Article

C−B−A Test of DNA Force Fields

Ivan A. Strelnikov, Natalya A. Kovaleva, Artem P. Klinov, and Elena A. Zubova*

Cite This: *ACS Omega* 2023, 8, 10253−10265

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The DNA duplex may be locally strongly bent in complexes with proteins, for example, with polymerases or in a nucleosome. At such bends, the DNA helix is locally in the noncanonical forms A (with a narrow major groove and a large amount of north sugars) or C (with a narrow minor groove and a large share of BII phosphates). To model the formation of such complexes by molecular dynamics methods, the force field is required to reproduce these conformational transitions for a naked DNA. We analyzed the available experimental data on the B−C and B−A transitions under the conditions easily implemented in modeling: in an aqueous NaCl solution. We selected six DNA duplexes which conformations at different salt concentrations are known reliably enough. At low salt concentrations, poly(GC) and poly(A) are in the B-form, classical and slightly shifted to the A-form, respectively. The duplexes ATAT and GGTATACC have a strong and salt concentration dependent bias toward the A-form. The polymers poly(AC) and poly(G) take the C- and A-forms, respectively, at high salt concentrations. The reproduction of the behavior of these oligomers can serve as a test for the balance of interactions between the base stacking and the conformational flexibility of the sugar−phosphate backbone in a DNA force field. We tested the AMBER bsc1 and CHARMM36 force fields and their hybrids, and we failed to reproduce the experiment. In all the force fields, the salt concentration dependence is very weak. The known B-philicity of the AMBER force field proved to result from the B-philicity of its excessively strong base stacking. In the CHARMM force field, the B-form is a result of a fragile balance between the A-philic base stacking (especially for G:C pairs) and the C-philic backbone. Finally, we analyzed some recent simulations of the LacI-, SOX-4-, and Sac7d-DNA complex formation in the framework of the AMBER force field.
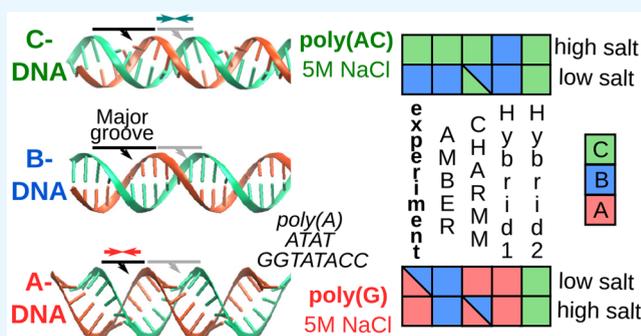
1. INTRODUCTION

Understanding the mechanisms of protein binding to DNA is a critical task not only in theoretical biophysics but also in medicine. To date, there appeared a lot of reports on modeling the processes of DNA−protein binding or dissociation, protein diffusion, and conformational dynamics in complexes (see reviews[1,2]). Usually, the researchers choose the AMBER force field. This force field allows simulating complexes even with highly deformed DNA (with kinks as a result of protein intercalation), for example, in cases of Sox protein,[3−5] Hbb,[6,7] Sac7d,[8−10] LacI,[11] and the others.[12,13] The CHARMM force field is being used much less frequently, with an important exception of nucleosomes.[14−16]

However, the reliability of the results of such modeling is limited by the accuracy of the phenomenological force fields used for protein, DNA and their interactions with each other and with the solvent. Both the AMBER and CHARMM DNA force fields are known to have several drawbacks. Under the conditions when the B-form of DNA is to be stable, on time intervals greater than a microsecond, a long DNA molecule loses its duplex structure in the framework of the CHARMM36 force field, and passes into an intermediate form between the

B- and A- forms in the framework of the AMBER bsc0 force field.[17] In the 2010s, several torsion angles were edited (primarily ϵ/ζ) to improve the representation of the BII conformation of phosphates[18,19] (AMBER bsc1 and CHARMM36). Recently, in 2021, a correction of α/γ angles has been proposed to more adequately reproduce Z-DNA in the framework of AMBER force field (OL21).[20]

More serious shortcomings for practical purposes are too weak base pairing and too strong base stacking in both the AMBER and CHARMM force fields, with the pairing weaker in the CHARMM force field, and the stacking stronger in the AMBER force field (see, for example, the discussion in ref 21). Of nonbonded interactions, DNA-ions interactions (in particular, with Mg^{2+}) also do not correspond to the experimental ones. The appropriate corrections for the
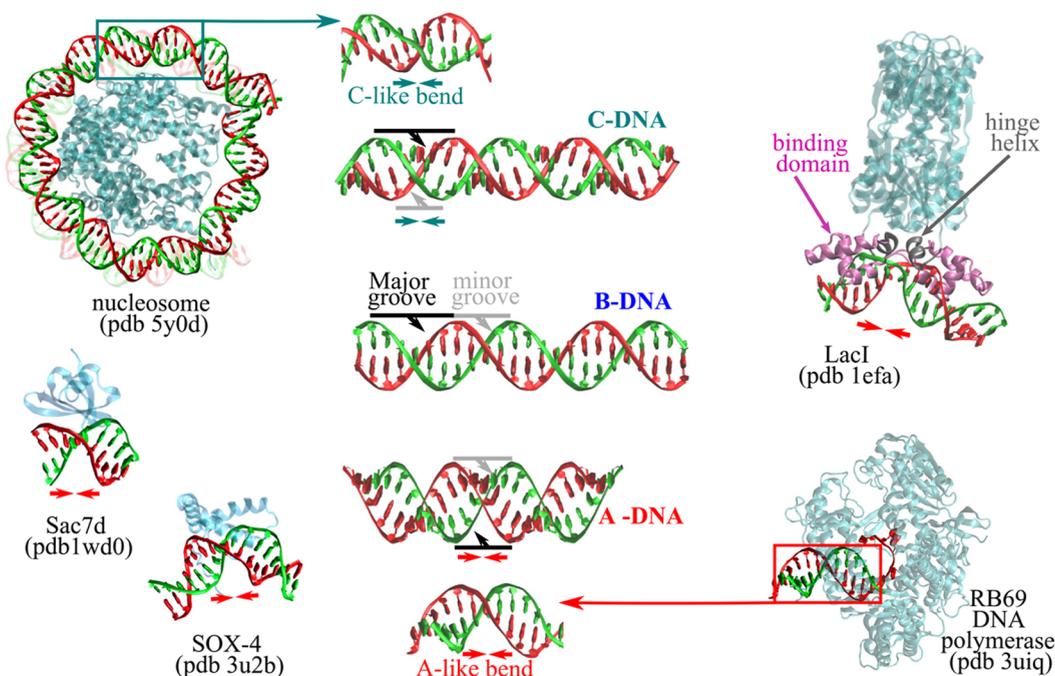
10253

Figure 1. Comparison of the ideal C-, B-, and A-forms of DNA (constructed using the w3DNA 2.0 server,[31,32] visualization: VMD program[33,34]). For C-DNA, we show the variant with 100% of BII phosphate conformations. We marked with green and red arrows the local narrowing of the DNA minor and major grooves on the C-like and A-like DNA bends on a nucleosome and in the complex with a polymerase, respectively. We also show several more protein complexes with A-like DNA bends.

AMBER force field has resulted in a new force field DES-Amber[22] with adjusted values for a number of van der Waals parameters and for some atomic partial charges. In the framework of the DES-Amber force field, one mainly uses a 'native' water model, TIP4P-D, since the TIP3P water is known to be a poor solvent for proteins.[23]

In this regard, one has to remember that it was the TIP3P water model that was used to derive the values of partial charges on DNA atoms in the CHARMM force field, and these values have never been revised. In the AMBER force field, a new set of partial atomic charges for DNA and RNA has been devised.[24] In order to amend excessively strong electrostatic interactions in DNA−protein complexes, a new Tumuc1 force field has been developed,[25] with new partial atomic charges and parameters of the valence bonds, angles, and torsion angles.

One can do the opposite and weaken too strong DNA−protein interactions by amending their van der Waals, rather than electrostatic, interactions. The CUFIX (or NBFIX) correction[26] adjusts the parameters σ of the Lennard-Jones potential between the atoms of the DNA phosphate group and some atoms on amino acids. The analogous corrections have been offered also for interactions with ions. It turned out that for each pair of atoms, the parameter σ has to be chosen independently of other pairs (combination rules do not apply). The CUFIX correction provides the correct values for the diffusion coefficients of proteins along DNA in both the cases of fast and slow diffusion.[7]

However, the most difficult task, currently unsolved, is the adequate modeling of A-DNA. DNA takes the A-form locally in complexes with endonucleases, transcription factors, and polymerases.[27,28] This is a form with a narrow major groove and sugars adopting the north conformation (C3′endo). The

geometry of A-DNA is opposite to that of C-DNA, in which the minor groove is narrow (see Figure 1). The signs of the Roll and Slide parameters are also opposite in the A- and C-forms. In C-DNA, a significant fraction (over 40%) of phosphates adopts the BII conformation.[29] DNA in nucleosomes has about 20% of phosphates in the BII conformation (see, for example, file 1kx5.pdb[30]).

