

# Development of Force Field Parameters for the Simulation of Single- and Double-Stranded DNA Molecules and DNA–Protein Complexes

Maxwell R. Tucker, Stefano Piana,\* Dazhi Tan, Michael V. LeVine, and David E. Shaw\*

 Cite This: *J. Phys. Chem. B* 2022, 126, 4442–4457

 Read Online

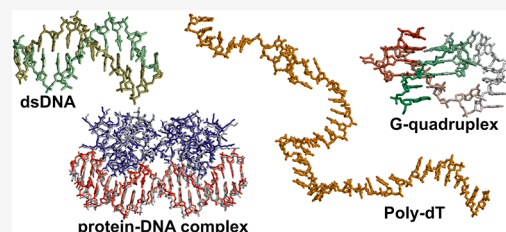
ACCESS |

 Metrics & More

 Article Recommendations

 Supporting Information

**ABSTRACT:** Although molecular dynamics (MD) simulations have been used extensively to study the structural dynamics of proteins, the role of MD simulation in studies of nucleic acid based systems has been more limited. One contributing factor to this disparity is the historically lower level of accuracy of the physical models used in such simulations to describe interactions involving nucleic acids. By modifying nonbonded and torsion parameters of a force field from the Amber family of models, we recently developed force field parameters for RNA that achieve a level of accuracy comparable to that of state-of-the-art protein force fields. Here we report force field parameters for DNA, which we developed by transferring nonbonded parameters from our recently reported RNA force field and making subsequent adjustments to torsion parameters. We have also modified the backbone charges in both the RNA and DNA parameter sets to make the treatment of electrostatics compatible with our recently developed variant of the Amber protein and ion force field. We name the force field resulting from the union of these three parameter sets (the new DNA parameters, the revised RNA parameters, and the existing protein and ion parameters) *DES-Amber*. Extensive testing of *DES-Amber* indicates that it can describe the thermal stability and conformational flexibility of single- and double-stranded DNA systems with a level of accuracy comparable to or, especially for disordered systems, exceeding that of state-of-the-art nucleic acid force fields. Finally, we show that, in certain favorable cases, *DES-Amber* can be used for long-timescale simulations of protein–nucleic acid complexes.



## INTRODUCTION

Molecular dynamics (MD) simulations have frequently been used to characterize the structural dynamics of proteins. The application of MD simulation to nucleic acid based systems, however, has received considerably less attention, and as a consequence, the energy functions and associated parameterizations (“force fields”) used in simulating nucleic acids have been subject to less development and validation than corresponding protein force fields. In recent years, however, there have been substantial improvements to both the Amber<sup>1–4</sup> and CHARMM<sup>5</sup> family of DNA force fields, and an entirely new DNA force field based on QM data has recently been proposed to complement existing models.<sup>6</sup> In the case of RNA, we recently revised the torsion and nonbonded parameters of an Amber RNA force field<sup>7</sup> with an emphasis on obtaining a force field that could, with an accuracy comparable to that of state-of-the-art protein force fields, describe both single- and double-stranded RNA structures.

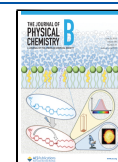
Here, we describe a revision of our previously published force field parameters for RNA aimed at achieving for DNA a similar level of accuracy as was obtained for simulations of single- and double-stranded RNA oligomers.<sup>7</sup> As RNA and DNA are chemically very similar, we postulated that previously adjusted parameters for RNA should be largely transferable to DNA with a similar effect. Indeed, we found that after transferring the modified RNA nonbonded parameters to the

DNA force field, only small adjustments to a few torsion parameters were required to make the force field suitable for DNA simulations. We also developed new parameters for the magnesium ion, a cofactor shown to often be critical for the stability of tertiary structures of nucleic acids, and for structural zinc ions, which are often found in protein motifs that bind to nucleic acids. The revised force field parameters were tested by comparing results from simulations of single-stranded (ssDNA) and double-stranded DNA (dsDNA) molecules (known to form a number of different secondary structures) to experiments. We found that the structure, dynamics, and thermodynamics of these systems were described with a level of accuracy comparable to that of state-of-the-art nucleic acid force fields. In particular, we found that this force field gives an excellent description of ssDNA structures while still providing a reasonably accurate description of dsDNA systems, with the notable exception that *DES-Amber* underestimates the relative stability of the BI/BII conformations.

**Received:** December 31, 2021

**Revised:** March 21, 2022

**Published:** June 12, 2022



**Table 1. Revised Torsion Parameters Obtained from a Fit to Torsion Energy Profiles Calculated at the MP2 Level of Theory<sup>a</sup>**

angle	$\phi_1$	$k_1$	$\phi_2$	$k_2$	$\phi_3$	$k_3$	$\phi_4$	$k_4$
$\chi_A$	-7.75	-0.5250	-4.44	1.9578	-35.39	-0.9569	-21.36	0.2058
$\chi_C$	138.56	-1.0002	-31.94	2.8162	34.57	-1.1862	-53.36	0.7164
$\chi_T$	-7.14	-0.1027	-2.45	2.2160	-13.96	-0.9599	-123.92	-0.2482
$\chi_G$	6.19	-0.5093	-0.54	1.7381	-29.00	-0.8156	-2.69	0.2408
$\zeta$	-154.60	-0.2379	-4.09	1.6200	-51.37	0.4252		

<sup>a</sup>Torsion angle potentials,  $V_{\phi}$ , are expressed as  $V_{\phi} = \sum_{n=1}^4 k_n \cos(n\phi - \phi_n)$ . The  $\chi$  torsion angle parameters were made nucleotide-specific (subscripts A, C, T, and G), while the same set of  $\zeta$  torsion angle parameters was used for all nucleotides. Units are degrees for angles and kcal mol<sup>-1</sup> for force constants.

A second, equally important aim of this work was to make our nucleic acid force field parameters compatible with our latest protein force field parameterization, which provides state-of-the-art accuracy.<sup>8</sup> In this protein force field, the effect of polarization on the dielectric environment is modeled by an empirical scaling and redistribution of the net charges for ionic residues. The same approach has been followed here to develop the DNA and RNA electrostatic parameters so that these force fields can be used in combination with the protein force field. We name the force field that includes all of these parameters (the new DNA parameters, the revised RNA parameters, and the existing protein and ion parameters) *DES-Amber*. (We note that, in our previous work,<sup>8</sup> “*DES-Amber*” referred to a force field consisting solely of the protein parameters.) Subsequent testing of *DES-Amber* in simulations of protein–DNA and protein–RNA complexes gave mixed results, yielding an accurate reproduction of the structures of some, but not all, of the complexes investigated.

Similar mixed results had been previously obtained for simulations of protein–protein complexes using the *DES-Amber* protein parameters.<sup>8</sup> A reparameterization of the phosphate backbone nonbonded interactions was attempted; this parameterization, described in detail in the [Supporting Information \(SI\)](#), involved the optimization of the phosphate–water interaction by modifications of the Lennard–Jones (LJ) pair potential followed by torsion refitting based on NMR data from short ssDNA fragments. The distribution of charges on the O2′ hydroxyl group and the aromatic carbon–water interactions were also optimized to reproduce QM data and the osmotic coefficient of purine in water. The resulting force field, named *DES-Amber 3.20* (from the value of the LJ  $\sigma$  parameter of the interaction between water and the phosphate oxygen), greatly improved the description of protein–nucleic acid complexes, though we could not achieve a consistent improvement across all systems. Although further parameter optimization may result in more consistent improvement, the simple, nonpolarizable functional form used to describe the nonbonded interactions may limit the ability of the force field to simultaneously describe with a high level of accuracy isolated proteins and nucleic acids, as well as protein–protein and protein–nucleic acid complexes, and it is possible that more sophisticated, polarizable force fields (which are outside the intended scope of this work) may be required to address this shortcoming.

## METHODS

**Ab Initio Calculations.** The previously introduced *DES-Amber* force field for RNA<sup>7</sup> is a modification of the Amber nucleic acid force field.<sup>1,9</sup> In this earlier work, nucleobase charges and LJ parameters were developed based on QM calculations of nucleobase stacking and pairing, and QM

torsion scans were used to develop new torsion parameters for the backbone and glycosidic torsions. For the development of a DNA force field, nucleobase nonbonded parameters are expected to be the same as in RNA, but the  $\chi$  (O4′-C1′-N9-C4 in purines and O4′-C1′-N1-C6 in pyrimidines) torsion and the  $\zeta$  (C3′-O3′-P′-O5′) torsion are expected to be affected to some extent by the absence of the OH2′ group. New torsion potentials were thus determined following our previous work on RNA.<sup>7</sup> In vacuo potential energy surface (PES) scans of the  $\chi$  torsion were performed for the four ribonucleosides. Torsion scans for the  $\zeta$  torsion were performed on a model molecule mimicking the DNA backbone. QM energies were calculated at the MP2 level of theory, using local and density-fitting approximations<sup>10</sup> with an augmented double-zeta basis set (aug-cc-pVDZ) and the COSMO model<sup>11</sup> to account for solvation effects, using the MOLPRO program.<sup>12</sup> The torsions were scanned between -180 and 180° in 10° increments. Each structure was relaxed at the MP2 level with a number of other dihedral angles constrained to prevent the formation of intramolecular hydrogen bonds, thus ensuring the smoothness of the potential energy surfaces. Only one-dimensional scans were performed, and the coupling between dihedral angles<sup>13</sup> was not explicitly considered. The total number of terms,  $N$ , in the cosine expansion was four for the  $\chi$  torsion and three for the  $\zeta$  torsions. The new  $k_n$  and  $\phi_n$  values are summarized in [Table 1](#).

**Assignment of Charges to the Terminal Residues.** In the Amber force field, the -1.0 charge of the 3′-terminal nucleobase, which is often phosphorylated, is split between the 3′-terminal residue (-0.6921) and the 5′-terminal residue (-0.3079). This charge distribution is the result of where the residue boundaries are defined in the Amber force field and does not follow the common approach of assigning integer charges to each residue in most fixed-charge force fields for biomolecules. We maintained it in our reparameterization for consistency with the original Amber force field. We also tested a force field (which we term *DES-Amber T*) in which the charge of the terminal groups was adjusted to localize all of the negative charge on the 3′-terminal nucleotide. For this force field, we performed simulations of short ssDNA fragments of two to six nucleotides (see the subsection Simulated Tempering Simulations of Short ssDNA Molecules below). Although we found some differences in the conformational ensembles sampled by simulations performed using these different charge distributions, both matched the experimental J-couplings well (RMSD <1 Hz, [Table S1](#)), suggesting that both charge distributions are reasonable choices.

**MD Simulations of dT<sub>40</sub>.** We performed MD simulations of dT<sub>40</sub> molecules starting from a structure in which all residues were in the B-DNA conformation. This structure was solvated in cubic 130<sup>3</sup> Å<sup>3</sup> water boxes containing 0.025, 0.1,

0.4, or 1.0 M NaCl. At each NaCl concentration, 100  $\mu$ s of simulation at 300 K and 1 atm in the NPT ensemble was performed using the DES-Amber and the DES-Amber SF1.0 force fields. For comparison, simulations of 100  $\mu$ s were also run using the Amber-bsc1 force field,<sup>1</sup> and simulations of 40  $\mu$ s were run using the CHARMM36 force field.<sup>5</sup> The persistence length  $L_p$  was calculated from  $\langle \cos \theta_{ij} \rangle = e^{-L|\mathbf{i}-\mathbf{j}|/L_p}$ , where, for a polymer formed by a chain of vectors of the same length  $L$ ,  $\theta_{ij}$  is the angle between two unit vectors  $i$  and  $j$  and  $|\mathbf{i}-\mathbf{j}|$  is their distance in sequence. The end-to-end distance was measured and FRET intensities were calculated assuming a Förster radius of 60 Å.<sup>14</sup>

**Simulated Tempering Simulations of Short ssDNA Molecules.** We performed simulated tempering simulations of 18 ssDNA molecules ranging from two to six nucleotides in length (Table S1). Each system was generated using Maestro<sup>15</sup> and solvated in a cubic 50<sup>3</sup> Å<sup>3</sup> box containing 0.1 M NaCl. Simulations were performed using the Amber-bsc1 force field<sup>1</sup> and the TIP3P water model,<sup>16</sup> the DES-Amber force field, the DES-Amber SF1.0 force field, and the DES-Amber T force field. The TIP4P-D water model<sup>17</sup> was used in all DES-Amber simulations. The systems were first equilibrated at 300 K for 100 ps using the Desmond software,<sup>18</sup> and production simulated tempering simulations<sup>19</sup> of 20  $\mu$ s were performed in the NPT ensemble<sup>20–22</sup> with 100 rungs spanning the 273–410 K range using Anton.<sup>23</sup> NMR J-couplings were calculated for the sugar ring protons using the Karplus relation as parameterized by Haasnoot<sup>24,25</sup> and compared to the J-couplings measured experimentally.<sup>26–35</sup>

**Simulated Tempering Simulations of the d-(GCGAAGC) Polynucleotide.** We performed simulated tempering simulations of the d(GCGAAGC) polynucleotide that experimentally folds into a hairpin structure. We solvated the experimentally determined structure (PDB entry 1kr8)<sup>36</sup> in a cubic 50<sup>3</sup> Å<sup>3</sup> box of TIP4P-D water containing 1.0 M NaCl. Simulations were performed using the DES-Amber force field and the DES-Amber SF1.0 force field. The systems were first equilibrated at 300 K for 100 ps using the Desmond software, and production simulated tempering simulations of 170  $\mu$ s were performed in the NPT ensemble with 100 rungs spanning the 278–410 K range using Anton. The free energy of the hairpin at 310 K was calculated as

$\Delta G_{310\text{K}} = -k_b T \ln \left( \frac{\phi_{310\text{K}}}{1 - \phi_{310\text{K}}} \right)$ , where  $\phi_{310\text{K}}$  is the fraction of frames where the hairpin was folded at 310 K, computed using a dual-cutoff<sup>37,38</sup> of 1 and 6 Å on the heavy-atom RMSD from the experimental structure, averaged over 50 ns.

