parametrization appears to overstabilize the rare  $\chi = syn$  state with respect to the dominant *anti* conformation. In addition, it underestimates the barrier height at around  $\chi = 0^\circ$ , thus allowing rapid interconversions between *anti* and *syn* regions. However, other force field imbalances, namely overestimation of the stacking energy between nucleic bases, may also contribute to formation of noncanonical structures. The refined glycosidic torsion potential  $\chi_{OL3}$  provides an improved description and substantially more stable MD simulations of RNA molecules ( $\chi_{OL3}$  is currently incorporated in the ff10 and f12SB default force fields for RNA simulations in AMBER). In the case of DNA, the  $\chi_{OL4}$  correction yields only partial improvement. It eliminates the distorted WC structure and reduces the rate of *anti/syn* interconversions, thus delaying formation of stable noncanonical structures from tens of nanoseconds in ff99bsc0 to hundreds of nanoseconds. None of the tested force fields provided a fully satisfactory description of the terminal regions, indicating that further improvement is needed to achieve realistic DNA and RNA MD simulations on microsecond time scales.

## ASSOCIATED CONTENT

### S Supporting Information

RMSD plots of simulations of A6, 3AT, 5C5G, and d(GCCAAGAACGGAGTGGC)<sub>2</sub> sequences and RMSD plots of additional simulations of DD with different water models and salt concentrations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: petr.jurecka@upol.cz.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the Operational Program Research and Development for Innovations of the European Regional Development Fund (CZ.1.05/2.1.00/03.0058), administered by the Ministry of Education, Youth and Sports of the Czech Republic, by the Grant Agency of the Czech Republic, project 14-29874P (M.Z.), and by grant no. RVO61388963 from the Academy of Sciences of the Czech Republic (F.L.).

## REFERENCES

- (1) Wong, K. Y.; Pettitt, B. M. *Biophys. J.* **2008**, 95 (12), 5618–5626.
- (2) Perez, A.; Orozco, M. *Angew. Chem., Int. Ed. Engl.* **2010**, 49 (28), 4805–4808.
- (3) Kornberg, A.; Baker, T. A. *DNA Replication*, Second ed.; University Science Books: Sausalito, CA, 2005.
- (4) Colizzi, F.; Bussi, G. *J. Am. Chem. Soc.* **2012**, 134 (11), 5173–5179.
- (5) Katz, R. A.; Merkel, G.; Andrasek, M. D.; Roder, H.; Skalka, A. M. *J. Biol. Chem.* **2011**, 286 (29), 25710–25718.
- (6) Nonin, S.; Leroy, J. L.; Gueron, M. *Biochemistry* **1995**, 34 (33), 10652–10659.
- (7) Kochyan, M.; Lancelot, G.; Leroy, J. L. *Nucleic Acids Res.* **1988**, 16 (15), 7685–7702.
- (8) Leroy, J. L.; Kochyan, M.; Huynh-Dinh, T.; Gueron, M. *J. Mol. Biol.* **1988**, 200 (2), 223–238.