The A- and C- forms of DNA are competitors under conditions with a low dielectric constant of a medium, or a low water activity, or a small available volume. These are the conditions in hydrophobic protein cavities. In vitro, the competition can be demonstrated in an experiment when one increases the alcohol proportion in a water solution of DNA.[35] At 78% of ethanol, the natural DNA (calf thymus) takes the A-form. With 0.1 M NaCl, the addition of methanol causes the transition to the C-form. In complexes with proteins, DNA can locally take the noncanonical A- and C-forms or be strongly bent.[36] In these cases, it is important that, for conditions corresponding to the experiment, a force field adequately reproduces the transition to noncanonical conformations of torsion angles on the DNA sugar−phosphate backbone, including the transition to noncanonical deoxyribose conformations (usually north, C3′endo).

Unfortunately, the transition to the A-form is not adequately modeled by both the AMBER and CHARMM force fields. They are considered to be B- and A-philic, correspondingly.[17,37−40] Since the balance of interactions in the DNA molecule depends on many factors, changing any of them can easily amend the situation in one particular case. It is more important to find out exactly which parameters of the force fields do not correspond to reality.

The answer to this question requires a direct comparison of the results of simulations and experiments on the B to C and B

to A transitions. In an adequate force field, the DNA molecule is to take both the C- and A- forms (depending on the base sequence) under the experimental conditions. For the test, one has to choose the experiments under the simplest possible conditions, without interactions of DNA with complex molecules (alcohol; spermine; drugs; polymers, including DNA: in fibers or in crystals). The simplest system for molecular modeling is a DNA aqueous solution with an addition of salt. We have analyzed (Section 2.1) the available experimental data on such transitions for NaCl salt (obtained by Raman and IR spectroscopy[41,42] and circular dichroism[43]).

Based on this analysis, we have formulated the "C−B−A" test for the balance of interactions between the base stacking and the backbone flexibility in DNA force fields. We have tested (Section 2.2) the AMBER bsc1 and CHARMM36 force fields and their "hybrids": the fields with potentials for the sugar−phosphate backbone from one force field and for the base stacking interactions from another field. For both the AMBER and CHARMM force fields, we have located the weak points not allowing one to reproduce the experimentally observed DNA behavior. Finally, in light of the results obtained, we have reviewed (Section 2.3) the possible influence of force-field choice on the results of modeling DNA complexes with three different proteins (LacI,[11] Sox-4,[5] and Sac7d[9]), where the DNA is bent toward the major groove and locally takes the A-form.

2. RESULTS AND DISCUSSION

2.1. C-DNA and A-DNA in Aqueous Solution: Experiment. Circular dichroism (CD) measurements show that, with a large amount of additional salt (4−6 M NaCl) in aqueous solution, natural DNA (calf thymus) takes the C-form (strictly speaking, not completely).[35] The regular polymer poly[d(AC)]·poly[d(GT)] (poly(AC)) has a CD spectrum very close to that of the calf thymus DNA. Poly(AC) also passes into the C-form upon addition of 4.5 M NaCl.[44] According to spectral measurements, the polymer poly[d-(GC)]₂ (poly(GC)) passes into the Z-form,[45] but at a low salt concentration (less than 0.2M), poly(GC) is in the classical B-form. Under the same conditions, the polymer poly(dA)·poly(dT) (poly(A)) also takes a slightly deformed B-form with some share of sugars in the north conformation (C3′endo).[46] The poly(A) double helix has not been observed in any other form under any conditions. A bias toward the A-form is observed in the (double stranded) oligomers d(ATAT)₂ (ATAT) and d(GGTATACC)₂ (GGTATACC)[47,48] (when they are stable: at low temperatures). Only the polymer poly(dG)·poly(dC) (poly(G)) demonstrates a complete transition to the A-form upon addition of 4 M NaCl.[45] Lowering the temperature facilitates this transition. Moreover, even almost without additional salt (0.03 M), 30% of this polymer is in the A-form.[45]

Although the picture of the conformational transitions described above seems clear, quantitative estimates can currently be obtained only from IR or Raman spectra. In these spectra, there are definite bands—markers of the A-form.[41,42] Unfortunately, C markers do not allow making reliable quantitative estimates. As we will see below, the A markers cannot always provide good accuracy as well. The accuracy can be increased by using 2D IR spectra.[49−51] However, such data are not available for the required DNA sequences in the required frequency range.

Certain vibrational modes of a DNA nucleotide change their frequencies when the sugar conformation changes. Among these modes are, for example, some deformation modes of the pyrimidine ring on nucleobases. Other modes change their frequencies when the geometry of the sugar−phosphate backbone changes. When some of DNA nucleotides are in the A-form, and the rest is in the B-form, one may estimate the fraction of the A-form as the ratio of the integral intensity of the A-form marker band to the sum of the intensities of the A- and B- form bands.

The transition from B-DNA to A-DNA is accompanied by the shift in the frequency of the mode OPOsym (symmetrical stretching vibrations in the phosphodiester group (C−O3′−P−O5′−C)) from 790 to 807 cm⁻¹. For DNA molecules consisting only of G:C pairs (for example, poly(G) and poly(GC)), this band can be used—but only at high salt concentrations—to fairly accurately estimate the degree of transition of the backbone to the A-form (see Table. 1 and

Table 1. DNA in Aqueous NaCl Solution[a]

DNA	salt, M	Markers[41]	Form
polyGC 20 °C[45]	0.2	OPOsym: 17%A g(664): 0% C3	B
polyA 30 °C[46]	0.2	OPOsym: 39%A? t(777): 22% C3? [a+t](642): 14% C3 t(1239): 53% C3	~70% B ~30% A
ATAT[47] −2 °C	0.15	OPOsym: 46%A? t(777): 25% C3? [a+t](642): 11% C3 t(1239): 55−62% C3	~65% B ~35% A
	5	OPOsym: 49%A? t(777): 55% C3? [a+t](642): 23% C3 t(1239):52−61% C3	~50% A ~50% B
GGTATACC[48] 0−15 °C	5−6	OPOsym: 34%A? t(777): 20% C3? g(1318): 52% C3 t(1239): 55% C3	~60% B ~40% A
polyG 20 °C[45]	0.01	OPOsym: 44%A g(664): 24% C3	B?
	0.2	OPOsym: 62%A g(664): 73% C3	~60% A ~40% B
polyG 30 °C[52]	0.03	OPOsym: 31%A? g(664): 0% C3 g(1318): 38% C3	B
	4	OPOsym: 78%A g(664): 100% C3 g(1318): 78% C3	A
10 °C[52]	1	OPOsym: 60%A g(664): 65% C3	~60% A ~40% B

[a]A: A-form according to backbone mode OPOsym. C3: C3′endo sugar pucker; (t): thymine, (g): guanine, (a): adenine; in brackets: wavenumbers for the bands in cm⁻¹.

Figure 2). Unfortunately, in this region of the spectrum, in addition to this A marker, there is also a band at a frequency of 780 cm⁻¹ corresponding to a certain deformation mode of the cytosine ring. It is possible to separate this band from the OPOsym band only if the salt concentration is high (see Figure 2). In other cases, this marker does not allow quantifying the fraction of the A-form without additional assumptions.
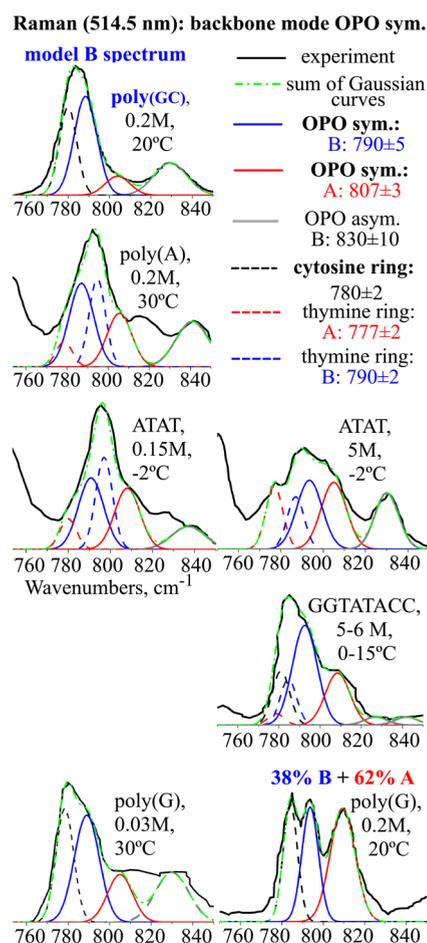
Figure 2. Raman spectra in the region of OPOsym (symmetrical stretching in the phosphodiester group) for some regular DNA polymers and oligomers (extracted from refs 45, 47, 52). A conjectured Gaussian peak fitting is done by us.

The situation is even worse for oligomers containing A:T pairs. There is a band of thymine similar to that of cytosine at 780 cm^{-1}, but the frequency of this band shifts after the B−A transition. It is possible that the intensity of this band also changes, as does the intensity of the OPOsym band. Therefore, the optical spectra (both IR and Raman) do not allow quantifying the fraction of the backbone that has passed into the A-form; only qualitative evaluation is possible. Therefore, for the majority of oligomers and polymers in the Table 1, we denote the estimates for the OPOsym mode and for the mode of thymine at 777 cm^{-1} with a question mark. In Figure 2, we give only one of many possible curve fitting of the experimental spectrum by a sum of Gaussians, it cannot be considered reliable.