**Simulated Tempering Simulations of DNA Duplex Formation.** We performed simulated tempering simulations of seven complementary DNA duplexes of sequence d-(CACAG), d(CGCGG), d(CGTACG), d(GAGTGAG), d(GCGC), d(GGATCC), and d(TCATGA). The double-stranded models of the seven DNA duplexes were generated with the program COOT.<sup>39</sup> Each duplex was solvated in a cubic 50<sup>3</sup> Å<sup>3</sup> box containing 0.1 M salt (Table S2). Simulations were performed using the Amber-bsc1 force field with the TIP3P water model, the DES-Amber force field, and the DES-Amber SF1.0 force field. The TIP4P-D water model was used in all DES-Amber simulations. The systems were first equilibrated at 300 K for 100 ps using the Desmond software, and production simulated tempering simulations of 500  $\mu$ s were performed in the NPT ensemble with 100 rungs spanning

the 278–410 K range using Anton. The fraction of frames containing B-DNA at each temperature ( $\phi$ ) was computed using a dual cutoff of 2 and 10 Å on the backbone RMSD from the ideal B-DNA conformation, averaged over 50 ns. The free energy of the B-DNA duplex at 310 K was calculated as

$\Delta G_{310\text{K}} = -k_b T \ln \left( \frac{2\phi_{310\text{K}}}{c(1 - \phi_{310\text{K}})^2} \right)$ , where  $c$  is the DNA concentration in the simulation box. The “experimental” free energy was estimated using the nearest-neighbor model of SantaLucia<sup>40</sup> with  $[\text{Na}^+] = 0.1$  M.

**MD Simulations of DNA Oligomers.** We performed simulations of 17 different sequences of lengths ranging from 10 to 17 amino acids (Table S4). The starting structure of each simulation was taken from the experimental structures deposited in the PDB (Table S4).<sup>41–57</sup> Three simulations were performed for the Drew–Dickerson dodecamer (DDD) starting from three different experimental structures. These simulations gave indistinguishable results, and only one is discussed in detail. Each oligomer was solvated by a 0.1 M NaCl water solution in a cubic box with a minimum distance between periodic images of 30 Å. For each system, 50  $\mu$ s of MD simulation in the NPT ensemble at 300 K was performed. Five microseconds of MD simulation was also performed using the Amber-OL15 force field<sup>4</sup> using the TIP3P<sup>16</sup> and the OPC<sup>58</sup> water models and the Joung–Cheatham ion parameters.<sup>59</sup> The base-step, base-pair, and nucleobase structural parameters of the oligomers were calculated for 5000 evenly spaced frames of each simulation using the DSSR software.<sup>60</sup>

**MD Simulations of Noncanonical DNA Structures.** We performed simulations of Z-DNA, two CEB2 and telomeric G-quadruplexes, and one DNA trimer. The initial structures were taken from PDB entries 4ocb,<sup>61</sup> 2lpw,<sup>62</sup> 1jrn,<sup>63</sup> and 1bwg.<sup>64</sup> Five independent simulations of Z-DNA at different salt concentrations were performed with the system solvated in 0.1, 0.5, 0.75, 1.0, or 1.5 M KCl in a 70<sup>3</sup> Å<sup>3</sup> cubic box. The quadruplexes were solvated in 0.1 M KCl in a 60<sup>3</sup> Å<sup>3</sup> cubic box. The trimer was solvated in 0.15 M KCl in a 80<sup>3</sup> Å<sup>3</sup> cubic box. To equilibrate the ion distribution and allow K<sup>+</sup> ions to migrate between the nucleobases, the telomeric G-quadruplex was equilibrated for 1  $\mu$ s using position restraints on the nucleic acid heavy atoms with a force constant of 0.1 kcal mol<sup>-1</sup> Å<sup>-1</sup>; CEB2 was equilibrated for 5  $\mu$ s using position restraints on the backbone atoms with a force constant of 0.2 kcal mol<sup>-1</sup> Å<sup>-1</sup>. Production runs of 50  $\mu$ s (with no position restraints) were then performed for each system.

**MD Simulations of Protein–Nucleic Acid Systems.** We performed simulations of six protein–RNA and six protein–DNA complexes. The starting structure for each simulation was the experimental NMR or X-ray structure as found in the PDB: 2dgc,<sup>65</sup> 3jxc,<sup>66</sup> 3zhm,<sup>67</sup> 4mhg,<sup>68</sup> 1tro,<sup>69</sup> 3bdn,<sup>70</sup> 2az0,<sup>71</sup> 4qoz,<sup>72</sup> 1urn,<sup>73</sup> 4qil,<sup>74</sup> 5dea,<sup>75</sup> and 3k5y.<sup>76</sup> Protein–nucleic acid complexes were solvated in a 0.1 M NaCl water solution in a cubic box with a minimum distance between periodic images of 30 Å. Production runs of 50  $\mu$ s were performed for each system.

## RESULTS

**Force Field Parameterization.** Development of the DES-Amber Force Field Parameters for DNA. As an initial step, we incorporated into the Amber-bsc1 DNA force field the revised nucleobase LJ and electrostatic nonbonded parameters that were derived previously for RNA.<sup>7</sup> For thymine, which is

typically not found in RNA—as it is replaced by uracil in most cases—the Amber charges were used.

We developed two force field parameter sets: one (which we name *DES-Amber*) with charge scaling consistent with the previously reported DES-Amber protein and ion force field parameters and a second (which we name *DES-Amber SF1.0*) with no charge scaling, consistent with the previously reported DES-Amber SF1.0 protein and ion force field parameters.<sup>8</sup> Our previously reported RNA parameter set<sup>7</sup> did not include charge scaling and is thus part of the DES-Amber SF1.0 force field; the revised RNA parameters reported in this work, with charge scaling applied to ionic residues, are part of the DES-Amber force field. Each force field (DES-Amber and DES-Amber SF1.0) thus contains parameters for DNA, RNA, proteins, and ions. Both DES-Amber and DES-Amber SF1.0 are based on procedures similar to those used when developing the DES-Amber protein force field,<sup>8</sup> with the only difference between the two parameter sets being the charge of the ionic residues, which are scaled by 0.9 in the case of DES-Amber and 1.0 (i.e., no rescaling) in the case of DES-Amber SF1.0.

Previous work has shown that glycosidic torsion parameters should not be shared between DNA and RNA molecules.<sup>9</sup> We thus focused the torsion optimization on the  $\chi$  (OS-CT-N-C) and  $\zeta$  (CT-OS-P-OS) angles; all other torsion parameters and bonded terms were kept the same as the RNA parameters in DES-Amber SF1.0 derived previously.<sup>8</sup> The torsion parameters describing the nucleotide-specific  $\chi$  angles were obtained from a fit to the torsion QM energy profiles calculated at the MP2 level using MolPro.<sup>12,77</sup> This level of theory has been found to accurately describe torsion angles in nucleobases,<sup>9</sup> and we found it to be sufficient to achieve a good level of accuracy for the parameterization of  $\chi$  angles in RNA.<sup>7</sup> Following previous work, single nucleosides in both C3'-endo and C2'-endo conformations were used as model compounds in the generation of  $\chi$ -related torsion energies.<sup>9</sup> Restraints were applied to the sugar hydroxyl group torsions to prevent the formation of intramolecular hydrogen bonds between the sugar and the nucleobase during geometry optimization. For the  $\zeta$  angle, a deoxyribose *N*-methyl phosphate molecule was used as a model compound in the generation of torsion energies. The revised torsion parameters are reported in Table 1.

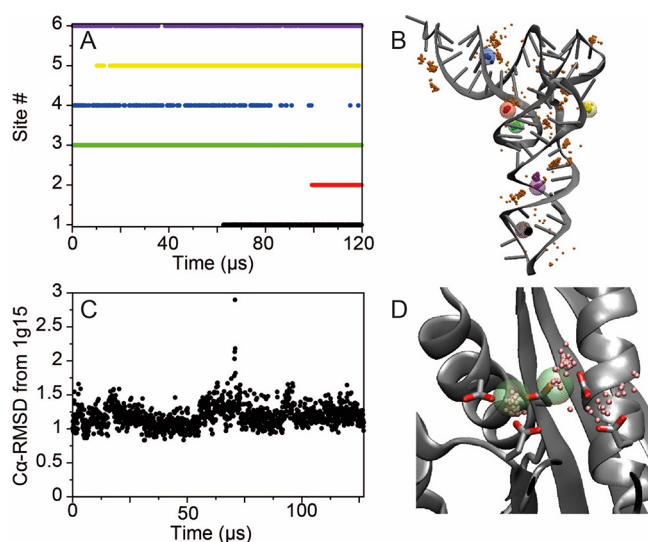
**Parameterization of the Magnesium Ion.** Magnesium ions are often very important for the stability of the tertiary structure of nucleic acids.<sup>78,79</sup> Previous studies have highlighted that, although the description of the hydrated Mg<sup>2+</sup> ion appears to be satisfactory,<sup>80</sup> the original parameterizations for the Mg<sup>2+</sup> ions in Amber<sup>81</sup> and CHARMM are suboptimal.<sup>82</sup> In both of these ion parameterizations, the ionic radius of the ion is too small, leading to the overbinding of Mg<sup>2+</sup> and slow exchange of the water shell. These deficiencies could result in simulation inaccuracies, especially in cases where the process of ion binding to the phosphate backbone is important. Subsequent studies have highlighted the importance of developing water-specific ion parameters,<sup>59</sup> in particular for divalent ions.<sup>83,84</sup> We thus revised the nonbonded parameters of the Mg<sup>2+</sup> ions with the aim of making them compatible with the TIP4P-D water model<sup>17</sup> and the DES-Amber force field. We focused our parameterization on trying to achieve a balance between the affinity of the Mg<sup>2+</sup> ion for water and for the negatively charged groups on proteins and nucleic acids (phosphates and carboxylates) while keeping the ionic radius close to the values estimated experimentally. Although we were able to achieve a reasonable compromise between these

competing objectives for the scaled-charge version of DES-Amber, we were unable to obtain a good fit for DES-Amber SF1.0, so in this force field, the original Amber parameters<sup>81</sup> were retained.

In our optimization of parameters for Mg<sup>2+</sup>, we used Mg<sup>2+</sup> binding to tRNA<sub>Phe</sub> as our benchmark system. This RNA molecule folds into a complex T-shaped structure stabilized by multiple Mg<sup>2+</sup> ions.<sup>85</sup> In two sites, Mg<sup>2+</sup> ions directly interact with the nucleotides (bridging the phosphate groups of two nucleotides), and in an additional three sites, the ions are completely solvated and only interact indirectly with the phosphate moieties through the hydration shell.<sup>86,87</sup> We expect that the ability to recapitulate both binding modes in a single simulation would provide an indication that the force field parameters achieve a reasonable balance between the affinity for water and the affinity for biomolecules. We were unable, however, to achieve this balance using standard combination rules for the LJ potential. We thus decided to first set pair-specific LJ interaction parameters between Mg<sup>2+</sup> and water to achieve an ionic radius consistent with the experimentally determined first peak of the radial distribution function of water and Mg<sup>2+</sup> and then scan for a suitable set of atomic LJ parameters for the Mg<sup>2+</sup> ion that would allow us to optimally describe the different ion-interaction sites in tRNA<sub>Phe</sub>.

Following previous work,<sup>82,88</sup> the LJ parameters for the Mg–O<sub>wat</sub> interaction were first determined by performing simulations of Mg<sup>2+</sup> in water, with  $\sigma$  parameters ranging from 2.9435 to 3.097 Å and  $\epsilon$  ranging from 0.015 to 0.045 kcal mol<sup>-1</sup>, ultimately choosing the parameters  $\epsilon = 0.025914$  kcal mol<sup>-1</sup> and  $\sigma = 3.004$  Å, as these values best reproduced the experimental Mg–O<sub>wat</sub> distance of 2.07–2.11 Å<sup>89,90</sup> and the water exchange rate of  $6.7 \times 10^5$  s<sup>-1</sup>.<sup>91</sup> Subsequently, simulations of 50–150  $\mu$ s were performed starting from a tRNA<sub>Phe</sub> structure (PDB entry 1ehz)<sup>87</sup> in which the six magnesium ions identified in the X-ray structure were removed and a total of 11 magnesium ions were added to the water solution together with 0.15 M KCl. Position restraints of 0.3 kcal mol<sup>-1</sup> Å<sup>-1</sup> were applied to the nucleic acid backbone to prevent the collapse of the structure when no structural magnesium ions were bound. During the simulations, the occupancy of the sites where Mg<sup>2+</sup> ions were observed in the X-ray structure and the coordination sphere of the ions were monitored. We gradually increased the value of the Mg<sup>2+</sup>  $\sigma$  from 2.77 Å, and for each increase, we tested values of the Mg<sup>2+</sup>  $\epsilon$  ranging from 0.009 to 0.002 kcal mol<sup>-1</sup>; we found, using LJ parameters for Mg<sup>2+</sup> (for all interactions except with water) of  $\epsilon = 0.003000$  kcal mol<sup>-1</sup> and  $\sigma = 3.075$  Å, that after 100  $\mu$ s of MD simulation starting from a tRNA<sub>Phe</sub> structure with no Mg<sup>2+</sup> ions bound, five of the six sites were occupied by Mg<sup>2+</sup> ions in the correct coordination pattern (Figure 1A,B). This suggests that this set of force field parameters results in a reasonable affinity for the phosphate backbone. In the case of the sixth site, a neighboring site was occupied instead. Finally, we verified that, following equilibration of the position of the Mg<sup>2+</sup> ions, the tRNA structure was also stable in the absence of position restraints (Figure S12).