- (9) Andreatta, D.; Sen, S.; Perez Lustres, J. L.; Kovalenko, S. A.; Ernsting, N. P.; Murphy, C. J.; Coleman, R. S.; Berg, M. A. *J. Am. Chem. Soc.* **2006**, *128* (21), 6885–6892.
- (10) Furse, K. E.; Corcelli, S. A. *J. Phys. Chem. B* **2010**, *114* (30), 9934–9945.
- (11) Egelle, C.; Schaub, E.; Piemont, E.; de Rocquigny, H.; Mely, Y. *C. R. Biol.* **2005**, *328* (12), 1041–1051.
- (12) Nikolova, E. N.; Al-Hashimi, H. M. *J. Biomol. NMR* **2009**, *45* (1–2), 9–16.
- (13) Nikolova, E. N.; Bascom, G. D.; Andricioaei, I.; Al-Hashimi, H. M. *Biochemistry* **2012**, *51* (43), 8654–8664.
- (14) Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37* (42), 14719–14735.
- (15) SantaLucia, J., Jr. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (4), 1460–1465.
- (16) Bommarito, S.; Peyret, N.; SantaLucia, J., Jr. *Nucleic Acids Res.* **2000**, *28* (9), 1929–1934.
- (17) Freier, S. M.; Kierzek, R.; Jaeger, J. A.; Sugimoto, N.; Caruthers, M. H.; Neilson, T.; Turner, D. H. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83* (24), 9373–9377.
- (18) Liu, J. D.; Zhao, L.; Xia, T. *Biochemistry* **2008**, *47* (22), 5962–5975.
- (19) Isaksson, J.; Chattopadhyaya, J. *Biochemistry* **2005**, *44* (14), 5390–5401.
- (20) Burkard, M. E.; Kierzek, R.; Turner, D. H. *J. Mol. Biol.* **1999**, *290* (5), 967–982.
- (21) Mohan, S.; Hsiao, C.; Vandeven, H.; Gallagher, R.; Krohn, E.; Kalahar, B.; Wartell, R. M.; Williams, L. D. *J. Phys. Chem. B* **2009**, *7*, 2614–2623.
- (22) Cheatham, T. E., 3rd; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435–471.
- (23) Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (24), 13922–13927.
- (24) Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2008**, *36* (7), 2379–2394.
- (25) Ma, H.; Wan, C.; Wu, A.; Zewail, A. H. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (3), 712–716.
- (26) Jung, J.; Van Orden, A. *J. Am. Chem. Soc.* **2006**, *128* (4), 1240–1249.
- (27) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U. S. A. (Biol. Sci.)* **1981**, *78* (4), 2179–2183.
- (28) 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. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.
- (29) Kannan, S.; Zacharias, M. *Biophys. J.* **2007**, *93* (9), 3218–3228.
- (30) Lin, M. M.; Meinhold, L.; Shorokhov, D.; Zewail, A. H. *Phys. Chem. Chem. Phys.* **2008**, *10* (29), 4227–4239.
- (31) Kuhrova, P.; Banas, P.; Best, R. B.; Sponer, J.; Otyepka, M. *J. Biomol. Struct. Dyn.* **2013**, *31*, 25–26.
- (32) Priyakumar, U. D.; MacKerell, A. D., Jr. *Chem. Rev. (Washington, DC, U. S.)* **2006**, *106* (2), 489–505.
- (33) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614.
- (34) Drsata, T.; Perez, A.; Orozco, M.; Morozov, A. V.; Sponer, J.; Lankas, F. *J. Chem. Theory Comput.* **2013**, *9* (1), 707–721.
- (35) Banáš, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T.; Sponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6* (12), 3836–3849.
- (36) Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T.; Jurecka, P.; Sponer, J. *J. Chem. Theory Comput.* **2012**, *8*, 2506–2520.
- (37) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16* (4), 845–862.
- (38) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92* (11), 3817–3829.
- (39) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902.
- (40) Mlynšky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2010**, *114* (19), 6642–6652.
- (41) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21* (12), 1049–1074.
- (42) Dockbregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; Debear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. *J. Mol. Biol.* **1989**, *209* (3), 459–474.
- (43) Klosterman, P. S.; Shah, S. A.; Steitz, T. A. *Biochemistry* **1999**, *38* (45), 14784–14792.
- (44) Zgarbova, M.; Luque, F. J.; Sponer, J.; Cheatham, T. E.; Otyepka, M.; Jurecka, P. *J. Chem. Theory Comput.* **2013**, *9* (5), 2339–2354.
- (45) Joung, I. S.; Cheatham, T. E. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041.
- (46) Joung, I. S.; Cheatham, T. E. *J. Phys. Chem. B* **2009**, *113* (40), 13279–13290.
- (47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
- (48) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.
- (49) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. E.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; 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.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, V.; Kovalenko, A.; Kollmann, P. A. *AMBER 11*; University of California: San Francisco, 2010.
- (50) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14*; University of California: San Francisco, 2014.
- (51) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20* (8), 786–798.
- (52) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31* (17), 5108–5121.
- (53) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., 3rd; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. *Nucleic Acids Res.* **2010**, *38* (1), 299–313.
- (54) Kührová, P.; Otyepka, M.; Šponer, J.; Banáš, P. *J. Chem. Theory Comput.* **2013**, *10*, 401–411.
- (55) Florova, P.; Sklenovsky, P.; Banas, P.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6* (11), 3569–3579.
- (56) Wu, Z. G.; Delaglio, F.; Tjandra, N.; Zhurkin, V. B.; Bax, A. *J. Biomol. NMR* **2003**, *26* (4), 297–315.
- (57) Schwieters, C. D.; Clore, G. M. *Biochemistry* **2007**, *46* (5), 1152–1166.
- (58) Zgarbova, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. *Phys. Chem. Chem. Phys.* **2010**, *12* (35), 10476–10493.
- (59) Banas, P.; Mladek, A.; Otyepka, M.; Zgarbova, M.; Jurecka, P.; Svozil, D.; Lankas, F.; Sponer, J. *J. Chem. Theory Comput.* **2012**, *8* (7), 2448–2460.
- (60) Chen, A. A.; Garcia, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (42), 16820–16825.
- (61) Cang, X. H.; Sponer, J.; Cheatham, T. E. *Nucleic Acids Res.* **2011**, *39* (10), 4499–4512.

- (62) Sponer, J.; Cang, X. H.; Cheatham, T. E. *Methods* **2012**, *57* (1), 25–39.
- (63) Sponer, J.; Mladek, A.; Spackova, N.; Cang, X. H.; Cheatham, T. E.; Grimme, S. *J. Am. Chem. Soc.* **2013**, *135* (26), 9785–9796.
- (64) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. *J. Chem. Theory Comput.* **2010**, *6* (5), 1520–1531.
- (65) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415.
- (66) Karsisiotis, A. I.; O’Kane, C.; Webba da Silva, M. *Methods* **2013**, *64* (1), 28–35.
- (67) Clark, G. R.; Pytel, P. D.; Squire, C. J. *Nucleic Acids Res.* **2012**, *40* (12), 5731–5738.
- (68) Tong, X.; Lan, W.; Zhang, X.; Wu, H.; Liu, M.; Cao, C. *Nucleic Acids Res.* **2011**, *39* (15), 6753–6763.
- (69) Sket, P.; Virgilio, A.; Esposito, V.; Galeone, A.; Plavec, J. *Nucleic Acids Res.* **2012**, *40* (21), 11047–11057.
- (70) Tjandra, N.; Tate, S.; Ono, A.; Kainoshio, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122* (26), 6190–6200.