Another problem with A:T pairs is the coincidence of the frequencies (642 cm^{-1}) of the pyrimidine ring breathing modes for adenine and thymine. It is impossible to separate these bands (and estimate their widths and intensities in different DNA forms) from Raman or IR spectra. Therefore, in Table 1 we present the total estimate for this marker. One cannot help but notice that the estimate by marker [a+t](642) is always very low as compared with the other markers, and can hardly be considered reliable. Guanine and, most likely, also cytosine have a similar vibrational mode, but in Raman spectra there is only the band (g(664) in Table 1) corresponding to

guanine vibrations (at a frequency of 664 cm^{-1} for C3′endo and 682 cm^{-1} for C2′endo). For the polymer poly(G), the estimate by g(664) is close to the estimate by backbone marker OPOsym (in the cases when this estimate can be considered reliable).

The polymer poly(GC) is believed to be in the B-form at low salt concentrations (0−0.2M). In the spectrum of poly(A), there is a noticeable shoulder (at 807 cm^{-1}) indicating the presence of north sugars. The spectrum of poly(A) also has a large peak at 820 cm^{-1} (not marked out in Figure 2). This band with different heights is present on all the Raman spectra from DNA samples. In the work,[46] the authors have even introduced a special designation, $a2$, for a DNA form with certain (noncanonical) torsion angles. The band at 820 cm^{-1} was a marker of this new form. Later, the form $a2$ disappears from the works on IR and Raman spectroscopy of DNA. There is no generally accepted assignment of this band.

The oligomer ATAT exists as a duplex only at low temperatures. From the broadening of the band near 790 cm^{-1}, one can definitely say that the addition of salt shifts the conformation toward the A-form, but one cannot give reliable quantitative estimates. The same can be said about the oligomer GGTATACC.

The most A-philic polymer is poly(G). Even at the lowest salt concentration (0.03 M) and a rather high temperature of 30 °C, the band OPOsym has a shoulder corresponding to the A-form. At room temperature and salt concentration 0.2 M, the backbone of this polymer is 62% in the A-form.

Let us compare the fractions of north sugar conformations calculated using different markers. The high-frequency thymine deformation mode at 1239 cm^{-1} gives approximately the same estimate of ∼55% for all the cases presented in Table 1. However, judging by the backbone marker OPOsym, the polymer poly(A) is in the (slightly deformed) B-form, and the oligomer ATAT undoubtedly contains a significant fraction of the A-form at the high salt concentration. In the works cited in Table 1, there is only one case when the estimate by this marker differs from ∼55%. It is a crystal GGTATACC in the A-form with t(1239): ∼67%. A similar band for the guanine deformation mode at 1318 cm^{-1} behaves quite differently. For the polymer poly(G), the estimate g(1318) practically repeats the estimate by the backbone marker OPOsym (in the cases when this estimate can be considered reliable).

As we have seen, the use of markers [a+t](642) and t(1239) cannot be considered justified without further analysis of their reliability. It should also be added that the appearance of a triplet instead of a very weak band of deoxyribose vibrations in the B-form at frequencies ∼860−900 cm^{-1} in both the Raman and IR spectra is also considered a marker of the transition to the A-form (see, for example, the work[52] and the review of markers[42]). But this band can not be used to quantify the fraction of the A-form. The PO$_2^-$ (OP1-P-OP2) vibrations at frequencies ∼1090 cm^{-1} and ∼1230 cm^{-1} also show some sensitivity to the B−A transition. However, it has recently been demonstrated for DNA films[53] that this sensitivity is not to the DNA conformation, but to hydration. Perhaps that is why the band of symmetric PO$_2^-$ vibrations at the frequency ∼1090 cm^{-1} sometimes is even indicated as being insensitive to the A−B transition.[42]

The practical conclusion from the experimental data reviewed above is as follows. Two DNA sequences, poly(AC) and poly(G), behave oppositely in aqueous NaCl solution at a high salt concentration. The first passes into the C-form (the

Table 2. Conformations of PolyGC, PolyA, ATAT, GGTATACC at Low Salt[a]

DNA, exp.% BII	AMBER	CHARMM	AMBER - *CHARMM's bases*	CHARMM - *AMBER's bases*
polyGC 34% BII	**B**	**B**	**B**	**C**
	g: 0(±1)% C3	g: 3(±4)% C3	g: 4(±4)% C3	g: 0.1(±0.6)% C3
	c: 2(±3)% C3	c: 16(±9)% C3	c: 23(±14)% C3	c: 3(±4)% C3
	36(±5)% BII	24(±7)% BII	22(±7)% BII	47(±9)% BII
	R = 1, S = 0	R = 5, S = 0.3	R = 4, S = −0.1	R = 2, S = 0.5
	MG = 18.6(±0.5)	MG = 17.1(±0.6)	MG = 17.1(±0.8)	MG = 18.2(±0.4)
	mG = 11.6(±0.5)	mG = 13.8(±0.5)	mG = 13.4(±0.5)	mG = 12.3(±0.5)
polyA 6% BII	**B**	**B**	**B**	**C/B**
	2(±2)% C3	a:4(±4)% C3	a:25(±14)% C3	a:1(±2)% C3
		t:23(±8)% C3	t:36(±14)% C3	t:6(±5)% C3
	12(±4)% BII	18(±7)% BII	6(±3)% BII	38(±9)% BII
	R = 1, S = −0.5	R = 6, S = 0.2	R = 4, S = −0.8	Roll = 6, S = 0.5
	MG = 19.1(±0.6)	MG = 16.6(±0.6)	MG = 17.4(±0.9)	MG = 17.5(±0.5)
	mG = 10.6(±0.4)	mG = 13.2(±0.5)	mG = 13.1(±0.7)	mG = 12.3(±0.5)
ATAT (−2 °C) 5% BII	**B**	**C**	**B**	**C**
		terminal H bonds		*terminal H bonds*
		broken aft. sev. ns		*off and on*
	1(±5)% C3	5(±9)% C3	12(±11)% C3	3(±6)% C3
	20(±12)% BII	76(±17)% BII	7(±10)% BII	57(±19)% BII
	R = 0, S = −0.5	R = 1, S = 1.2	R = 3, S = −0.4	R = 0, S = 1.3
	mG = 11(±2)	mG = 15(±1)	mG = 12(±2)	mG = 14(±1)
GGTATACC 17% BII	**B**	**C**	**B**	**C**
	1(±3)% C3	3(±5)% C3	a,t: 19(±19)% C3	1(±3)% C3
			g: 30(±29)% C3	
			c: 4(±10)% C3	
	25(±8)% BII	61(±17)% BII	11(±8)% BII	62(±18)% BII
	R = 0, S = −0.4	R = 1, S = 1.5	R = 4, S = −0.7	R = 1, S = 1
	MG = 18(±1)	MG = 17.1(±0.7)	MG = 17(±2)	MG = 17.8(±0.7)
	mG = 10(±1)	mG = 13(±2)	mG = 13(±1)	mG = 12(±1)

[a]Almost no concentration dependence in the simulations, see the text. Initially, all the oligomers were in the B-form. Unless otherwise indicated, the transition to the final conformation occurred within 5 ns after the system preparation (see Section 4.2). Temperature: 300 K (unless otherwise stated), additional salt: ~0.15M NaCl, water: TIP4P/long;[54] ions: Joung-Cheatham.[55] For each DNA sequence, the average experimental value of the fraction of BII phosphate conformations in solutions is given (by ^{31}P NMR chemical shift[56]). C3: north sugar pucker (C3'endo) of thymine (t), cytosine (c), guanine (g), or adenine (a). If only one number is given, the share of north sugars did not depend on the type of nucleobase. The unit of measure for Roll (R) is degree; Slide (S), Major groove width (MG) and minor groove width (mG)—Angstrom.

minor groove narrows), the second into the A-form (the major groove narrows). Strictly speaking, one may limit oneself to these two sequences to test a DNA force field for the balance between the flexibility of phosphates and sugar puckers. Other oligomers can serve as reference points. The polymers poly(GC) and poly(A) at low salt concentrations are the variants of the B-form (the classical form and a form with a shift to A-DNA). For the A-philic oligomers ATAT and GGTATACC, the fraction of the A-form is to increase with the addition of salt.

2.2. C-DNA and A-DNA in Aqueous Solution: Modeling. We have simulated all the six DNA sequences discussed in the previous section: poly(GC) and poly(A) (20 bp duplexes); ATAT and GTATACC; poly(AC) and poly(G) (20 bp duplexes) at two salt concentrations (see Table 5 in Section 4). We have tested four force fields: AMBER (bsc1), CHARMM (36), and two hybrid fields, in which the sugar—phosphate backbone was modeled in the framework of one field, and the base stacking—in the framework of the other force field (for details, see Section 4.1). In addition, we have tested the Tumuc1 force field for two sequences, poly(AC) and poly(G). In all the cases, we used the same water and ion models (TIP4P/long[54] water, Joung-Cheatham ions,[55] see Section 4.2). The duration of the observed conformational

changes and the stability of the final conformations are discussed in Section 4.3.

Table 2 and Figure 3 show the simulation results for the first four oligomers at a low salt concentration. The first defect of all the tested force fields is that none of them (except for the hybrid 'AMBER—CHARMM's bases' in the case of the oligomer ATAT) showed any dependence on salt concentration for the oligomers ATAT and GGTATACC. This contradicts the experiment (higher salt concentration—larger fraction of the A-form). All the system parameters at the high salt proved out to be the same as at the low salt, so they are not listed in Table 2 and in Figure 3.