As an additional test, we performed a simulation of the *Escherichia coli* ribonuclease H (RNaseH). The active site of this protein contains four negatively charged residues (Asp10, Glu48, Asp70, and Asp134) and one Mg<sup>2+</sup> ion.<sup>92</sup> At high Mg<sup>2+</sup> concentration, binding of a second Mg<sup>2+</sup> ion inactivates the enzyme. We performed a simulation starting from the apo crystal structure (PDB entry 2rn2),<sup>93</sup> in which Mg<sup>2+</sup> ions are



**Figure 1.** Binding of  $\text{Mg}^{2+}$  to tRNA and RNaseH. (A) Occupancy of the six  $\text{Mg}^{2+}$  sites in tRNA as a function of simulation time. After 120  $\mu\text{s}$  of simulation, five of the binding sites observed experimentally<sup>87</sup> were occupied. (B)  $\text{Mg}^{2+}$  ion positions between 100 and 120  $\mu\text{s}$  of simulation; 25 snapshots of the positions of the  $\text{Mg}^{2+}$  ions occupying the six binding sites are reported as small solid pink spheres. The positions of the crystallographic  $\text{Mg}^{2+}$  ions are highlighted as large transparent multicolored spheres. (C)  $C\alpha$ -RMSD in Å from the X-ray structure (PDB entry 1g15)<sup>92</sup> in a 127  $\mu\text{s}$  simulation of RNaseH started from the apo enzyme. (D) Snapshots of the positions of the  $\text{Mg}^{2+}$  ions from the last 20  $\mu\text{s}$  of simulation (small pink spheres). The positions of the crystallographic  $\text{Mn}^{2+}$  ions are reported as large green transparent spheres.

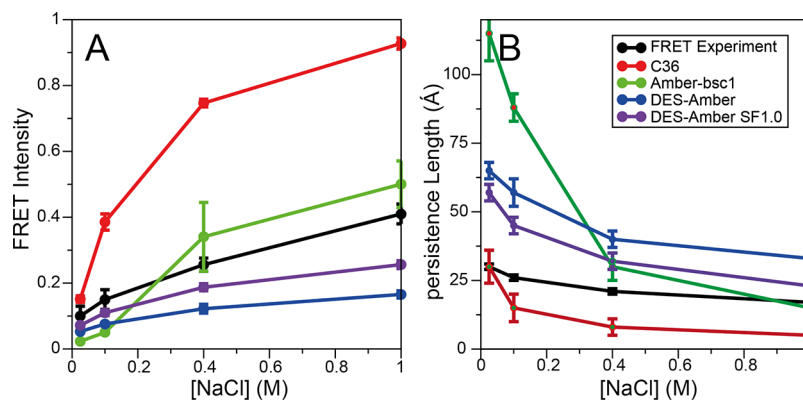
not present, in a 0.1 M  $\text{MgCl}$  and 0.1 M  $\text{NaCl}$  water solution. Position restraints were not necessary in this case, as  $\text{Mg}^{2+}$  is not required for the structural stability of the RNaseH protein. Indeed, the global protein structure was remarkably stable during the 127  $\mu\text{s}$  of simulation, as indicated by the average  $C\alpha$  root-mean-square deviation (RMSD) of 1.2 Å with respect to the holo structure (PDB entry 1g15)<sup>92</sup> (Figure 1B). After approximately 50  $\mu\text{s}$  of simulation, a  $\text{Mg}^{2+}$  ion occupied one of the binding sites occupied by a divalent ion in the  $\text{Mn}^{2+}$ -bound X-ray structure, and at approximately 100  $\mu\text{s}$  of simulation, a second ion bound to the active site (Figure 1D). Both ions remained bound for the rest of the simulation, with the second

ion being more loosely coordinated (Figure 1C), an observation that may be consistent with its lower experimentally determined affinity.<sup>92</sup>

In our simulation setup for the RNaseH system, there were 18  $\text{Mg}^{2+}$  ions in solution and 20 negatively charged residues in the protein. During the course of the simulation,  $\text{Mg}^{2+}$  ions transiently interacted with other protein residues, but despite the large number of possible interactions, only two additional persistent interactions were observed (with Glu6 and Glu154). The significance of these interactions is unclear, as the binding–unbinding kinetics are slow with respect to the timescale of the simulation, but the low number of alternative persistent interactions observed suggests that this set of  $\text{Mg}^{2+}$  parameters does not lead to strong overbinding of the magnesium ions to the negatively charged protein residues.

**Single-Stranded DNA Tests. Solution Conformation of Long Single-Stranded DNA Molecules.** In testing DES-Amber, our first goal was to verify that the force field is capable of reproducing the conformational properties of long, disordered single-stranded nucleic acids. Previously, we have observed that simulations of a long, disordered RNA molecule (rU40) performed with the Amber and, to a lesser extent, the CHARMM force fields resulted in highly compact conformations, at odds with experimental observations, and found that this issue was resolved in the DES-Amber SF1.0 force field.<sup>7</sup> For this work, we used a 40-residue poly-thymine ( $\text{dT}_{40}$ ), as our test system. Although this molecule is not of particular biological significance, it is known to be mostly disordered in solution<sup>94,95</sup> and thus may be particularly sensitive to force field errors. With both DES-Amber and DES-Amber SF1.0, we performed 100  $\mu\text{s}$  of simulation at 300 K in 0.025, 0.1, 0.4, and 1.0 M  $\text{NaCl}$  solutions; for comparison, we also performed simulations with the Amber-bsc1<sup>1</sup> and CHARMM36<sup>5</sup> force fields. These two force fields were chosen because they are generally believed to represent the state of the art for the Amber and CHARMM family of force fields.<sup>5,96</sup> As a first step, for each simulation, we compared the persistence length ( $L_p$ ) and the FRET intensities calculated assuming a Förster radius of 56.4 Å<sup>14,94</sup> (Figure 2A,B) to the reported experimental values.

At low salt concentrations, simulations with both DES-Amber and DES-Amber SF1.0 resulted in relatively expanded conformational ensembles that are generally consistent with the FRET intensity values measured experimentally at low salt



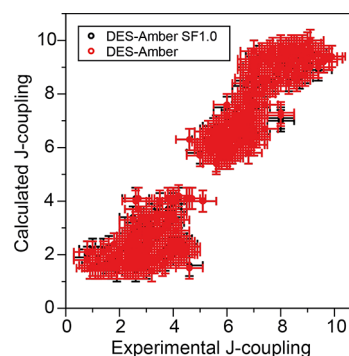
**Figure 2.** Comparison of calculated and experimental (A) FRET intensities and (B) persistence lengths for  $\text{dT}_{40}$  as function of  $\text{NaCl}$  concentration for simulations run with CHARMM36 (C36, red), Amber-bsc1 (green), DES-Amber (blue), and DES-Amber SF1.0 (purple). Experimentally measured FRET intensities and persistence lengths<sup>94</sup> are reported in black.

concentrations,<sup>14</sup> although the persistence length appears to be overestimated due to the prevalence of relatively rigid B-like structures. This effect was even more pronounced in the Amber-bsc1 simulations, where the conformational ensemble was dominated by the presence of a long B-like structure with an end-to-end distance of >120 Å and a persistence length of 85 Å. This finding is at odds with experimental observations indicating that  $dT_{40}$  should be somewhat disordered, with an end-to-end distance of  $\sim 65$  Å<sup>94</sup> and a persistence length of 30 Å.<sup>94</sup> This discrepancy may be the result of a slight overfitting of torsion parameters to helical structures.

The FRET intensity and persistence length calculated from the CHARMM36 simulations were in excellent agreement with experiment at the lowest salt concentration investigated here (0.025 M, Figure 2), but at physiological salt concentrations, the simulations resulted in FRET intensities that are much higher than the experimental values due to the strong tendency of the CHARMM36 force field to collapse into molten globule structures. In simulations at high salt concentration, the extended structures observed in DES-Amber, DES-Amber SF1.0, and Amber-bsc1 had a greater tendency to collapse, resulting in a smaller persistence length and higher FRET intensities, in fair agreement with experiment. In the case of Amber-bsc1, the collapsed species in simulation were characterized by metastable hairpin structures (as indicated by the large number of base pairs that are transiently formed, Figure S1). The significance of this finding is unclear, as FRET and SAXS data suggest that poly-dT should be largely disordered in solution,<sup>14,94,95,97</sup> but an atomic-level structural characterization of this system using experimental approaches is difficult.<sup>97</sup>

Overall, these results suggest that the dimensions and flexibility of the disordered state, a quantity that can be strongly affected by force field artifacts and is difficult to predict in simulation,<sup>7</sup> are recapitulated reasonably well by both DES-Amber force fields. It should be noted that the DES-Amber force fields appear to underestimate the level of compaction that occurs at high salt concentration (Figure 2), which results in disordered ensembles and persistence lengths that are slightly larger than the experimental estimates, possibly due to some degree of overstabilization of B-like structures.

**Solution Conformation of Short Single-Stranded DNA Molecules.** To test the description of local conformational fluctuations in short ssDNA, we performed simulated tempering simulations of 18 ssDNA molecules composed of two to six nucleotides (Table S1). Each system was subjected to 20  $\mu$ s of simulation using DES-Amber and DES-Amber SF1.0, and we compared the resulting <sup>3</sup>J-couplings for the sugar ring protons, calculated using the Haasnoot Karplus equation,<sup>24,25</sup> to a dataset of 644 <sup>3</sup>J-couplings measured using NMR spectroscopy (Figure 3).<sup>26–35</sup> Using both revised force fields, most of the systems investigated had mean squared errors <1 Hz<sup>2</sup>, indicating good agreement between the calculated and experimental <sup>3</sup>J-couplings (for comparison, the average root-mean-square error of Amber-bsc1 across all systems is 1.70 Hz; Table S1). The largest deviations were observed in the H1'–H2' <sup>3</sup>J-coupling of the first nucleobase; the calculated coupling for this pair was often larger than the experimental value, suggesting that, in simulation, the sugar ring slightly overpopulates the S conformation.<sup>25,26</sup> These deviations, however, are generally not much larger than the expected accuracy of the Haasnoot parameterization of the Karplus relation, and to avoid the risk of overfitting, we refrained from performing



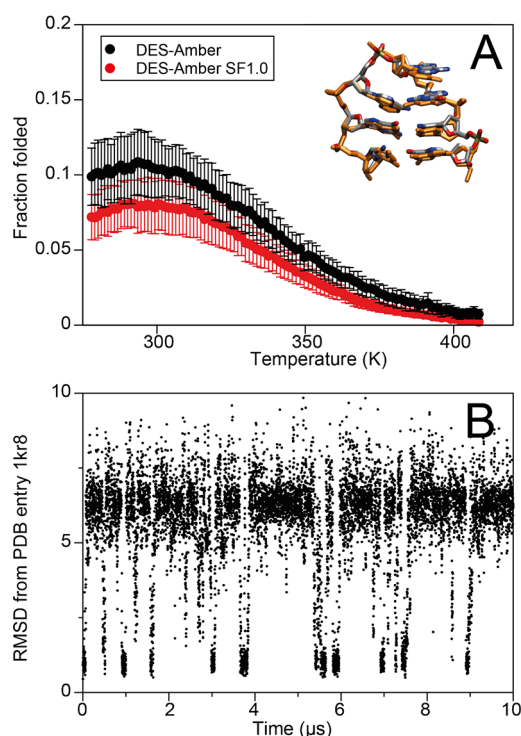
**Figure 3.** Comparison of 644 calculated and experimental J-couplings for the deoxyribose protons in 18 short ssDNA molecules, ranging from two to six nucleotides in length, in simulations using the DES-Amber (red) and DES-Amber SF1.0 (black) force fields.

additional tuning of the force field parameters based on these data.

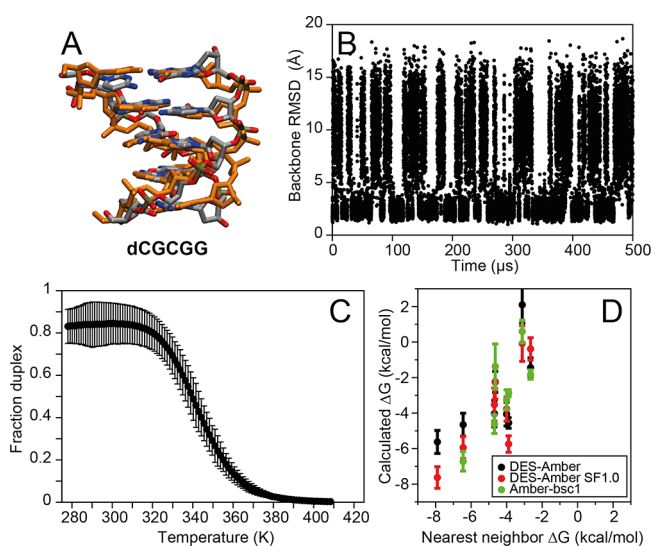
**Double-Stranded DNA Tests. Structure and Stability of a DNA Hairpin.** Having established that DES-Amber provides a reasonable description of disordered single-stranded DNA molecules, our next goal was to test whether the force field could describe molecules that fold into a well-determined structure, such as a DNA hairpin. We performed 170  $\mu$ s of simulated tempering simulations of the d(GCGAAGC) polynucleotide using DES-Amber and DES-Amber SF1.0. This sequence has been found experimentally to fold into a hairpin in solution (PDB entry 1kr8).<sup>36</sup> On the  $\mu$ s timescale in our simulations, the polynucleotide folded into structures with heavy-atom RMSDs of <1 Å from the experimentally determined structure (Figure 4), suggesting a folding rate roughly 2 orders of magnitude faster than that of RNA tetraloops.<sup>7</sup> Similarly to what we observed for a number of RNA tetraloops, the hairpin was less stable than what is observed experimentally (the calculated  $\Delta G_{\text{fold}}$  at 310 K was 1.3 and 1.5 kcal mol<sup>-1</sup> for DES-Amber and DES-Amber SF1.0, respectively, compared to the experimentally determined<sup>98</sup>  $\Delta G_{\text{fold}}$  of  $-3.2$  kcal mol<sup>-1</sup>). This discrepancy may be the result of residual inaccuracy in the description of the nonbonded interactions.

**Thermodynamic Stability of DNA Duplexes.** B-DNA, as the most common double helix found in nature, is arguably the most important structure that is formed by DNA molecules. As a test of the ability of the force field to correctly identify double-stranded B-DNA as the thermodynamically most stable structure, and to recapitulate the observed experimental dependence of stability on the DNA sequence, we performed 500  $\mu$ s simulated tempering simulations of seven pairs of DNA sequences with lengths ranging between four and seven nucleotides. We found that the correct B-DNA structure formed reversibly (Figure 5A and Figures S2–S4) in simulations performed with all three force fields investigated (DES-Amber, DES-Amber SF1.0, and Amber-bsc1). The B-DNA structure was also correctly identified as the thermodynamically most stable conformation at low temperatures (Figure 5B,C).

Some notable differences were observed between Amber-bsc1 and the DES-Amber force fields in the non-B-DNA conformational ensembles. Whereas in DES-Amber and DES-Amber SF1.0 the two DNA oligomers were noninteracting 60–70% of the time when B-DNA was not formed, in Amber-



**Figure 4.** (A) Folding free energy as a function of temperature for the DNA hairpin d(GCGAAGC) with the DES-Amber (red) and DES-Amber SF1.0 (black) force fields. The structure obtained by averaging the frames between 9.2 and 9.5  $\mu\text{s}$  of simulation time is shown superimposed with the experimental structure. (B) Backbone RMSD trace ( $\text{\AA}$ ) for the first 10  $\mu\text{s}$  of the simulated tempering simulation of the d(GCGAAGC) DNA sequence performed with the DES-Amber force field.



**Figure 5.** Simulated tempering simulations of short B-DNA fragments. (A) Snapshot from the d(CGCGG) simulation superimposed on a canonical B-DNA reference structure. (B) Backbone RMSD from a canonical B-DNA structure in the simulation of d(CGCGG) performed with DES-Amber. (C) Fraction of B-DNA duplex frames as a function of temperature. (D) Calculated melting free energy at 300 K for short B-DNA duplexes from simulations performed with DES-Amber, DES-Amber SF1.0, and Amber-bsc1 compared to the melting free energy predicted using a nearest-neighbor model.

bsc1, the two strands rarely fully separated and, when B-DNA was not formed, were bound in molten globule states with noncanonical structure 90–97% of the time (Table S3). Such states have not been reported in the literature and are probably an artifact of the tendency of this force field to form disordered structures that are overcollapsed.

There was generally a reasonable correlation between the calculated relative stability and the stability predicted by a nearest-neighbor model<sup>40</sup> (Figure S5D and Table S2), although the absolute stability at 310 K was systematically underestimated (by approximately 0.2 kcal mol<sup>-1</sup> per nucleobase for DES-Amber and 0.15 kcal mol<sup>-1</sup> per nucleobase for DES-Amber SF1.0 and Amber-bsc1). The origin of this discrepancy is unclear. It would be possible, in principle, to achieve much better agreement with experiment for these systems through a suitable modification of torsion parameters, as is often performed in the optimization of protein force fields.<sup>37,99,100</sup> We decided, however, not to pursue this route, as the simulations of short ssDNA molecules indicate that there is already a strong tendency to populate B-DNA conformations, suggesting that errors in the torsion parameters may not be responsible for the behavior observed here.

**Solution Structure of Long DNA Duplexes.** It is important that a DNA force field have the ability to reproduce the subtle deviations from the canonical B-DNA conformation that are induced by changes to local sequence, as these differences can be relevant for function and sequence recognition. Following previous work on DNA force field development,<sup>1</sup> we have performed simulations of 19 dsDNA molecules containing 10 to 17 nucleobase pairs using the DES-Amber and DES-Amber SF1.0 force fields. Each simulation was run in the NPT ensemble at 300 K for 50  $\mu\text{s}$ , and the structure observed in simulation was compared to the corresponding experimentally determined structure. For all but one of the systems investigated, the DNA structure remained close to the experimental structure (Table S4), although in some cases fraying of the termini was observed. The RMSDs were comparable to those observed for Amber-bsc1 (e.g., average backbone RMSD was 2.47  $\text{\AA}$  for DES-Amber and 2.49  $\text{\AA}$  for Amber-bsc1), a force field that has been shown to describe these systems with extremely high accuracy.<sup>1,101</sup>

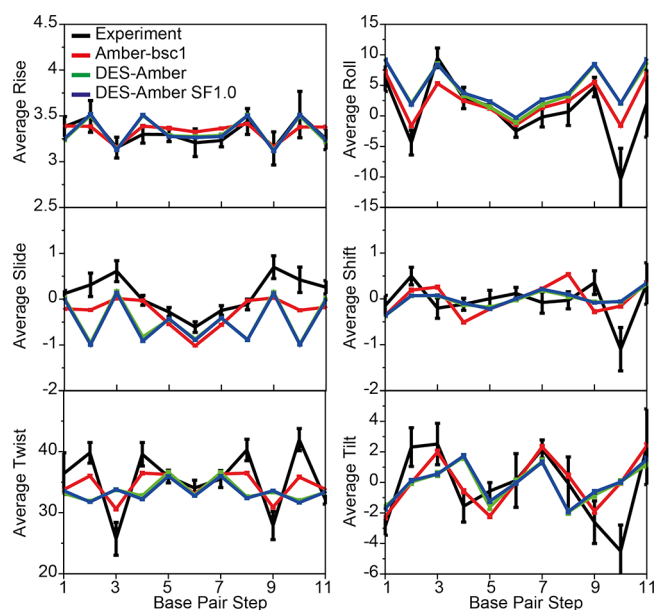
A somewhat larger structural deviation from the X-ray structure was observed for d(CTTTTAAAAG) using all force fields, in particular when using DES-Amber SF1.0; this is possibly because this sequence was crystallized with seven Ca<sup>2+</sup> ions bound, and these ions were not present in the MD simulation. In simulations of d(GGATATATCC) performed using DES-Amber SF1.0, the duplex dissociated. This is not entirely unexpected given that nearest-neighbor models predict this sequence to be only marginally stable at 300 K,<sup>102,103</sup> and our calculations of the free energy of melting of short duplexes suggest that the DES-Amber SF1.0 force field may slightly underestimate duplex stability.

Although RMSD is a useful metric for monitoring large conformational transitions such as DNA melting or the B-A transition, it is relatively insensitive to the details of the fine structure of the DNA double helix. We thus calculated average base-step and base-pair parameters, as well as structural parameters for each nucleobase. The full set of data is available in the SI. DES-Amber, DES-Amber SF 1.0, and DES-Amber 3.20 results are almost indistinguishable; this is not unexpected, as they largely share the same torsion potentials.

Previous studies have indicated that force field inaccuracies can result in the over-stabilization of the  $\gamma$ -trans conformation, which can become the dominant conformation at long timescales.<sup>104–106</sup> Our structural analyses indicate that in simulations performed with the DES-Amber force fields, the  $\gamma$ -trans conformation is observed in the terminal base pairs 10% of the time, possibly due to occasional fraying, and 1% of the time in the rest of the sequence. (The corresponding results are 8 and 4% in Amber-bsc1 and 3 and 0.4% in Amber-OL15, respectively). The deviation of the step and pair parameters from the X-ray data was calculated for all force fields (Tables S9–S14). Although it is unclear how closely the simulation of the duplex in solution should resemble the X-ray structure, we found that the parameters in simulations performed using the DES-Amber force fields deviate from the X-ray structure by about the same amount as the reference force fields (Amber-bsc1 and Amber-OL15), suggesting that the DES-Amber force fields provide a description of the average nucleobase conformation with about the same accuracy as these state-of-the-art force fields. Averages were also calculated for the helical rise and twist per base step (Table S15). We found that, for all force fields, the calculated average rise was  $\sim 3.35$ ,  $\sim 0.1$  Å larger than that in the X-ray structures. The average helical twist was  $33^\circ$  for DES-Amber and  $34^\circ$  for Amber-OL15 and Amber-bsc1, values that are comparable to the  $35^\circ$  average value in the X-ray structures.

Finally, we analyzed in more detail the self-complementary Drew–Dickerson dodecamer (DDD) of sequence d(CGCGAATTCGCG), a system that has previously been very well-characterized by both simulations and experiments.<sup>5,41–43,96,107,108</sup> In all three force fields, simulations of DDD resulted in a structure closer to the structures determined by NMR in aqueous solution (e.g., DES-Amber average RMSD was 1.93 Å from PDB entry 1duf<sup>41</sup> and 2.14 Å from 1naj<sup>42</sup>) than to the crystal structure (e.g., DES-Amber average was RMSD 2.61 Å from PDB entry 1bna<sup>43</sup>) (Table S4).

For the six base-step parameters (twist, roll, slide, rise, shift, and tilt),<sup>60</sup> we compared the values to the range of values observed in experimentally determined structures<sup>101</sup> (Figure 6). All base-step parameters were well converged in 50  $\mu$ s of simulation, with DES-Amber and DES-Amber SF1.0 giving almost identical results. Agreement with experiment for most of the base-step parameters was satisfactory, though the local description of the GC base step appeared to be inaccurate; this is consistent with the observation that the helical twist of the GC step appears to be underestimated in DES-Amber simulations (Tables S9–S11). Interestingly, simulations performed with Amber-bsc1, despite having an overall RMSD from the experimental structures (1.83 Å from 1duf) similar to that observed in simulations performed with DES-Amber (1.93 Å) (Table S4), typically yielded a better description of the local geometry of the GC base step. The CGCG region of the DDD dodecamer has been shown to exhibit unusual flexibility due to the population of a large fraction of the BII state at 300 K (34% for the GC step and 56% for the CG step).<sup>109</sup> The BI/BII population of these steps in the DDD simulations performed with the DES-Amber force fields was only 1% for the GC step and 3% for the CG step, whereas it was  $\sim 30$  and  $\sim 50\%$ , respectively, for Amber-bsc1 and Amber-OL15. More generally, we found that the BII population in the DES-Amber duplex simulations rarely exceeded a few percent, while it was almost an order of



**Figure 6.** MD simulations of DDD. Comparison of the average six base-step parameters (twist, roll, slide, rise, shift, and tilt) measured in simulations performed with DES-Amber (green), DES-Amber SF1.0 (blue), and Amber-bsc1 (red) to the values observed in seven crystal and NMR structures deposited in the protein data bank (PDB entries 1bna, 2bna, 7bna, 9bna, 1naj, 1jgr, and 4c64)<sup>101</sup> (black). Error bars for the experimental values report the variance of the data. Error bars for the simulated data are estimated using block averaging of the trajectory data.

magnitude larger in simulations performed using Amber-bsc1 and Amber-OL15. This suggests that there is some imbalance in describing the relative stability of the BI and BII conformations and points to an area in which the DES-Amber force field could be improved, possibly by suitable optimization of the backbone torsions or the guanine nonbonded parameters that for this work were simply transferred, without further refinement, from the RNA parameters previously derived for the DES-Amber SF1.0 force field.

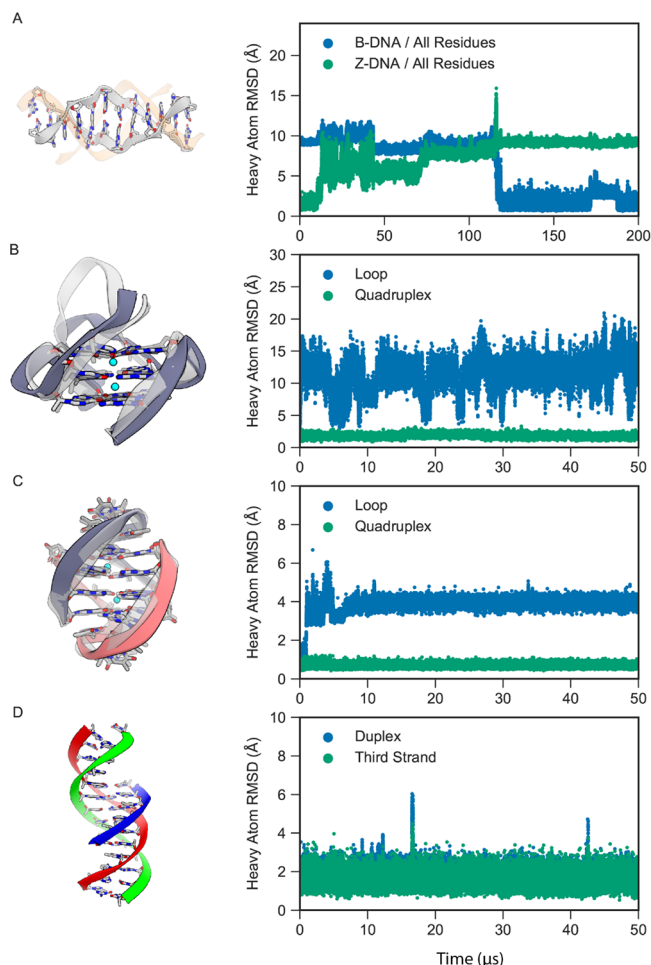
#### Structural Stability of Noncanonical DNA Structures.