In the AMBER force field, at any salt concentration, these four oligomers retain the B-form, the oligomer ATAT exhibiting an excess of BII phosphate conformations (at low temperature). In the CHARMM force field, the oligomers GGTATACC and ATAT are in the C-form even at low salt. The bias toward the C-form is the defect of the sugar—phosphate backbone of the CHARMM force field, because this bias is absent in the hybrid force field 'AMBER—CHARMM's bases' but is present in the force field 'CHARMM—AMBER's bases'.

On the other hand, in the CHARMM force field the base stacking is A-philic. Indeed, in the hybrid force field
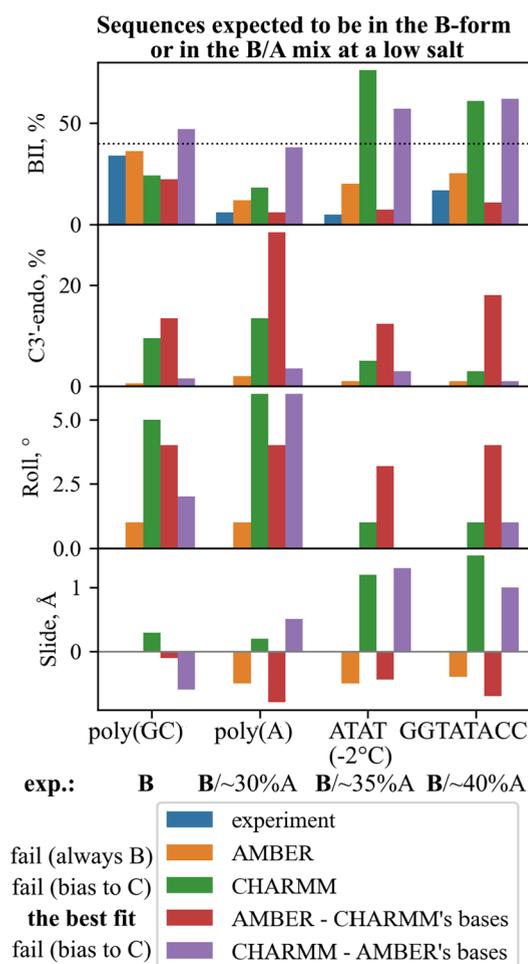
Figure 3. Graphical representation of Table 2. The experimental data on the fraction of the A-form are taken from Table 1.

As the salt concentration increases, the fraction of the A-form decreases (also in contrast to the experiment), while the fraction of BII conformations increases.

In the force field 'AMBER—CHARMM's bases', the polymer poly(G) is in the A-form, regardless of the amount of salt, which confirms the bias of the base stacking (rather than the backbone) in the CHARMM force field to the A-form. In the CHARMM force field at the high salt, the narrowing of the minor groove and the partial transition to the B- and then to the C-form (the fraction of phosphates in the BII conformation increases from 3 to 16%) is provided by the sugar—phosphate backbone. Indeed, in the force field 'CHARMM—AMBER's bases' (in which the AMBER bases do not tend to A-like base stacking), even at the low salt, the polymer poly(G) is in the C-form.

The bias of the DNA backbone in the CHARMM force field toward the C-form may be provided by the interaction of charges on the backbone with ions in solution. At the high salt, a large amount of sodium ions appear in the minor groove of poly(G) near the atom O4′ of the sugar ring (direct contact) and near the atom N2 of guanine. Perhaps the reason is that the atom O4′ in the CHARMM force field has a charge of −0.5e versus −0.3691e in the AMBER force field. These additional sodium ions in the minor groove, attracting phosphate groups, seem to provide narrowing of the minor groove and the shift in the balance toward the C-form.

In summary, the base stacking of the AMBER force field is excessively B-philic with a slight bias toward the C-form. Contrary to experiment, there is no A-DNA for the polymer poly(G) in this force field. However, it reproduces the transition to the C-form for the polymer poly(AC). In the CHARMM field, the base stacking provides the bias to the A-form for poly(G), while the backbone is excessively C-philic. It can be said that B-DNA in the CHARMM force field is the result of a fluid balance between the shift of the base stacking toward the A-form and the backbone toward the C-form. In contrast, B-DNA in the AMBER force field is a stable inflexible construct due to the B-philic excessively strong base stacking.

The main difference in stacking in the AMBER and CHARMM force fields results from a very large difference in partial charges on the nucleobases. It is with the assignment of a set of charges that the construction of a force field for DNA begins. This is the most difficult task, and the developers of the two force fields solved it in fundamentally different ways. The stacking of the CHARMM force field turned out to be strongly A-philic, and therefore, in order to obtain the B-form, the sugar—phosphate backbone had to be made C-philic.

We did not test the DES-AMBER force field because the developers themselves found that the fraction of BII-conformations was greatly underestimated. Accordingly, the DES-AMBER is guaranteed to fail the C—B part of the C—B—A test.

In the framework of the Tumuc1 force field, we modeled the behavior of the polymers poly(AC) and poly(G) for two salt concentrations: 0 M and 4 M. There was no concentration dependence. For poly(AC), we observed transitions between B- and C-forms ($1(\pm2)$% C3, $32(\pm6)$% BII, Roll = 2, Slide = −0.2, MG = 19.7, mG= 11.2), as in the CHARMM force field at low salt. The polymer poly(G) took the C-form instead of the A-form, although with a not very large fraction of BII phosphate conformations ($1(\pm1)$% C3, $37(\pm6)$% BII, Roll = 4, Slide = 0, MG = 19.2, mG = 11.9). The helix parameters turned out to be close to the parameters in the AMBER force

'AMBER—CHARMM's bases', the amount of north sugars is much higher for all four oligomers and is in the best agreement with the experiment. Moreover, this force field reproduces an increase in the fraction of north sugars with an increase in the salt concentration for the oligomer ATAT. At salt concentration 5 M, the replicas (we had four of them) pass from one form to another, being in the B-form more often. So one can say that the result is a mixture of the B- and A-DNA. On average, the fraction of north sugars is $24(\pm16)$% (at $5(\pm9)$% BII phosphate conformations; Roll = 5°, Slide = −0.7 Å).

For the polymers poly(AC) and poly(G), a salt dependence does take place (see Tables 3 and 4, and Figure 4). The AMBER force field reproduces the transition of the polymer poly(AC) from the B- to the C- form with the addition of salt but does not reproduce the bias of the polymer poly(G) to the A-form, nor the complete transition to the A-form at high salt concentration. In the AMBER force field, this polymer retains the B-form at any salt. Moreover, a decrease in temperature increases the bias of poly(G) in the opposite direction: toward the C-form.

In the CHARMM force field, the polymer poly(AC) is never in the B-form, even with zero additional salt. There are successive transitions from B to C and vice versa. At the high salt concentration, a complete transition to the C-form occurs. The polymer poly(G) is completely in the A-form already at a salt of 0.15 M (which does not correspond to the experiment).

Table 3. Dependence of the Conformation of the Polymer Poly(AC) on NaCl Concentration in Different Force Fields[a]

DNA, exp.% BII	AMBER	CHARMM	AMBER - *CHARMM's* bases	CHARMM - *AMBER's* bases
polyAC 0 M	**B**	alternate B−C and C−B trans.	**B**	**C**
	after 20 ns failed B−C trans. attempt			*after 15 ns relax.*
	2(±2)% C3	6(±5)% C3	13(±7)% C3	0(±1)% C3
	27(±5)% BII	40(±10)% BII	18(±5)% BII	65(±10)% BII
	R = 3, S = −0.3	R = 4, S = 0.8	R = 5, S = −0.4	R = 1, S = 1
	MG = 19.3(±0.6)	MG = 18.1(±0.5)	MG = 18.3(±0.8)	MG = 18.8(±0.3)
	mG = 11.8(±0.4)	mG = 13.0(±0.7)	mG = 13.1(±0.5)	mG = 11.0(±0.7)
polyAC 0.1 M 23% BII	**C**/∼3 ns jumps to **B**, after 30 ns in B, 5 ns B−C trans.	alternate B−C and C−B trans.	**B**	**C**
				after 15 ns in B, 5 ns B−C trans.
	1(±1)% C3	6(±6)% C3	16(±6)% C3	1(±1)% C3
	37(±5)% BII	41(±12)% BII	16(±5)% BII	68(±10)% BII
	R = −1, S = 0	R = 3, S = 0.8	R = 5, S = −0.4	R = 0, S = 1.1
	MG = 19.8(±0.5)	MG = 18.2(±0.5)	MG = 18.1(±0.8)	MG = 18.8(±0.3)
	mG = 10.4(±0.5)	mG = 13(±1)	mG = 13.1(±0.4)	mG = 10.8(±0.5)
polyAC 4.5 M	**C**	**C**	**B**	**C**
	after 13 ns B−C trans., 22 ns relax.	*after 45 ns B−C trans.*		*after 20 ns B−C trans.*
	1(±1)% C3	1(±1)% C3	17(±5)% C3	0(±1)% C3
	46(±2)% BII	65(±7)% BII	21(±6)% BII	85(±8)% BII
	R = −3, S = 0.2	R = −3, S = 1.8	R = 3, S = −0.3	R = −1, S = 1.6
	MG = 20.1(±0.3)	MG = 18.0(±0.4)	MG = 18.3(±0.7)	MG = 18.5(±0.3)
	mG = 9.4(±0.2)	mG = 8.8(±0.5)	mG = 12.6(±0.6)	mG = 9.4(±0.5)

[a]In the experiment, one observes the transition from the B-form to the C-form with increasing salt. Simulation parameters and notation are the same as in Table 2.