Although B-DNA is the predominant form of DNA under physiological conditions, particular sequences or conditions can stabilize alternate DNA structures. We investigated the ability of DES-Amber to recapitulate some of these structures by performing simulations of (i) Z-DNA at varying salt concentrations, (ii) two DNA G-quadruplexes, and (iii) a DNA triplex.

Z-DNA is a left-handed helical form of DNA that is stabilized by GC-rich sequences containing 8-methylguanine,<sup>110</sup> or 5-methyl cytosine in water/methanol solution<sup>111</sup> or at very high salt concentrations ( $>3$  M NaCl).<sup>112,113</sup> For d(CG)<sub>n</sub>, the Z-DNA conformation can be crystallized under physiological conditions, suggesting that it should be at least marginally stable in solution; improving the stability of this structure has been the focus of a recent Amber force field optimization.<sup>114</sup> We have performed 200  $\mu$ s simulations starting from a d(CGCGCGCGCGCG) Z-DNA dodecamer X-ray structure (PDB entry 4ocb)<sup>61</sup> at salt concentrations ranging from 0.1 to 1.5 M NaCl. Simulations could not be performed at higher concentrations, as the solubility of NaCl and KCl in DES-Amber is  $<2$  M. Simulations were fairly stable



on the timescale of a few microseconds, with backbone RMSDs from the starting X-ray structure of 1–2 Å (see Figure 7A for a representative simulation at 0.75 M), but extensive



**Figure 7.** MD simulations of other forms of DNA with DES-Amber. (A) A Z-DNA dodecamer (gray) in 0.75 M KCl spontaneously transitioned to B-DNA (orange). The heavy-atom RMSDs with respect to the initial Z-DNA structure (green) and final B-DNA structure (blue) are shown on the right. (B) The CEB25 human minisatellite locus G-quadruplex, with the initial structure shown in light gray and the final structure shown in dark gray. The heavy-atom RMSDs of the loop (blue) and quadruplex (green) segments (in the reference alignment to the quadruplex core) are shown on the right. (C) The *Oxytricha nova* telomeric DNA G-quadruplex, with the initial structure shown in light gray and the final structure shown in dark gray and red. The heavy-atom RMSDs of the loop (blue) and quadruplex (green) segments (in the reference alignment to the quadruplex core) are shown on the right. (D) Triple-stranded DNA, with the duplex shown in red and green and the third strand in blue. The heavy-atom RMSDs of the duplex (blue) and third strand (green) (in the reference alignment of the duplex) are shown on the right.

fraying of the duplex terminal base pairs was observed on longer timescales in all simulations. In most cases, one-third to one-half of the initial Z-DNA base pairs were still formed at the end of the simulation (Figure S8), a result suggesting that Z-DNA is only partially stable under the simulation conditions over long timescales; this finding is consistent with the experimental observation that rather extreme conditions are required to stabilize Z-DNA in solution.

In high-resolution X-ray structures of Z-DNA, two main backbone rotamer states (ZI and ZII) have been observed. Previous studies have shown that simulations of Z-DNA in 2 M NaCl performed with several force fields of the Amber family can populate rotamer conformations (ZI' and ZII') not observed in the X-ray structures.<sup>3,4</sup> Although the solution structure of Z-DNA in 2 M NaCl is not known, as it is only marginally stable,<sup>113</sup> the ZI' and ZII' rotamers are believed to be force field artifacts.<sup>4</sup> We have calculated the torsion angles from the first 2  $\mu$ s of the Z-DNA simulations at 1 M NaCl (Figure S21). On this timescale and at this salt concentration, the Z-DNA structure was largely preserved in the DES-Amber simulations, and the ZI (60–65%) and ZII (30–35%) rotamers dominate. We observed a small amount of the ZI' conformation (~5%) and none of the ZII' conformation.

Interestingly, in the 0.75 M simulation, the Z-DNA spontaneously transitioned to B-DNA without fully dissociating. The process began with terminal base-pair fraying, which led to the distortion of the full Z-DNA duplex after approximately 10  $\mu$ s. The structure then began to adopt a B-DNA-like structure on one end, while a DNA-bubble-like structure formed in the center of the duplex. The gap formed by the bubble transiently opened and closed by the formation of mismatched base pairs over the course of 100  $\mu$ s until the entire structure eventually underwent a cooperative transition to B-DNA at roughly 120  $\mu$ s (Figure 7A).

Another important noncanonical structure is the G-quadruplex, a conformational form of DNA commonly found in telomeric DNA. We performed 50  $\mu$ s simulations of two DNA G-quadruplexes: a repeat unit of the CEB25 human minisatellite locus (PDB entry 2lpw)<sup>62</sup> and the *Oxytricha nova* telomeric DNA G-quadruplex (PDB entry 1jrn).<sup>63</sup> The G-quadruplex structure is generally composed of a quadruplex region stabilized by potassium ions that intercalate between four guanine bases arranged in a plane separated by loops: a single nine-nucleotide-long loop in CEB25 and two three-nucleotide-long loops in the telomeric DNA system. All crystallographic ions and water molecules were removed during simulation setup, and the systems were solvated in 0.1 M KCl. We first performed 1  $\mu$ s of restrained equilibration before running the 50  $\mu$ s simulations.

During equilibration, potassium ions intercalated between the guanine bases, stabilizing the complex; as a result, in both simulations, the quadruplex region was fairly stable (average heavy-atom RMSDs of 0.8 and 1.7 Å, as shown in Figure 7B,C). The loop region in CEB25, however, reversibly sampled several different conformations, including the experimentally determined structure—a result consistent with the higher degree of flexibility observed experimentally for this region.<sup>62</sup> The two shorter loops in the telomeric DNA were more rigid; in all simulations, however, the ion sites at the stem–loop junction were rarely occupied, and flipping of T7 after a few microseconds of simulation induced a slight rearrangement in the loop structures, with formation of a T5–T8 and T17–T20 pair (Figure 7C). This conformation has been observed in previous studies using enhanced sampling methods.<sup>115</sup> Although accurate parameterization of torsion angles can improve the loop stability, such as in the Amber-OL15 force field,<sup>4</sup> it has been hypothesized that an accurate description of this loop requires a polarizable force field.<sup>116</sup>

Finally, we performed simulations of triple-stranded DNA (also known as H-DNA or triplex DNA). Triplex DNA is a complex formed between a B-DNA duplex and a third

oligonucleotide.<sup>64</sup> The complex is formed by Hoogsteen base pairs<sup>117</sup> and is stabilized under acidic conditions due to the protonation of cytosines to create a C-G<sup>+</sup>C<sup>+</sup> triad.<sup>117</sup> We parameterized the protonated cytosine based on our unprotonated cytosine parameters (see [Methods](#)) and simulated the d(GACTGAGAGACGTA-TACCGTCTCT-CAGTC-CTCTCT) triple-stranded DNA complex<sup>64</sup> for 50  $\mu$ s. We find that in both DES-Amber and DES-Amber SF1.0, the complex was stable on the timescale of the simulation, with marginally higher stability in DES-Amber ([Figure 7D](#)). In particular, we found that even when the B-DNA duplex underwent terminal fraying, the third strand remained stable in the major groove. These results indicate that the DES-Amber force field can accurately capture the Hoogsteen base pairing necessary for stabilizing triple-stranded DNA.

**Protein–Nucleic Acid Complexes. Structural Stability of Protein–DNA Complexes.** Finally, we investigated the ability of DES-Amber to maintain the structure of protein–nucleic acid complexes. This is a particularly demanding test, as it requires an accurate balance of the interactions both within and between force fields for each biomolecular component.<sup>118,119</sup> We selected the complexes of DNA with the GNC4 leucine zipper (PDB entry 2dgc),<sup>65</sup> the P22 c2 repressor (PDB entry 3jxc),<sup>66</sup> the CI repressor from bacteriophage TP901-1 (PDB entry 3zhm),<sup>67</sup> the ETS transcriptional repressor ETV6 (PDB entry 4mhg),<sup>68</sup> the Trp repressor (PDB entry 1tro),<sup>69</sup> and the lambda repressor (PDB entry 3bdn)<sup>70</sup> as our test systems. Each system was run at 300 K for 50  $\mu$ s using DES-Amber and DES-Amber SF1.0, and the deviation from the starting structure was measured ([Figure 8](#)).

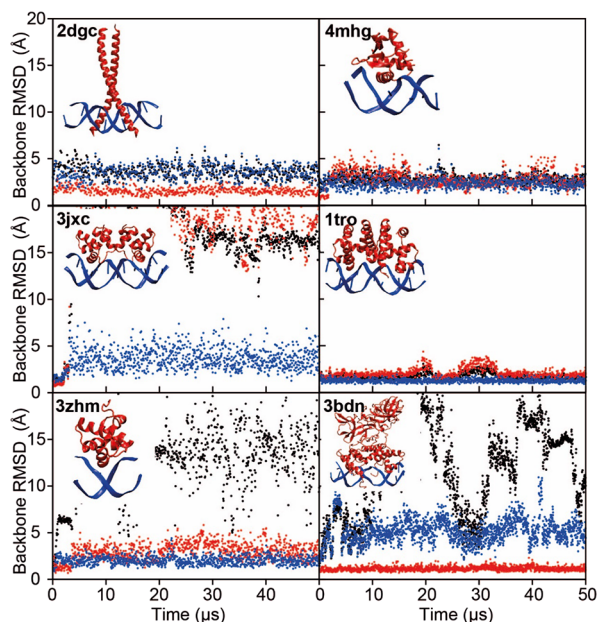
The Trp repressor, GNC4 leucine zipper, and ETS transcriptional repressor ETV6 complexes were stable in both force fields throughout the 50  $\mu$ s simulations. The lambda repressor–operon complex displayed a large amount of

flexibility when DES-Amber SF1.0 was used, with reversible fraying at the DNA termini and reversible binding/unbinding of one of the two lambda repressor units. This complex was more stable with DES-Amber, although transient binding/unbinding was also observed in this force field.

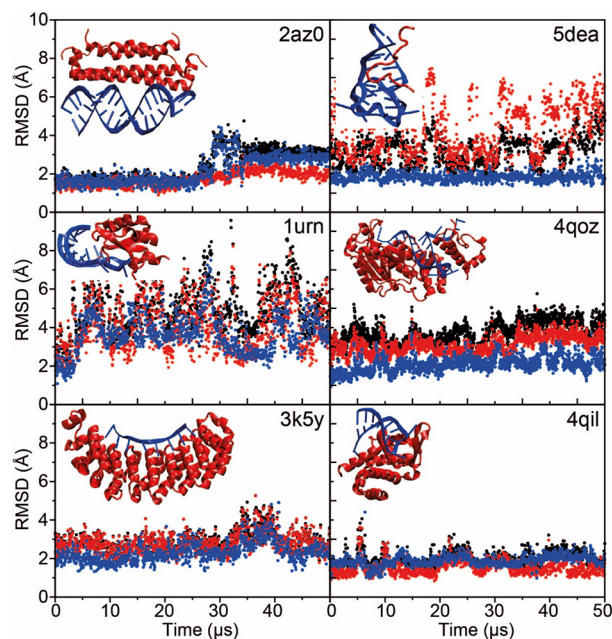
The P22 c2 repressor and CI repressor systems were stable only for a few microseconds, after which the complexes dissociated in both force fields. The P22 c2 repressor is a protein dimer with two independent binding sites, and dimerization is associated with the bending of the DNA operon due to a local B-DNA to B'-DNA transition.<sup>120</sup> In the simulation, the protein–protein dimer interface broke, leading to a straightening of the DNA double helix and disruption of the original geometry of the complex. The CI repressor is monomeric and binds selectively to a short double-stranded DNA stretch. This complex dissociated in the simulation, although both the DNA and the repressor are structurally very stable, as indicated by the small RMSDs with respect to the starting structure for these individual pieces of the complex.

**Structural Stability of Protein–RNA Complexes.** We also performed simulations of a set of six protein–RNA complexes with various structures and binding modes using DES-Amber; some of these were previously investigated by Krepl et al.<sup>119</sup> In particular, we conducted 50  $\mu$ s of simulations of the complexes of dsRNA–B2 protein (2az0),<sup>71</sup> tertiarily structured mRNA stem–loop, stem–loop binding protein and 3'hExoribonuclease 1 (4qoz),<sup>72</sup> RNA hairpin–U1A protein (1urn),<sup>73</sup> the ROQ domain–Hmg19 stem–loop RNA (4qil),<sup>74</sup> RNA duplex–quadruplex junction–RGG peptide (5dea),<sup>75</sup> and ssRNA–Pumilio FBF-2 protein (3k5y).<sup>76</sup>

None of the complexes dissociated, and four complexes (2az0, 4qoz, 3k5y, and 4qil) were remarkably stable for the entire 50  $\mu$ s of simulation ([Figure 9](#)). We did observe,



**Figure 8.** MD simulations of six protein–DNA complexes performed with the DES-Amber force field. RMSD trace as a function of simulation time for the complex (black), the protein (red), and the DNA (blue). Each panel is labeled with the PDB entry of the starting structure. The PDB structure of the complex is displayed as an inset in each plot.



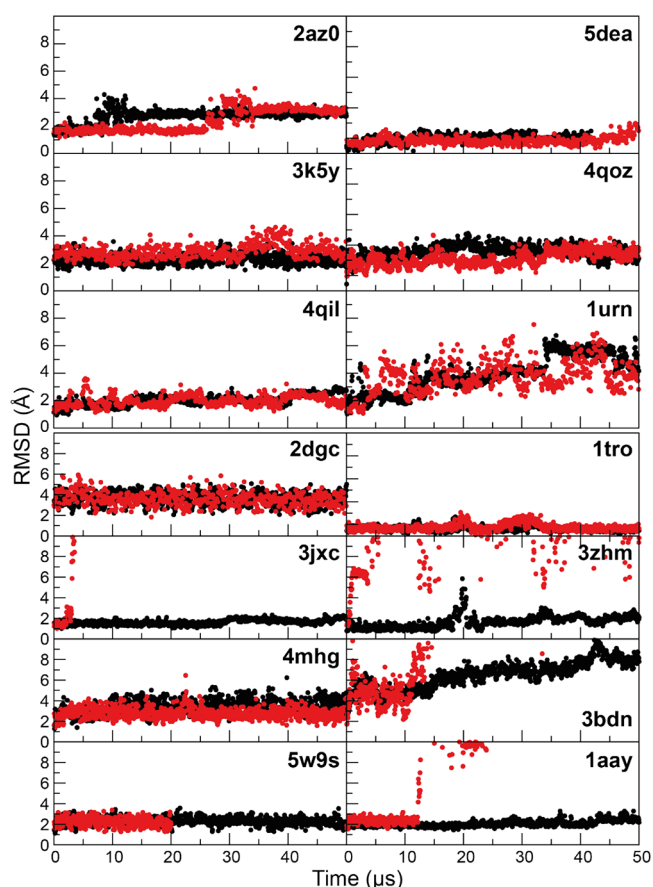
**Figure 9.** MD simulations of six protein–RNA complexes performed with the DES-Amber force field. RMSD trace as a function of simulation time for the complex (black), the protein (red), and the RNA (blue). Each panel is labeled with the PDB entry of the starting structure. The experimentally determined structure of the complex is displayed as an inset in each plot.

however, that a few key protein–nucleic acid interactions were unstable in the simulations. In 5dea, for example, the C- and N-terminal residues of the peptide were fairly flexible and only loosely bound, a result consistent with the NMR observation that only residues 8 to 17 are essential for binding.<sup>121</sup> During the simulation, however, the contact between RNA and Arg10 broke; this contact is considered critical for recognition<sup>121</sup> and is thus not expected to be broken.

Overall, the protein–nucleic acid simulations performed using DES-Amber on the 50  $\mu$ s timescale resulted in stable simulations for all protein–RNA complexes considered and for about half of the protein–DNA complexes. Complexes were sometimes found to be more flexible than indicated by experiment, and certain key interactions were poorly described. It is possible that, even in such cases, the simulations may still be informative, but care should be taken in the interpretation of the results.

**Reparameterization of Nucleic Acid Backbone Nonbonded Parameters to Improve the Description of Protein–Nucleic Acid Complexes.** Overall, we found that the DES-Amber force fields provided a fairly accurate description of nucleic acids but that their performance was less satisfactory in simulations of certain protein–nucleic acid complexes. We notice that often, in cases where the dissociation of a complex was observed, structures of the individual components (protein and nucleic acid) were reasonably well-maintained, suggesting that the main issue is related to inaccuracies in the description of nonbonded interactions at the protein–nucleic acid interface. We speculate that the use of a water model like TIP4P-D (or OPC, which has very similar parameters<sup>58</sup>) that reduces artifacts related to overcompaction may magnify force field inaccuracies that in simulations performed with water models like TIP3P would normally be masked by the strong stabilization of compact structures. Simulations performed with TIP4P-D or OPC may thus be less forgiving toward inaccuracies in the macromolecular force field. In an attempt to improve the description of protein–nucleic acid complexes, we have performed a reparameterization, using experimental data (Figure S6), of the nonbonded parameters of the phosphate group in DES-Amber, which we had not previously optimized (see SI for a full description of these reparameterization and validation efforts). We did not attempt the same reparameterization for DES-Amber SF1.0.

The reparameterized force field, which we term *DES-Amber 3.20* (from the value in Å of the  $\sigma$  parameter in the optimized LJ interaction between the phosphate oxygen atoms and water<sup>122–128</sup>), is comparably accurate to DES-Amber in simulations of nucleic acids (see, e.g., Figures S5, S7, and S9), and we tested it in simulations of protein–nucleic acid complexes by performing 50  $\mu$ s simulations of eight DNA–protein complexes (the six complexes listed above and two zinc fingers), as shown in Figure 8 and in Figures S13–20, and six RNA–protein complexes, as shown in Figure 9. We found that on the 50  $\mu$ s timescale investigated, all the complexes studied were stable and none dissociated (Figure 10), a clear improvement compared to using DES-Amber without the reparameterization. This finding suggests that DES-Amber 3.20 may be suitable for long-timescale simulations of systems containing both proteins and nucleic acids. (The relatively large RMSD of the 3bdn complex was mostly attributable to rigid pivot motions of the two large protein domains that are not involved in DNA binding and are thus free to move in



**Figure 10.** MD simulations of six protein–RNA complexes (top) and eight protein–DNA complexes (bottom) performed with the DES-Amber 3.20 (black) and DES-Amber (red). Backbone RMSD traces are reported as a function of simulation time. Flexible terminal residues and regions that were not resolved in the X-ray structures were omitted from the RMSD calculations. Each panel is labeled with the PDB entry of the starting structure.

solution, but the complex itself was stable.) Although it is possible that the improved performance of DES-Amber 3.20 might be the result of having two additional parameters compared to DES-Amber (specifically, two additional, explicit LJ pairs for nucleic acid–water interactions; see SI), the improvements observed in the description of protein–nucleic acid interactions were substantial, and we recommend the use of DES-Amber 3.20 in simulations of such systems.

## DISCUSSION

We have performed an optimization of the Amber DNA force field parameters using our previously reported nonbonded parameters for RNA as a starting point. We also optimized parameters for magnesium ions and structural zinc ions. The new DNA force field parameters are part of the DES-Amber force field, which also contains RNA and protein parameters, and were tested on both single- and double-stranded DNA systems. Comparison of DES-Amber simulation results to experiment indicated that, in most cases, the structure and local flexibility of both single- and double-stranded DNA systems were reproduced with a level of accuracy similar to state-of-the-art force fields, with the notable exception of the BII population, which appears to have been underestimated in the DES-Amber DNA duplex simulations.

Despite the adjustment of DNA-specific parameters, we observed that the stability of the DNA B-helix was underestimated by  $\sim 0.1$ – $0.2$  kcal mol<sup>-1</sup> per nucleobase. While it would be possible to correct this deficiency by further optimizing torsion parameters, we decided against this strategy, as it appeared that structures consistent with the B-helix conformation were already well-populated in single-stranded DNA. We thus suspect that some error in the balance of base stacking and hydrogen bonding or backbone electrostatic interactions may be responsible. We also observed some deviation from experiment in the local structure of CG base pairs, a deficiency resulting from the strong repulsion between the two guanine nucleobases on opposing strands, which induced local and global distortions in the helix. Such an imbalance is also characteristic of simulations performed with other Amber force fields,<sup>129,130</sup> and it is unclear if a simple reparameterization of nonbonded interactions and torsion angles would be sufficient to address it or if a more sophisticated functional form might be required.

Finally, we performed a set of simulations of protein–DNA and protein–RNA complexes, with somewhat mixed results. Some complexes appeared to be well-described, while others became unstable and dissociated during simulation. Discrepancies from experiment were also observed in previous simulations of a number of protein–protein complexes.<sup>8</sup> Based on the simulations presented here, this instability could not simply be ascribed to internal strain resulting from the instability of one of the biomolecular components, as in some cases of unstable complexes the isolated proteins or nucleic acids were structurally stable. To further investigate this issue, we performed an analysis of the backbone nonbonded interactions and attempted a reparameterization of the phosphate group based on experimental data (see SI). The resulting DES-Amber 3.20 force fields improved the description of the complexes while yielding results roughly equivalent to the DES-Amber force fields for systems containing nucleic acids only. To achieve these results, it was necessary to introduce two additional nonbonded pairs, which may in effect be compensating for the lack of an explicit polarization model; such adjustments may not be required in the context of a polarizable force field. Achieving an even higher degree of accuracy may require the use of more complex functional forms. Further studies will be required to fully address this issue.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.1c10971>.

Development of a generalized force field for the description of structural zinc ions compatible with DES-Amber; parameterization strategy for nonstandard residues; optimization of nucleic acid backbone nonbonded parameters against experimental data; assessment of the accuracy of DES-Amber 3.20 for DNA simulations; assessment of the accuracy of DES-Amber 3.20 for RNA simulations; simulations of short DNA oligomers and comparison to NMR J-couplings (Table S1); simulations of dsDNA reversible duplex formation (Table S2); simulations of dsDNA reversible duplex formation (Table S3); long DNA oligomers simulations (Table S4); deviations from experimental J-couplings for

simulations of ssRNA tetranucleotides (Table S5); magnitude of torsion corrections from refitting based on experimental NMR J-coupling data (Table S6); modified force field parameters for DES-Amber 3.20 (Table S7); simulations of dsRNA reversible duplex formation (Table S8); DES-Amber simulations of DNA duplexes (Table S9); DES-Amber 1.0 simulations of DNA duplexes (Table S10); DES-Amber 3.20 simulations of DNA duplexes (Table S11); Amber-bsc1 simulations of DNA duplexes (Table S12); Amber-OL15 simulations in TIP3P water of DNA duplexes (Table S13); Amber-OL15 simulations in OPC water of DNA duplexes (Table S14); average helical rise and twist in simulations of DNA duplexes (Table S15); dT40 simulations (Figure S1); simulated tempering simulations of DNA duplex formation performed with the Amber-bsc1 force field (Figure S2); simulated tempering simulations of DNA duplex formation performed with the DES-Amber SF1.0 force field (Figure S3); simulated tempering simulations of DNA duplex formation performed with the DES-Amber force field (Figure S4); simulated tempering simulations of DNA duplex formation performed with the DES-Amber 3.20 force field (Figure S5); comparison of calculated and experimental osmotic coefficients for DMP, purine, cytidine, and uridine (Figure S6); comparison of the disordered states of rU40 and dT40 simulated with different force fields with FRET data (Figure S7); MD simulations of the d(CGCGCGCGCGCG) Z-DNA dodecamer at KCl concentrations ranging between 0.1 and 1.5 M (Figure S8); MD simulations of a d(CGCGCGCGCGCG) Z-DNA dodecamer performed with the DES-Amber 3.20 force field (Figure S9); MD simulations of three complex RNA folds (Figure S10); MD simulations of six proteins containing structural zinc ions (Figure S11); MD simulations of tRNA (Figure S12); MD simulations of a leucine zipper/DNA complex (Figure S13); MD simulations of the Trp-repressor/operator complex (Figure S14); MD simulations of the P22 c2 repressor protein operator complex (Figure S15); MD simulations of CI repressor in complex with the OL2 operator half-site (Figure S16); MD simulations of ETV6 bound to DNA (Figure S17); MD simulations of the  $\lambda$ -repressor (Figure S18); MD simulations of the zinc finger of human CXXC5 in complex with CpG DNA (Figure S19); MD simulations of the ZIF268 zinc finger-DNA complex (Figure S20); MD simulations of Z-DNA in 1 M NaCl (Figure S21); and MD simulations of complex RNA folds (Figure S22) (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Stefano Piana** – D. E. Shaw Research, New York, New York 10036, United States; Phone: (212) 403-8165; Email: [Stefano.Piana-Agostinetti@DEShawResearch.com](mailto:Stefano.Piana-Agostinetti@DEShawResearch.com); Fax: (646) 873-2165

**David E. Shaw** – D. E. Shaw Research, New York, New York 10036, United States; Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States; [orcid.org/0000-0001-8265-](https://orcid.org/0000-0001-8265-)

5761; Phone: (212) 478-0260; Email: David.Shaw@DEShawResearch.com; Fax: (212) 845-1286

## Authors

Maxwell R. Tucker – D. E. Shaw Research, New York, New York 10036, United States

Dazhi Tan – D. E. Shaw Research, New York, New York 10036, United States

Michael V. LeVine – D. E. Shaw Research, New York, New York 10036, United States

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jpcc.1c10971>

## Notes

The authors declare the following competing financial interest(s): This study was conducted and funded internally by D. E. Shaw Research, of which D.E.S. is the sole beneficial owner and Chief Scientist, and with which all authors are affiliated.