Table 4. Dependence of the Conformation of the Polymer Poly(G) on NaCl Concentration in Different Force Fields[a]

DNA, exp.% BII	AMBER	CHARMM	AMBER—CHARMM's bases	CHARMM—AMBER's bases
polyG 0.15 M 42% BII	**B**	**A**	**A**	**C**
		after 30 ns B−A trans.		
	2(±2)% C3	g: 93(±6)% C3	g: 87(±6)% C3	1(±1)% C3
		c: 82(±7)% C3	c: 34(±16)% C3	
	30(±6)% BII	3(±3)% BII	1(±2)% BII	58(±10)% BII
	R = 3, S = −0.3	R = 9, S = −1.6	R = 5, S = −1.6	R = 3, S = 0.9
	MG = 19.1(±0.7)	MG = 14(±1)	MG = 16(±2)	MG = 17.1(±0.4)
	mG = 12.1(±0.4)	mG = 15.9(±0.2)	mG = 14.9(±0.3)	mG = 12.2(±0.5)
polyG 4 M	**B**	**A/B** mix	**A**	
		after 35 ns trans.		
	3(±3)% C3	g: 59(±11)% C3	g: 86(±6)% C3	
		c: 54(±13)% C3	c: 42(±14)% C3	
	26(±6)% BII	16(±7)% BII	1(±1)% BII	
	R = 3, S = −0.4	R = 7, S = −0.8	R = 5, S = −1.6	
	MG = 18.3(±0.6)	MG = 15(±1)	MG = 16(±1)	
	mG = 11.7(±0.5)	mG = 15.5(±0.3)	mG = 14.9(±0.3)	
polyG 1 M, 10 °C	**B/C**			**B/C**
	1(±2)% C3			1(±1)% C3
	36(±5)% BII			40(±6)% BII
	R = 3, S = −0.1			R = 3, S = 0.6
	MG = 18.9(±0.6)			MG = 18.2(±0.4)
	mG = 11.8(±0.4)			mG = 12.5(±0.4)

[a]In the experiment, one observes the full transition to the A-form with increasing salt. Simulation parameters and notation are the same as in Table 2.

field. Thus, the Tumuc1 force field turned out to be excessively C-philic.

2.3. DNA Force Field and Modeling of DNA−Protein Complexes. Let us consider some simulation results on the formation of DNA−protein complexes in the case of a strong A-like DNA bending (toward the major groove, see Figure 1). If the DNA molecule, in the framework of the used force field, is not able to adopt the required conformation under experimental conditions, then the complex may not form, even in the case of excessively strong DNA−protein interactions. On the other hand, the complex may not form for other reasons, including technical ones related to other shortcomings of the force field. But the researcher may erroneously conclude that the binding mechanism is a
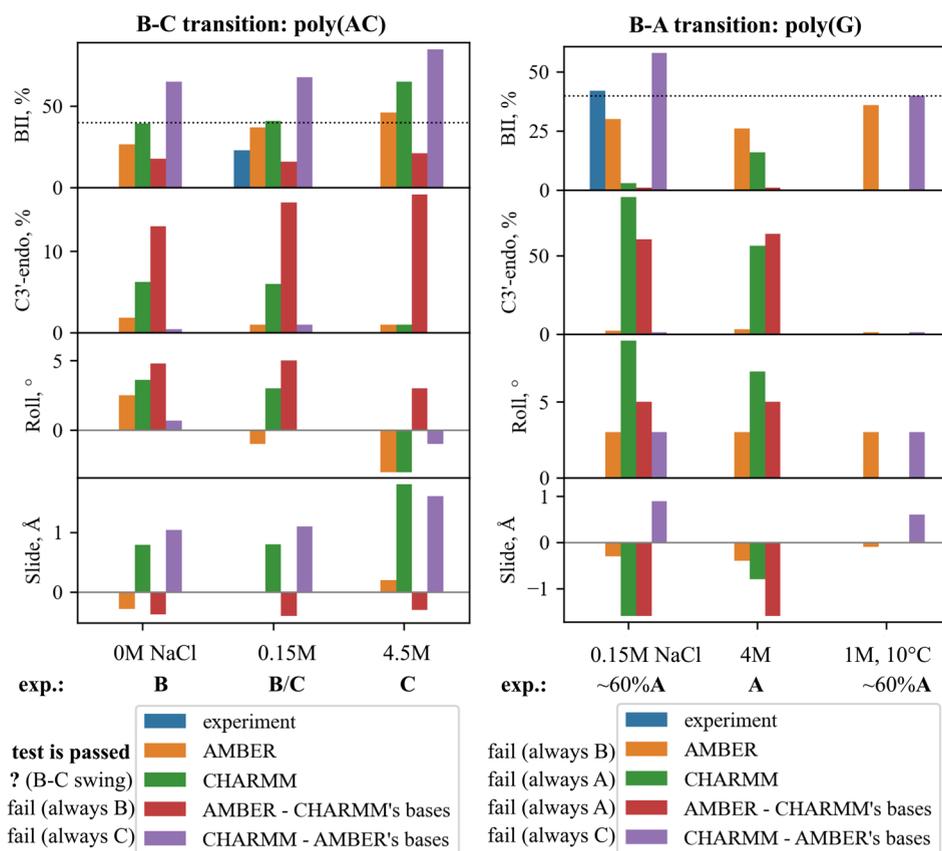
Figure 4. Graphical representation of Tables 3 and 4. The experimental data on the fraction of the A-form are taken from Table 1.

conformational selection (a protein binds only to a certain noncanonical form of DNA), while in reality an induced fit takes place (a protein induces the needed change in DNA conformation during or after binding, see review[1]).

Let us compare the results of modeling of the DNA/LacI,[11] DNA/SOX-4,[5] and DNA/Sac7d[9] complexes. All the authors use the AMBER force field (variants bsc1 in ref 11 and bsc0 in refs 5 and 9.) with TIP3P water. The bsc1[18] modification increases the fraction of BII phosphate conformations (underestimated in bsc0) and facilitates the transition to the C-form.

In the experiment, in the process of the complex formation, all three the proteins LacI, SOX-4, and Sac7d bind to the minor groove and, expanding it, form an A-like bend toward the major groove (see Figure 1). DNA locally takes the A-form. In the case of the protein LacI, the binding (and presumably the target recognition) also occurs in the major groove.

In the modeling, the protein Sac7d (albeit at an elevated temperature of 338 K) quickly (within tenths of a nanosecond) penetrates the DNA minor groove, moves along it, and induces a bend on the target, showing classical induced fit within a time interval less than one microsecond. The protein SOX-4 also demonstrates fairly fast (within 5 ns at 300 K) nonspecific binding. However, for this protein at a lower temperature—and, perhaps, due to too strong DNA−protein interactions in the AMBER force field—the transition from the nonspecific to specific binding seems impossible to observe by ordinary molecular dynamics. The authors do not do this and study the specific binding mechanism with the use of the metadynamics method.

It should be noted that, in most works on DNA−protein complex formation, the authors use enhanced sampling techniques: steered molecular dynamics,[4] metadynamics,[5,10] umbrella sampling,[13] adaptive biasing force,[3] and meta-eABF[6] calculations. These approaches make it possible to build free energy profiles and compare the obtained energy of complex formation with experiment. But one can not follow the process of complex formation in real time. In case of insufficient number or poor choice of reaction coordinates, one may face a difficulty in finding the real pass to the global minimum. The discrepancies may also result from an inadequate balance of interactions in other parts of the system (for example, incorrect DNA−protein binding energy, inappropriate water model, and so on, see Section 1).

In the case of the protein LacI, after a microsecond ordinary MD calculation, the authors failed to obtain the native state of the DNA−protein complex with an A-like bend. Contrary to the proteins SOX-4 and Sac7d, the transition to this state is impeded by a steric clash between LacI and the straight DNA. Because of the known B-philicity of the AMBER force field, the authors hypothesized that the DNA molecule has to be already bent before the complex formation (conformational selection). However, there are two other factors (specifics of the complex formation and the features of the force field not related to DNA) that could contribute to the negative result of the simulation.

There are experimental data[57] that two "hinge" alpha-helices, bound with the DNA minor groove in the LacI/DNA complex, are disordered in a free and nonspecifically bound protein. More exactly, these data refer to a truncated protein, the

"headpiece", consisting of the hinge helices and the binding domains (see Figure 1). But, since the affinity of the headpiece for DNA is the same as of the whole LacI protein, the headpiece is believed to be an adequate model system for binding of the whole protein. In the simulations by ref 11, the hinge helices in a free (whole) protein are almost ordered. In the CHARMM force field (also with the TIP3P water), isolated hinge helices in solution are disordered.[58] But they are ordered as a part of a free LacI headpiece,[58] contrary to experiment. Such an excessive stability of the hinge helices may result from using the TIP3P water model which is a poor solvent for proteins.[23] The same may be true also for the modeling of the whole LacI protein in ref 11.

The ordered state of the hinge helices may not facilitate, but, on the contrary, hinder the formation of the complex. Indeed, the folded hinge helices cannot penetrate the minor groove at the early stages of binding. So they cannot destroy the spine of hydration[59] in the minor groove, which stabilizes the B-form and prevents binding as well as bending. Without additional research, it is impossible to say what exactly—B-philic DNA in the AMBER force field, inappropriate water model, both together, or something else in addition (as a high barrier for the binding)—caused the negative result of modeling in ref 11.