The MD trajectories of the simulations described in this manuscript are available for noncommercial use through contacting [trajectories@deshawresearch.com](mailto:trajectories@deshawresearch.com). The force field parameters are available in Desmond format on GitHub: <https://github.com/DEShawResearch/viparr-ffpublic>. (The force field parameters are also available in Gromacs format through contacting [trajectories@deshawresearch.com](mailto:trajectories@deshawresearch.com).)

## ACKNOWLEDGMENTS

We thank Giovanni Bussi for helpful discussion on the importance of nonbonded backbone parameters, John L. Klepeis for helpful discussions and a critical reading of the manuscript, and Berkman Frank for editorial assistance. This study was conducted and funded internally by D. E. Shaw Research, of which D.E.S. is the sole beneficial owner and Chief Scientist and with which all authors are affiliated.

## REFERENCES

- (1) Ivani, I.; et al. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, *13*, 55–58.
- (2) Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J. Reference Simulations of Noncanonical Nucleic Acids with Different  $\chi$  Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506–2520.
- (3) Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham, T. E.; Otyepka, M.; Jurečka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339–2354.
- (4) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.
- (5) Hart, K.; Foppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348–362.
- (6) Liebl, K.; Zacharias, M. Tumuc1: A new accurate DNA force field consistent with high-level quantum chemistry. *J. Chem. Theory Comput.* **2021**, *17*, 7096–7105.
- (7) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E1346–E1355.
- (8) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a Force Field for the Simulation of Single-Chain

Proteins and Protein–Protein Complexes. *J. Chem. Theory Comput.* **2020**, *16*, 2494–2507.

(9) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.

(10) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149–8160.

(11) Klamt, A.; Schuurmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

(12) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: a general-purpose quantum chemistry program package. *Molpro. WIREs Comput. Mol. Sci.* **2012**, *2*, 242–253.

(13) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., III Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(14) Murphy, M. C.; Rasnik, I.; Cheng, W.; Lohman, T. M.; Ha, T. Probing Single-Stranded DNA Conformational Flexibility Using Fluorescence Spectroscopy. *Biophys. J.* **2004**, *86*, 2530–2537.

(15) Maestro Schrödinger Release 2018–4: *Maestro*, Schrödinger, LLC: New York, NY (2018).

(16) Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.

(17) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B.* **2015**, *119*, 5113–5123.

(18) K. J., Bowers et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, 43–43 (2006). IEEE: New York, NY.

(19) Pan, A. C.; Weinreich, T. M.; Piana, S.; Shaw, D. E. Demonstrating an Order-of-Magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems. *J. Chem. Theory Comput.* **2016**, *12*, 1360–1367.

(20) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(21) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(22) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics.* **1994**, *101*, 4177–4189.

(23) D. E., Shaw et al. Millisecond-Scale Molecular Dynamics Simulations on Anton. In *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)* (2009). ACM Press: New York, NY.

(24) Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. The relationship between proton-proton NMR coupling constants and substituent electronegativities—I. *Tetrahedron* **1980**, *36*, 2783–2792.

(25) Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; de Leeuw, H. P. M.; Altona, C. The relationship between proton-proton NMR coupling constants and substituent electronegativities. II—conformational analysis of the sugar ring in nucleosides and nucleotides in solution using a generalized Karplus equation. *Org. Magn. Reson.* **1981**, *15*, 43–52.

(26) Mellema, J.-R.; Pieters, J. M. L.; Marel, G. A.; Boom, J. H.; Haasnoot, C. A. G.; Altona, C. Sequence-dependent structural variation in single-helical DNA Proton NMR studies of d(T-A-T-A) and d(A-T-A-T) in aqueous solution. *Eur. J. Biochem.* **1984**, *143*, 285–301.

- (27) Olsthoorn, C. S. M.; Bostelaar, L. J.; Boom, J. H.; Altona, C. Conformational Characteristics of the Trinucleoside Diphosphate dApdApdA and Its Constituents from Nuclear Magnetic Resonance and Circular Dichroism Studies. Extrapolation to the Stacked Conformers. *Eur. J. Biochem.* **1980**, *112*, 95–110.
- (28) Mellema, J. R.; Haasnoot, C. A. G.; Van Boom, J. H.; Altona, C. Complete assignment and conformational analysis of a deoxyribotranucleotide, d(TAAT). A 360 and 500 MHz NMR study. *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.* **1981**, *655*, 256–264.
- (29) Mellema, J.-R.; Jellema, A. K.; Haasnoot, C. A. G.; Boom, J. H.; Altona, C. Conformational analysis of the single-helical DNA fragment d(T-A-A-T) in aqueous solution. The combined use of NMR proton chemical shifts and coupling constants obtained at 300 MHz and 500 MHz. *Eur. J. Biochem.* **1984**, *141*, 165–175.
- (30) Ts'o, P. O. P.; Cheng, D. M.; Frechet, D.; Kan, L. S. NMR Studies on Short Helices. In *Nucleic Acids: The Vectors of Life* (B., Pullman; J., Jortner, eds.), 201–206 (1983). D. Reidel Publ. Co.: Dordrecht.
- (31) Cheng, D. M.; Kan, L. S.; Leutzinger, E. E.; Jayaraman, K.; Miller, P. S.; Ts'o, P. O. P. Conformational study of two short pentadeoxyribonucleotides, CpCpApApG and CpTpTpGpG, and their fragments by proton NMR. *Biochemistry* **1982**, *21*, 621–630.
- (32) Neumann, J.-M.; Huynh-Dinh, T.; Kan, S. K.; Genissel, B.; Igolen, J.; Tran-Dinh, S. DNA Fragment Conformations: 1H-NMR and Relaxation Studies of 2'-Deoxyadenyl(3'-5')thymidyl(3'-5')deoxyguanosyl(3'-5')thymidine, d(A-T-G-T), in Neutral Aqueous Solution. *Eur. J. Biochem.* **1982**, *121*, 317–323.
- (33) Tran-Dinh, S.; Neumann, J.-M.; Huynh-Dinh, T.; Genissel, B.; Igolen, J.; Simonnot, G. DNA Fragment Conformations: 1H-NMR Studies of Helix-Coil Transition, Conformations and Dynamic Structures of the Self-Complementary Deoxyhexanucleotide d(A-C-A-T-G-T) in Aqueous. *Eur. J. Biochem.* **1982**, *124*, 415–425.
- (34) Hartog, J. H. J.; Altona, C.; Boom, J. H.; Marcelis, A. T. M.; Marel, G. A.; Rinkel, L. J.; Wille-Haezeleger, G.; Reedijk, J. Cis-Platinum Induced Distortions in DNA. Conformational Analysis of d(GpCpG) and cis-Pt(NH<sub>3</sub>)<sub>2</sub>{d(GpCpG)}, Studied by 500-MHz NMR. *Eur. J. Biochem.* **1983**, *134*, 485–495.
- (35) Tran-Dinh, S.; Neumann, J.-M.; Taboury, J.; Huynh-Dinh, T.; Renous, S.; Genissel, B.; Igolen, J. DNA Fragment Conformations. 1H-NMR Comparative Studies of Helix-Coil Transition and Conformation of d(C-A-C-G-T-G) and d(G-T-G-A-C). Influence of Helix Formation on Proton Chemical Shifts. *Eur. J. Biochem.* **1983**, *133*, 579–589.
- (36) Padrta, P.; Štefl, R.; Králík, L.; Židek, L.; Sklenář, V. Refinement of d(GCGAAGC) Hairpin Structure Using One- and Two-Bond Residual Dipolar Couplings. *J. Biomol. NMR* **2002**, *24*, 1–14.
- (37) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47–L49.
- (38) Northrup, S. H.; Hynes, J. T. The stable states picture of chemical reactions. I. Formulation for rate constants and initial condition effects. *J. Chem. Phys.* **1980**, *73*, 2700–2714.
- (39) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **2004**, *60*, 2126–2132.
- (40) SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci.* **1998**, *95*, 1460–1465.
- (41) Tjandra, N.; Tate, S.; Ono, A.; Kainosho, M.; Bax, A. The NMR Structure of a DNA Dodecamer in an Aqueous Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc.* **2000**, *122*, 6190–6200.
- (42) Wu, Z. Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and 31P chemical shift anisotropy. *J. Biomol. NMR* **2003**, *26*, 297–315.
- (43) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences.* **1981**, *78*, 2179–2183.
- (44) Rettig, M.; Germann, M. W.; Wang, S.; Wilson, W. D. Molecular Basis for Sequence-Dependent Induced DNA Bending. *ChemBioChem* **2013**, *14*, 323–331.
- (45) Johnson, C. N.; Spring, A. M.; Desai, S.; Cunningham, R. P.; Germann, M. W. DNA Sequence Context Conceals  $\alpha$ -Anomeric Lesions. *J. Mol. Biol.* **2012**, *416*, 425–437.
- (46) Barbic, A.; Zimmer, D. P.; Crothers, D. M. Structural origins of adenine-tract bending. *Proc. Natl. Acad. Sci.* **2003**, *100*, 2369–2373.
- (47) DiGabriele, A. D.; Steitz, T. A. A DNA Dodecamer Containing an Adenine Tract Crystallizes in a Unique Lattice and Exhibits a New Bend. *J. Mol. Biol.* **1993**, *231*, 1024–1039.
- (48) MacDonald, D.; Herbert, K.; Zhang, X.; Polgruto, T.; Lu, P. Solution structure of an A-tract DNA bend. *J. Mol. Biol.* **2001**, *306*, 1081–1098.
- (49) Bae, S.-H.; Cheong, H.-K.; Cheong, C.; Kang, S.; Hwang, D. S.; Choi, B.-S. Structure and Dynamics of Hemimethylated GATC Sites: Implications for DNA-SeqA Recognition. *J. Biol. Chem.* **2003**, *278*, 45987–45993.
- (50) Wang, F.; DeMuro, N. E.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. Base-Displaced Intercalated Structure of the Food Mutagen 2-Amino-3-methylimidazo[4,5-f]quinoline in the Recognition Sequence of the Nar I Restriction Enzyme, a Hotspot for –2 bp Deletions. *J. Am. Chem. Soc.* **2006**, *128*, 10085–10095.
- (51) Bhattacharyya, D.; et al. Flanking Bases Influence the Nature of DNA Distortion by Platinum 1,2-Intrastrand (GG) Cross-Links. *PLoS One* **2011**, *6*, No. e23582.
- (52) Julien, O.; Beadle, J. R.; Magee, W. C.; Chatterjee, S.; Hostetler, K. Y.; Evans, D. H.; Sykes, B. D. Solution Structure of a DNA Duplex Containing the Potent Anti-Poxvirus Agent Cidofovir. *J. Am. Chem. Soc.* **2011**, *133*, 2264–2274.
- (53) Stefl, R.; Wu, H.; Ravindranathan, S.; Sklenar, V.; Feigon, J. DNA A-tract bending in three dimensions: Solving the dA4T4 vs. dT4A4 conundrum. *Proc. Natl. Acad. Sci.* **2004**, *101*, 1177–1182.
- (54) Sathyamoorthy, B.; Shi, H.; Zhou, H.; Xue, Y.; Rangadurai, A.; Merriman, D. K.; Al-Hashimi, H. M. Insights into Watson–Crick/Hoogsteen breathing dynamics and damage repair from the solution structure and dynamic ensemble of DNA duplexes containing m1A. *Nucleic Acids Res.* **2017**, *45*, 5586–5601.
- (55) Ghosh, A.; et al. Indolicidin Targets Duplex DNA: Structural and Mechanistic Insight through a Combination of Spectroscopy and Microscopy. *ChemMedChem* **2014**, *9*, 2052–2058.
- (56) McAteer, K.; Aceves-Gaona, A.; Michalczyk, R.; Buchko, G. W.; Isern, N. G.; “Pete” Silks, L. A.; Miller, J. H.; Kennedy, M. A. Compensating bends in a 16-base-pair DNA oligomer containing a T<sub>3</sub>A<sub>3</sub> segment: A NMR study of global DNA curvature. *Biopolymers* **2004**, *75*, 497–511.
- (57) Renisio, J.-G. Pre-organized structure of viral DNA at the binding-processing site of HIV-1 integrase. *Nucleic Acids Res.* **2005**, *33*, 1970–1981.
- (58) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.
- (59) Joung, I. S.; Cheatham, T. E., III Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (60) Lu, X.-J.; Bussemaker, H. J.; Olson, W. K. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **2015**, gkv716.
- (61) Luo, Z.; Dauter, M.; Dauter, Z. Phosphates in the Z-DNA dodecamer are flexible, but their P-SAD signal is sufficient for structure solution. *Acta Crystallogr D Biol Crystallogr.* **2014**, *70*, 1790–1800.
- (62) Amrane, S.; Adrian, M.; Heddi, B.; Serero, A.; Nicolas, A.; Mergny, J.-L.; Phan, A. T. Formation of Pearl-Necklace Monomeric G-Quadruplexes in the Human CEB2S Minisatellite. *J. Am. Chem. Soc.* **2012**, *134*, 5807–5816.
- (63) Haider, S.; Parkinson, G. N.; Neidle, S. Crystal Structure of the Potassium Form of an Oxytricha nova G-quadruplex. *J. Mol. Biol.* **2002**, *320*, 189–200.

- (64) Asensio, J. L.; Brown, T.; Lane, A. N. Solution conformation of a parallel DNA triple helix with 5' and 3' triplex–duplex junctions. *Structure* **1999**, *7*, 1–11.
- (65) Keller, W.; König, P.; Richmond, T. J. Crystal Structure of a bZIP/DNA Complex at 2.2 Å: Determinants of DNA Specific Recognition. *J. Mol. Biol.* **1995**, *254*, 657–667.
- (66) Watkins, D.; Mohan, S.; Koudelka, G. B.; Williams, L. D. Sequence Recognition of DNA by Protein-Induced Conformational Transitions. *J. Mol. Biol.* **2010**, *396*, 1145–1164.
- (67) Frandsen, K. H.; Rasmussen, K. K.; Jensen, M. R.; Hammer, K.; Pedersen, M.; Poulsen, J.-C. N.; Arleth, L.; Lo Leggio, L. Binding of the N-Terminal Domain of the Lactococcal Bacteriophage TP901-1 CI Repressor to Its Target DNA: A Crystallography, Small Angle Scattering, and Nuclear Magnetic Resonance Study. *Biochemistry* **2013**, *52*, 6892–6904.
- (68) De, S.; Chan, A. C. K.; Coyne, H. J.; Bhachech, N.; Hermsdorf, U.; Okon, M.; Murphy, M. E. P.; Graves, B. J.; McIntosh, L. P. Steric Mechanism of Auto-Inhibitory Regulation of Specific and Non-Specific DNA Binding by the ETS Transcriptional Repressor ETV6. *J. Mol. Biol.* **2014**, *426*, 1390–1406.
- (69) Otwinowski, Z.; Schevitz, R. W.; Zhang, R.-G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. Crystal Structure of TRP Repressor/Operator Complex at Atomic Resolution. *Nature* **1988**, *335*, 321–329.
- (70) Stayrook, S.; Jaru-Ampornpan, P.; Ni, J.; Hochschild, A.; Lewis, M. Crystal structure of the  $\lambda$  repressor and a model for pairwise cooperative operator binding. *Nature* **2008**, *452*, 1022–1025.
- (71) Chao, J. A.; Lee, J. H.; Chapados, B. R.; Debler, E. W.; Schneemann, A.; Williamson, J. R. Dual modes of RNA-silencing suppression by Flock House virus protein B2. *Nat. Struct. Mol. Biol.* **2005**, *12*, 952–957.
- (72) Zhang, J.; Tan, D.; DeRose, E. F.; Perera, L.; Dominski, Z.; Marzluff, W. F.; Tong, L.; Hall, T. M. T. Molecular mechanisms for the regulation of histone mRNA stem-loop-binding protein by phosphorylation. *Proc. Natl. Acad. Sci.* **2014**, *111*, E2937–E2946.
- (73) Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C.-H.; Nagai, K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **1994**, *372*, 432–438.
- (74) Tan, D.; Zhou, M.; Kiledjian, M.; Tong, L. The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nat. Struct. Mol. Biol.* **2014**, *21*, 679–685.
- (75) Vasilyev, N.; Polonskaia, A.; Darnell, J. C.; Darnell, R. B.; Patel, D. J.; Serganov, A. Crystal structure reveals specific recognition of a G-quadruplex RNA by a  $\beta$ -turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E5391–E5400.
- (76) Wang, Y.; Opperman, L.; Wickens, M.; Hall, T. M. T. Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proceedings of the National Academy of Sciences*. **2009**, *106*, 20186–20191.
- (77) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schultz, M. MOLPRO version 2018.1, a package of ab initio programs. (2018).
- (78) Draper, D. E. RNA Folding: Thermodynamic and Molecular Descriptions of the Roles of Ions. *Biophys. J.* **2008**, *95*, 5489–5495.
- (79) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.* **2012**, *22*, 262–272.
- (80) Bergonzo, C.; Hall, K. B.; Cheatham, T. E. Divalent Ion Dependent Conformational Changes in an RNA Stem-Loop Observed by Molecular Dynamics. *J. Chem. Theory Comput.* **2016**, *12*, 3382–3389.
- (81) Åqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **1990**, *94*, 8021–8024.
- (82) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493–1502.
- (83) Mamatkulov, S.; Fyta, M.; Netz, R. R. Force fields for divalent cations based on single-ion and ion-pair properties. *J. Chem. Phys.* **2013**, *138*, No. 024505.
- (84) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733–2748.
- (85) Romer, R.; Hach, R. tRNA Conformation and Magnesium Binding. A Study of Yeast Phenylalanine-Specific tRNA by a Fluorescent Indicator and Differential Melting Curves. *Eur. J. Biochem.* **1975**, *55*, 271–284.
- (86) Quigley, G. J.; Teeter, M. M.; Rich, A. Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. *Proceedings of the National Academy of Sciences*. **1978**, *75*, 64–68.
- (87) Shi, H.; Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: A classic structure revisited. *RNA* **2000**, *6*, 1091–1105.
- (88) Grotz, K. K.; Cruz-León, S.; Schwierz, N. Optimized Magnesium Force Field Parameters for Biomolecular Simulations with Accurate Solvation, Ion-Binding, and Water-Exchange Properties. *J. Chem. Theory Comput.* **2021**, *17*, 2530–2540.
- (89) Ohtaki, H.; Radnai, T. Structure and dynamics of hydrated ions. *Chem. Rev.* **1993**, *93*, 1157–1204.
- (90) Caminiti, R.; Licheri, G.; Piccaluga, G.; Pinna, G. X-ray diffraction study of MgCl<sub>2</sub> aqueous solutions. *J. Appl. Crystallogr.* **1979**, *12*, 34–38.
- (91) Bleuzen, A.; Pittet, P. A.; Helm, L.; Merbach, A. E. *Magn. Reson. Chem.* **1997**, *35*, 765–773.
- (92) Goedken, E. R.; Marqusee, S. Co-crystal of *Escherichia coli* RNase HI with Mn<sup>2+</sup> Ions Reveals Two Divalent Metals Bound in the Active Site. *J. Biol. Chem.* **2001**, *276*, 7266–7271.
- (93) Katayanagi, K.; Miyagawa, M.; Matsushima, M.; Ishikawa, M.; Kanaya, S.; Nakamura, H.; Ikehara, M.; Matsuzaki, T.; Morikawa, K. Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.* **1992**, *223*, 1029–1052.
- (94) Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proceedings of the National Academy of Sciences*. **2012**, *109*, 799–804.
- (95) Goddard, N. L.; Bonnet, G.; Krichevsky, O.; Libchaber, A. Sequence Dependent Rigidity of Single Stranded DNA. *Phys. Rev. Lett.* **2000**, *85*, 2400–2403.
- (96) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127.
- (97) Plumridge, A.; Meisburger, S. P.; Pollack, L. Visualizing single-stranded nucleic acids in solution. *Nucleic Acids Res.* **2016**, *45*, No. e66.
- (98) Yoshizawa, S.; Kawai, G.; Watanabe, K.; Miura, K.; Hirao, I. GNA Trinucleotide Loop Sequences Producing Extraordinarily Stable DNA Minihairpins. *Biochemistry* **1997**, *36*, 4761–4767.
- (99) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix–Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (100) Best, R. B.; Mittal, J. Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *J. Phys. Chem. B* **2010**, *114*, 14916–14923.
- (101) Dans, P. D.; Ivani, I.; Hospital, A.; Portella, G.; González, C.; Orozco, M. How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.* **2017**, *45*, 4217.
- (102) SantaLucia, J.; Allawi, H. T.; Seneviratne, P. A. Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability<sup>†</sup>. *Biochemistry* **1996**, *35*, 3555–3562.
- (103) Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **2007**, *35*, W43–W46.

- (104) Varnai, P. DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res.* **2004**, *32*, 4269–4280.
- (105) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of  $\alpha/\gamma$  Conformers. *Biophys. J.* **2007**, *92*, 3817–3829.
- (106) Zgarbová, M.; Jurečka, P.; Banáš, P.; Havrila, M.; Šponer, J.; Otyepka, M. Noncanonical  $\alpha/\gamma$  Backbone Conformations in RNA and the Accuracy of Their Description by the AMBER Force Field. *J. Phys. Chem. B.* **2017**, *121*, 2420–2433.
- (107) Dans, P. D.; Danišáň, L.; Ivani, I.; Dršata, T.; Lankaš, F.; Hospital, A.; Walther, J.; Pujagut, R. L.; Battistini, F.; Gelpí, J. L.; Lavery, R. Long-timescale dynamics of the Drew–Dickerson dodecamer. *Nucleic Acids Res.* **2016**, *44*, 4052–4066.
- (108) Dršata, T.; Pérez, A.; Orozco, M.; Morozov, A. V.; Šponer, J.; Lankaš, F. Structure, Stiffness and Substates of the Dickerson–Drew Dodecamer. *J. Chem. Theory Comput.* **2013**, *9*, 707–721.
- (109) Tian, Y.; Kayatta, M.; Shultis, K.; Gonzalez, A.; Mueller, L. J.; Hatcher, M. E.  $^{31}\text{P}$  NMR Investigation of Backbone Dynamics in DNA Binding Sites. *J. Phys. Chem. B.* **2009**, *113*, 2596–2603.
- (110) Sugiyama, H.; Kawai, K.; Matsunaga, A.; Fujimoto, K.; Saito, I.; Robinson, H.; Wang, A. H.-J. Synthesis, Structure and Thermodynamic Properties of 8-Methylguanine-Containing Oligonucleotides: Z-DNA under Physiological Salt Conditions. *Nucleic Acids Res.* **1996**, *24*, 1272–1278.
- (111) Feigon, J.; Wang, A. H.-J.; van der Marel, G. A.; Van Boom, J. H.; Rich, A. A one- and two-dimensional NMR study of the B to Z transition of  $(\text{m}^5\text{dC-dG})_3$  in methanolic solution. *Nucl Acids Res.* **1984**, *12*, 1243–1263.
- (112) Tran-Dinh, S.; Taboury, J.; Neumann, J. M.; Huynh-Dinh, T.; Genissel, B.; Langlois d'Estaintot, B.; Igolen, J. Proton NMR and circular dichroism studies of the B and Z conformations of the self-complementary deoxyhexanucleotide d(m<sup>5</sup>C-G-C-G-m<sup>5</sup>C-G): mechanism of the Z-B-coil transitions. *Biochemistry* **1984**, *23*, 1362–1371.
- (113) Pohl, F. M.; Jovin, T. M. Salt-induced co-operative conformational change of a synthetic DNA: Equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.* **1972**, *67*, 375–396.
- (114) Zgarbová, M.; Šponer, J.; Jurečka, P. Z-DNA as a touchstone for additive empirical force fields and a refinement of the alpha/gamma DNA torsions for AMBER. *J. Chem. Theory Comput.* **2021**, *17*, 6292–6301.
- (115) Fadrná, E.; Špačková, N.; Sarzyńska, J.; Koča, J.; Orozco, M.; Cheatham, T. E.; Kulinski, T.; Šponer, J. Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. *J. Chem. Theory Comput.* **2009**, *5*, 2514–2530.
- (116) Gkionis, K.; Kruse, H.; Platts, J. A.; Mládek, A.; Koča, J.; Šponer, J. Ion Binding to Quadruplex DNA Stems. Comparison of MM and QM Descriptions Reveals Sizable Polarization Effects Not Included in Contemporary Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1326–1340.
- (117) Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Cryst.* **1963**, *16*, 907–916.
- (118) MacKerell, A. D.; Nilsson, L. Molecular Dynamics Simulations of  $\phi$ . *Curr. Opin. Struct. Biol.* **2008**, *18*, 194–199.
- (119) Krepl, M.; Havrila, M.; Stadlbauer, P.; Banas, P.; Otyepka, M.; Pasulka, J.; Stefl, R.; Šponer, J. Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein–RNA Complexes? *J. Chem. Theory Comput.* **2015**, *11*, 1220–1243.
- (120) Watkins, D.; Hsiao, C.; Woods, K. K.; Koudelka, G. B.; Williams, L. D. P22 c2 Repressor–Operator Complex: Mechanisms of Direct and Indirect Readout †. *Biochemistry* **2008**, *47*, 2325–2338.
- (121) Phan, A. T.; et al. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* **2011**, *18*, 796–804.
- (122) Yoo, J.; Aksimentiev, A. Improved Parameterization of Amine–Carboxylate and Amine–Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J. Chem. Theory Comput.* **2016**, *12*, 430–443.
- (123) Yoo, J.; Aksimentiev, A. New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 8432–8449.
- (124) Yoo, J.; Aksimentiev, A. Improved Parametrization of  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Mg}^{2+}$  Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *J. Phys. Chem. Lett.* **2012**, *3*, 45–50.
- (125) Mlýnský, V.; Kührová, P.; Kühn, T.; Otyepka, M.; Bussi, G.; Banáš, P.; Šponer, J. Fine-Tuning of the AMBER RNA Force Field with a New Term Adjusting Interactions of Terminal Nucleotides. *J. Chem. Theory Comput.* **2020**, *16*, 3936–3946.
- (126) Kührová, P.; Mlýnský, V.; Zgarbová, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Šponer, J.; Banáš, P. Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **2019**, *15*, 3288–3305.
- (127) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. Simulations of Anionic Lipid Membranes: Development of Interaction-Specific Ion Parameters and Validation Using NMR Data. *J. Phys. Chem. B.* **2013**, *117*, 10183–10192.
- (128) Luo, Y.; Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *J. Phys. Chem. Lett.* **2010**, *1*, 183–189.
- (129) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E.; Giudice, E.; Lankaš, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. II: Sequence Context Effects on the Dynamical Structures of the 10 Unique Dinucleotide Steps. *Biophys. J.* **2005**, *89*, 3721–3740.
- (130) Hospital, A.; et al. BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.* **2016**, *44*, D272–D278.