In the cases of SOX-4 and Sac7d, such a complicated rearrangement of the protein is not needed for the complex formation. Therefore, excessively strong DNA−protein interactions may outweigh the DNA rigidity in the AMBER force field. As a result, the protein Sac7d easily induces the required DNA bend. And, for the native DNA/SOX-4 complex, there is a corresponding (global) minimum and a not very high barrier to pass from the (local) minimum for the nonspecifically bound complex.

Thus, when analyzing results of modeling the formation of DNA−protein complexes, it is necessary to take into account both the specifics of the possible mechanism of binding of a particular protein to DNA, and the features of the force fields used. Unfortunately, the choice of the force field is almost never substantiated in the corresponding works. In the case of modeling the complex with the protein LacI, it would be better to choose the CHARMM force field for the DNA. In it, the DNA duplex is more flexible and, for sequences with a large number of G:C pairs, more A-philic. This choice would allow eliminating at least one of the uncertainty factors in the analysis of the obtained simulation results.

3. CONCLUSION

The DNA duplex in the DNA−protein formation process often locally bends (nucleosome, complexes with polymerases) and passes into the C (narrow minor groove, negative Roll, positive Slide) or A (narrow major groove, positive Roll, negative Slide) forms. The B−C transition is accompanied by the switch of a significant fraction (more than ∼40%) of phosphates from the BI to the BII conformation, and the B−A transition is accompanied by the switch of sugars from the south (C2′endo) to the north (C3′endo) conformation. To study the formation of protein complexes with bent DNA by molecular dynamics modeling, the used DNA force field is needed to adequately reproduce both the conformational transitions, which requires appropriate testing.

In the present work, we have analyzed the available experimental data on the B−C and B−A transitions under the conditions most conveniently implemented in molecular dynamics modeling: in an aqueous NaCl solution. Unfortu-

nately, from the currently available spectroscopic data (Raman and IR), one cannot in most cases extract reliable quantitative estimates of the fraction of north sugars or of the fraction of the DNA backbone in the A conformation. For the B−C transition, only circular dichroism data are reliable. And it is also currently impossible to extract the exact fraction of the BII conformations of phosphates from them. However, there are DNA sequences that can be used to test DNA force fields.

We have selected six DNA oligomers that respond differently to an increase in salt concentration. At a low salt concentration, two polymers, poly(GC) and poly(A), take the B-form, classical and slightly deformed with some fraction of north sugars, respectively. The oligomers ATAT (at a very low temperature, when it is stable) and GGTATACC (also at low temperature) have an appreciable bias toward the A-form. The fraction of the A-form increases with increasing salt concentration. At a high salt concentration 4−5M, the polymers poly(AC) and poly(G) take the C- and A- forms, respectively.

Modeling these six DNA duplexes at the two (low and high) salt concentrations is a good test for the balance between the base stacking and the backbone (phosphate groups and deoxyriboses) conformational mobility in DNA force fields. We have tested the two most commonly used force fields AMBER bsc1 and CHARMM36. To find out how exactly the balance is upset in each case, we also had to test the hybrid force fields, with backbone interactions taken from one of the force fields and the base stacking taken from another.

We showed that the "B-philicity" of the AMBER force field is a consequence of the "B-philicity" of its base stacking. The stacking is excessively strong, and it does not allow the polymer poly(G) to pass to the A-form. The transition to the C-form for the polymer poly(AC) is possible in the framework of the AMBER force field. In the CHARMM force field, the balance between the interactions in the DNA duplex is fundamentally different. Namely, the B-form is the result of a fragile balance between the A-philic base stacking (especially for G:C base pairs) and the C-philic backbone. This equilibrium is strongly shifted toward the C-form with the predominance of A:T pairs in the DNA sequence, and toward the A-form with the predominance of G:C pairs. Therefore, in the CHARMM force field, the polymer poly(G) is always in the A-form, and the fraction of sugars in the north conformation for poly(GC) and poly(A) approximately coincide (in the experiment, poly(A) is more A-philic). The question on the balance between the base stacking and the backbone conformational flexibility in the real DNA duplex remains open.

In both the force fields, the interactions between the DNA and the ions make much smaller contribution to the overall balance of interactions in the system as compared to the experiment. In simulation of the oligomers ATAT and GGTATACC the experimental dependence on salt concentration is not observed. The influence of ions in both the force fields is especially weak under the conditions of the B−A transition. The fraction of north sugars in poly(G) does not depend on the salt concentration in the framework of neither force field (except for the wrong dependence in the CHARMM force field); while the B−C transition for poly(AC) is reproduced in the AMBER force field, and almost reproduced in the CHARMM force field. Due to the "dynamic" nature of the B-form in the CHARMM force field, at low salt concentrations the polymer poly(AC)

incessantly passes between the B- and C- forms, the C-form having too large percentage of BII phosphate conformations.

Thus, we have found a discrepancy in the DNA conformational flexibility (transitions to the C- and A-forms) between modeling in the framework of the AMBER and CHARMM force fields and the experimental data. A researcher choosing a force field might want to take into account the revealed features of the DNA duplex in these two commonly used force fields, along with the specifics of the possible mechanism of the complex formation, especially if the formation requires A-like bending of the DNA duplex. For example, in the case of the LacI/DNA complex, the CHARMM force field would be a better choice than the AMBER force field. In the CHARMM force field, the DNA duplex is more flexible and, with the predominance of G:C pairs, more A-philic. However, the reliability of the results can only be ensured by a force field that passes the C−B−A test.

4. METHODS: DETAILS OF MD CALCULATIONS

4.1. Force Fields and Their Hybrids, Simulation Parameters.

We simulated DNA oligomers in water with ions using the LAMMPS[60,61] package in the framework of the force fields AMBER bsc1[18] and CHARMM36[19] and their hybrids, four force fields in total. For some oligomers, the Tumuc1 force field was also tested.[25] In the hybrid force fields, the bonded interactions belonged to the first force field (for example, AMBER in the force field 'AMBER═CHARMM's bases'). The nonbonded interactions in the sugar−phosphate backbone belonged to the same force field, while the van der Waals parameters and partial electric charges on the bases were replaced by their counterparts from the other force field (CHARMM in the force field 'AMBER—CHARMM's bases'). Thus, the base stacking (van der Waals and Coulomb interactions between bases) was modeled in the framework of the other force field. We have compensated the difference in total charges on the bases (they are neutral in the CHARMM and negative in the AMBER) by a change in the charge on the hydrogen atom H1′. In such a splicing of the force fields, the effective potential of base rotation around the glycosidic bond χ turns out to be hybrid. Indeed, the potential depends on hybrid van der Waals and Coulomb interactions between the atoms on the sugar ring (with the parameters of the first force field) and on the bases (with the parameters of the second force field). We have plotted potential energy curves for rotation around the glycosidic bond χ for all four bases bound to the deoxyribose locked in the south conformation in a vacuum. It proved out that these curves lie between the corresponding curves for the CHARMM and AMBER force fields in the area of the *anti-* conformation (see as an example Figure 5 for adenine).

For calculation of the electrostatic interactions, we used the PPPM (particle−particle particle-mesh Ewald)[62] method with an accuracy of six significant digits (the dimensionless LAMMPS parameter 10^{-5}). We chose to cut off the short-range part of the potential at the distance optimal for the calculation speed (8 Å). In the AMBER force field, we calculated the van der Waals forces with an accuracy of 10^{-5} kcal/mol/Å (i.e., we considered the force equal to zero at a distance larger than 8−20 Å depending on the type of atoms). In the CHARMM force field, we applied this approach only to ions and water (to their interactions with each other and with the DNA). The van der Waals interactions between the DNA
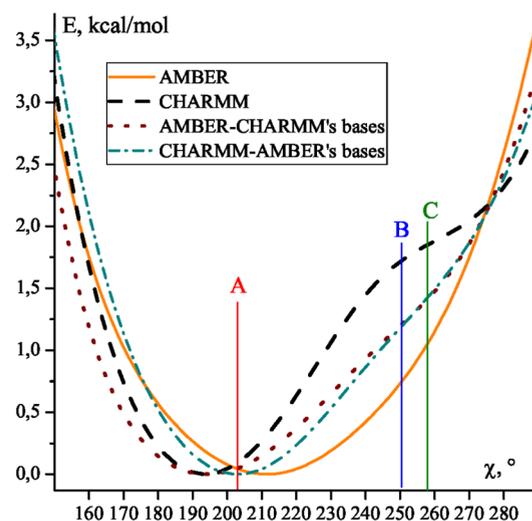

Figure 5. Potential energy for rotation around the glycosidic bond χ for adenine in a vacuum (deoxyribose in south conformation). We marked the values of χ in the A-, B-, and C-forms.

atoms smoothly vanished in the range from 12 to 13 Å, regardless of the type of atoms.

We used the Langevin thermostat and the Nose-Hoover barostat (pressure of 1 atm) with a damping parameter of one picosecond. We restrained X−H valence bonds and H−X−H valence angles (X is any atom) with 10^{-5} kcal/mol tolerance by the SHAKE algorithm.[63] The integration step was equal to 2 fs.

4.2. Initial and Boundary Conditions, the Problem of Terminal Base Pairs; Choice of Models for Water and Ions; Helix parameters.

We set up the initial state of the system as follows. We generated the DNA molecule in the B-form by the code of the D.Case group.[64] We placed sodium and chloride ions at random positions around the DNA no closer than 3 Å to any other ion or DNA atom. Randomly oriented water molecules were placed in positions with oxygen atom not closer than 2.5 Å to any other atom in the system. Then we corrected the atom positions by energy minimization using the conjugate-gradient method (until the force vector norm became less than 10^3 kcal/mol/Å). After that, we launched one MD run at a temperature $T = 300$ K for 10 ps in the NVT ensemble and then another run at the same temperature for 10 ps in the NPT ensemble. One could start the MD relaxation of the system from this state, but the cloud of ions usually comes to equilibrium much later than other parts of such systems. In order to speed up this process, we performed an additional MD run with the fixed DNA molecule in the NPT ensemble for 2 ns. For the first nanosecond, the thermostat temperature for water molecules was 300 K, and for ions, it was 500 K. During the next nanosecond, we gradually cooled the ions (and the water if needed) down to the simulation temperature (300 K, 283.15 K, or 271.15 K). The productive runs were carried out in the NPT ensemble.

In long-term all atom simulations, one observes too often and too long openings of the terminal base pairs[65,66] creating defective duplexes and impeding the cooperative B−A transition. To avoid the fraying, we linked the bases in the terminal base pairs by a parabolic potential. More exactly, we replaced the energy of one of the hydrogen bonds of the base pair (the central one: N1−H3 on A:T and N3−H1 on G:C) with a parabola. In an unrestrained simulation, the nitrogen

and hydrogen atoms are at a distance of ~1.7−2.2 Å. The parabola approximates the sum of their van der Waals and Coulomb interactions in this region. For long oligomers, such restraints on the terminal base pairs should have little effect on the result of simulations. Because the B−C transition is noncooperative, the fraction of BII phosphate conformations on the central part of the duplexes is very close for duplexes with restrained and unrestrained terminal base pairs.[66]

We simulated the tetramer ATAT without the restraints on the terminal base pairs. It is known that at room temperature this oligomer is unstable and decomposes into two strands. At the experimental temperature of −2 °C, when the oligomer is intact, in the force fields AMBER and 'AMBER—CHARMM's bases', the hydrogen bonds did not break during the simulation (55−219 ns in total from 4 replicas). In the CHARMM force field, the terminal hydrogen bonds broke after a few nanoseconds. In the case of the force field 'CHARMM—AMBER's bases' (with stronger AMBER's base stacking), the breaks occurred less frequently and healed in most cases. However, in both the cases, we calculated the average values of the parameters Roll and Slide only on the central step of the oligomer ATAT.

In all the DNA force fields we used the same TIP4P/long water[54] and the same Joung-Cheatham ions.[55] Due to the specifics of the choice of partial charges in the CHARMM force field, the water model could have a significant impact on the calculation results. Therefore, we carried out an additional simulation with the inherent models for ions and water (file toppar_water_ions.str in CHARMM36). We simulated the oligomer GGTATACC with unrestrained hydrogen bonds on the terminal base pairs at salt concentration 4.3M. The replacement of water and ions led to a decrease in the fraction of BII phosphate conformations, but only of 20 percentage points from ~70 to ~50%. The fraction of north sugars increased from ~2 to ~8%. Both the changes were in the correct direction, but the observed values are still far from the experiment (17% BII and ~(20−50)% C3′endo at 5 M NaCl).

The computational cell was cubic, with periodic boundary conditions. The edge of the cell was about 15 Å greater than the axis of the extended oligomer in the B-form. We carried out the calculations for the oligomers with counterions, and in weak and strong salt solutions, with concentrations of approximately 0.1−0.15 M and 4−5M. Table 5 lists the total simulation time and the characteristics of the modeled systems depending on the salt concentration. In some cases, we collected statistics from several (usually 4) replicas.

We calculated the helix parameters Roll and Slide using the algorithms adopted from the 3DNA source code.[31,67] We estimated the major and minor groove widths by measuring the distances between phosphorus atoms on the two antiparallel DNA strands. Our algorithm was different from the refined algorithm by 3DNA, but it gives close results.

4.3. Conformational Changes and Stability of Final Conformations.
For all the results from Tables 2, 3, and 4, the oligomers were initially in the B-form. We stopped the calculations after the transition to another conformation (if any) was completed, and the lifetime in the new conformation was long enough to reliably calculate the equilibrium average values and their dispersions for the parameters of the DNA helix. We believed that an equilibrium had been reached if there was no change in the average values of the parameters, and only thermal fluctuations were observed. After this, we collected the data from the rest of the trajectory (or from the

Table 5. Parameters of the Simulated Systems: The Number of Water Molecules and Pairs of Additional Salt Ions for Each Concentration, and the Total Simulation Time

Sequence	NaCl, M	Water	Ions	Simul. time, ns
(GC)$_{10}$ (polyGC)	0.15	19705	53	48−63
A$_{20}$ (polyA)	0.15	19705	53	53−55
	4.1	17493	1412	54−59
ATAT^a	0.13	1469	4	55−202
	4.5	1275	125	58−219
GGTA-TACC	0.13	2864	7	140−224
	4.2	2562	196	149−234
(AC)$_{10}$ (polyAC)	0	19362	0	52−97
	0.1	19308	34	51−152
	4.7	16895	1543	56−160
G$_{20}$ (polyG)	0.15	19705	53	34−70
	1^b	19508	356	35−42
	4.1	17493	1412	25−101

^{a}Temperature −2 °C. ^{b}Temperature 10 °C.

final parts of all the trajectories, if there were several replicas of the system), and calculated the equilibrium average values and their standard deviations.

One may think that the productive runs (see Table 5) are too short. Therefore, the molecules may have passed into a metastable form closer to the initial B-form, and after longer calculations we will see a different conformation (the global minimum). This possibility seems indeed not excluded for long oligomers, G$_{20}$ (poly(G)) and (AC)$_{10}$ (poly(AC)). Therefore, for doubtful cases, we tested the stability of the alternative DNA forms.

In the framework of the AMBER force field at the high salt concentration, we calculated the trajectories for the polymers poly(AC) starting from the A-form (B-DNA passed into C-DNA) and poly(G) starting from both the A- and the C-forms (B-DNA did not change the conformation). In all these cases, the molecule passed into the B-form (initial conformation for simulations from Tables 3 and 4) within 14−38 ns.

In the framework of the CHARMM force field, we obtained the trajectories for poly(AC) starting from the A-form (at the high salt concentration, B-DNA passed into C-DNA) and poly(G) starting from the C-form (at the low salt concentration, B-DNA passed into A-DNA). The oligomer poly(AC) took the B-form within 83 ns. Within 60 ns, the oligomer poly(G) passed into a form having 65% north sugar conformations and 10% BII phosphate conformations. This is an intermediate form we saw in the process of transition from the B- to the A-form. So, poly(G) will definitely end up in the A-form.

Thus, in all the doubtful cases, none of the possible alternative forms is the global minimum. Under the considered conditions, the stable conformations are those listed in Tables 2, 3, and 4, although the shapes of the molecules still may change after microseconds-long simulations (see Section 1).

■ AUTHOR INFORMATION

Corresponding Author

Elena A. Zubova − *N.N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, Moscow 119991, Russia;* ◉ orcid.org/0000-0001-8329-0645; Email: zubova@chph.ras.ru, elena.0.zubova@gmail.com

Authors

Ivan A. Strelnikov − *N.N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, Moscow 119991, Russia;* ● orcid.org/0000-0003-3430-7425

Natalya A. Kovaleva − *N.N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, Moscow 119991, Russia;* ● orcid.org/0000-0002-0621-9682

Artem P. Klinov − *N.N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, Moscow 119991, Russia;* ● orcid.org/0000-0002-8653-4972

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c07781

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Van Der Vaart, A. Coupled binding-bending-folding: The complex conformational dynamics of protein-DNA binding studied by atomistic molecular dynamics simulations. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1091−1098.

(2) Yoo, J.; Winogradoff, D.; Aksimentiev, A. Molecular dynamics simulations of DNA-DNA and DNA-protein interactions. *Curr. Opin. Struct. Biol.* **2020**, *64*, 88−96.

(3) Bouvier, B.; Zakrzewska, K.; Lavery, R. Protein-DNA recognition triggered by a DNA conformational switch. *Angew. Chem., Int. Ed.* **2011**, *50*, 6516−6518.

(4) Yesudhas, D.; Anwar, M. A.; Panneerselvam, S.; Kim, H. K.; Choi, S. Evaluation of Sox2 binding affinities for distinct DNA patterns using steered molecular dynamics simulation. *FEBS Open Bio* **2017**, *7*, 1750−1767.

(5) Singh, R. K.; Mukherjee, A. Molecular Mechanism of the Intercalation of the SOX-4 Protein into DNA Inducing Bends and Kinks. *J. Phys. Chem. B* **2021**, *125*, 3752−3762.

(6) Hognon, C.; Garaude, S.; Timmins, J.; Chipot, C.; Dehez, F.; Monari, A. Molecular Bases of DNA Packaging in Bacteria Revealed by All-Atom Molecular Dynamics Simulations: The Case of Histone-Like Proteins in Borrelia burgdorferi. *J. Phys. Chem. Lett.* **2019**, *10*, 7200−7207.

(7) Yoo, J. On the stability of protein−DNA complexes in molecular dynamics simulations using the CUFIX corrections. *J. Korean Phys. Soc.* **2021**, *78*, 461−466.

(8) Spiriti, J.; van der Vaart, A. DNA Binding and Bending by Sac7d is Stepwise. *ChemBioChem.* **2013**, *14*, 1434−1437.

(9) Zacharias, M. Atomic Resolution Insight into Sac7d Protein Binding to DNA and Associated Global Changes by Molecular Dynamics Simulations. *Angew. Chem., Int. Ed.* **2019**, *58*, 5967−5972.

(10) Singh, R. K.; Mukherjee, A. Molecular Mechanism of Dual Intercalation in Sac7d−DNA Complexation. *J. Phys. Chem. B* **2022**, *126*, 1682−1690.

(11) Liao, Q.; Lüking, M.; Krüger, D. M.; Deindl, S.; Elf, J.; Kasson, P. M.; Lynn Kamerlin, S. C. Long Time-Scale Atomistic Simulations of the Structure and Dynamics of Transcription Factor-DNA Recognition. *J. Phys. Chem. B* **2019**, *123*, 3576−3590.

(12) Mondal, M.; Mukherjee, S.; Bhattacharyya, D. Contribution of phenylalanine side chain intercalation to the TATA-box binding protein-DNA interaction: molecular dynamics and dispersion-corrected density functional theory studies. *J. Mol. Model.* **2014**, *20*, 2499.

(13) Singh, R. K.; Sasikala, W. D.; Mukherjee, A. Molecular Origin of DNA Kinking by Transcription Factors. *J. Phys. Chem. B* **2015**, *119*, 11590−11596.

(14) Shaytan, A. K.; Armeev, G. A.; Goncearenco, A.; Zhurkin, V. B.; Landsman, D.; Panchenko, A. R. Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Micro-second Timescale: Atomistic Insights into Nucleosome Functions. *J. Mol. Biol.* **2016**, *428*, 221−237.

(15) Elbahnsi, A.; Retureau, R.; Baaden, M.; Hartmann, B.; Oguey, C. Holding the Nucleosome Together: A Quantitative Description of the DNA-Histone Interface in Solution. *J. Chem. Theory Comput.* **2018**, *14*, 1045−1058.

(16) Sun, T.; Minhas, V.; Mirzoev, A.; Korolev, N.; Lyubartsev, A. P.; Nordenskiöld, L. A Bottom-Up Coarse-Grained Model for Nucleosome-Nucleosome Interactions with Explicit Ions. *J. Chem. Theory Comput.* **2022**, *18*, 3948−3960.

(17) Minhas, V.; Sun, T.; Mirzoev, A.; Korolev, N.; Lyubartsev, A. P.; Nordenskiold, L. Modeling DNA Flexibility: Comparison of Force Fields from Atomistic to Multiscale Levels. *J. Phys. Chem. B* **2020**, *124*, 38−49.

(18) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55−58.

(19) Hart, K.; Foloppe, 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.

(20) 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.

(21) Gallardo, A.; Bogart, B. M.; Dutagaci, B. Protein−Nucleic Acid Interactions for RNA Polymerase II Elongation Factors by Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2022**, *62*, 3079−3089.

(22) Tucker, M. R.; Piana, S.; Tan, D.; LeVine, M. V.; Shaw, D. E. Development of Force Field Parameters for the Simulation of Single- and Double-Stranded DNA Molecules and DNA−Protein Complexes. *J. Phys. Chem. B* **2022**, *126*, 4442−4457.

(23) 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.

(24) Janeček, M.; Kührová, P.; Mlýnský, V.; Otyepka, M.; Šponer, J.; Banáš, P. W-RESP: Well-Restrained Electrostatic Potential-Derived Charges. Revisiting the Charge Derivation Model. *J. Chem. Theory Comput.* **2021**, *17*, 3495−3509.

(25) 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.

(26) 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.

(27) Lu, X.-J.; Shakked, Z.; Olson, W. A-form Conformational Motifs in Ligand-bound DNA Structures. *J. Mol. Biol.* **2000**, *300*, 819−840.

(28) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233−269.

(29) van Dam, L.; Levitt, M. H. BII Nucleotides in the B and C Forms of Natural-sequence Polymeric DNA: A New Model for the C Form of DNA. *J. Mol. Biol.* **2000**, *304*, 541−561.

(30) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. *J. Mol. Biol.* **2002**, *319*, 1097−1113.

(31) http://web.x3dna.org/.

(32) Li, S.; Olson, W. K.; Lu, X.-J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **2019**, *47*, W26–W34.

(33) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(34) VMD (http://www.ks.uiuc.edu/Research/vmd/) was developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

(35) Ivanov, V. I.; Minchenkova, L. E.; Schyolkina, A. K.; Poletayev, A. I. Different conformations of double-stranded nucleic acid in solution as revealed by circular dichroism. *Biopolymers* **1973**, *12*, 89–110.

(36) Strelnikov, I. A.; Kovaleva, N. A.; Zubova, E. A. Variability of the DNA Backbone Geometry in DNA–Protein Complexes: Experimental Data Analysis. *J. Chem. Inf. Model.* **2021**, *61*, 4783–4794.

(37) Cheatham, T. E.; Crowley, M. F.; Fox, T.; Kollman, P. A. A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 9626–9630.

(38) Zgarbová, M.; Jurečka, P.; Šponer, J.; Otyepka, M. A- to B-DNA Transition in AMBER Force Fields and Its Coupling to Sugar Pucker. *J. Chem. Theory Comput.* **2018**, *14*, 319–328.

(39) Zhang, H.; Fu, H.; Shao, X.; Dehez, F.; Chipot, C.; Cai, W. Changes in Microenvironment Modulate the B- to A-DNA Transition. *J. Chem. Inf. Model.* **2019**, *59*, 2324–2330.

(40) Pastor, N. The B- to A-DNA transition and the reorganization of solvent at the DNA surface. *Biophys. J.* **2005**, *88*, 3262–3275.

(41) Benevides, J. M.; Overman, S. A.; Thomas, G. J., Jr Raman, polarized Raman and ultraviolet resonance Raman spectroscopy of nucleic acids and their complexes. *J. Raman Spectrosc.* **2005**, *36*, 279–299.

(42) Banyay, M.; Sarkar, M.; Gräslund, A. A library of IR bands of nucleic acids in solution. *Biophys. Chem.* **2003**, *104*, 477–488.

(43) Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37*, 1713–1725.

(44) Vorlíčková, M.; Kypr, J. Conformational Variability of Poly(dA-dT)· Poly(dA-dT) and Some Other Deoxyribonucleic Acids Includes a Novel Type of Double Helix. *J. Biomol. Struct. Dyn.* **1985**, *3*, 67–83.

(45) Nishimura, Y.; Torigoe, C.; Tsuboi, M. An A-form poly-(dG)*poly(dC) in H_2O solution. *Biopolymers* **1985**, *24*, 1841–1844.

(46) Katahira, M.; Nishimura, Y.; Tsuboi, M.; Sato, T.; Mitsui, Y.; Iitaka, Y. Local and overall conformations of DNA double helices with the AT base pairs. *Biochim. Biophys. Acta, Gene Struct. Expression* **1986**, *867*, 256–267.

(47) Katahira, M.; Nishimura, Y.; Tsuboi, M.; Matsuzaki, J.-i.; Hotoda, H.; Sekine, M.; Hata, T. A Raman Spectroscopic Study on Conformations of DNA Oligomers: A Dominant Effect of an AA:TT Sequence Over Those of AT:AT and TA:TA Sequences on Determining Conformations of DNA Duplexes. *Nucleosides, Nucleotides Nucleic Acids* **1994**, *13*, 1467–1481.

(48) Wang, Y.; Thomas, G. A.; Peticolas, W. L. Sequence Dependent Conformations of Oligomeric DNA's in Aqueous Solutions and in Crystals. *J. Biomol. Struct. Dyn.* **1987**, *5*, 249–274.

(49) Lee, C.; Cho, M. Vibrational dynamics of DNA: IV. Vibrational spectroscopic characteristics of A-, B-, and Z-form DNA's. *J. Chem. Phys.* **2007**, *126*, 145102.

(50) Siebert, T.; Guchhait, B.; Liu, Y.; Costard, R.; Elsaesser, T. Anharmonic Backbone Vibrations in Ultrafast Processes at the DNA-Water Interface. *J. Phys. Chem. B* **2015**, *119*, 9670–9677.

(51) Elsaesser, T.; Schauss, J.; Kundu, A.; Fingerhut, B. P. Phosphate Vibrations Probe Electric Fields in Hydrated Biomolecules: Spectroscopy, Dynamics, and Interactions. *J. Phys. Chem. B* **2021**, *125*, 3899–3908.

(52) Nishimura, Y.; Torigoe, C.; Tsuboi, M. Salt induced B–A transition of poly(dG)*poly(dC) and the stabilization of A form by its methylation. *Nucleic Acids Res.* **1986**, *14*, 2737–2748.

(53) Serec, K.; Šegedin, N.; Krajačić, M.; Dolanski Babić, S. Conformational Transitions of Double-Stranded DNA in Thin Films. *Appl. Sci.* **2021**, *11*, 2360.

(54) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

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

(56) Heddi, B.; Oguey, C.; Lavelle, C.; Foloppe, N.; Hartmann, B. Intrinsic flexibility of B-DNA: the experimental TRX scale. *Nucleic Acids Res.* **2010**, *38*, 1034–1047.

(57) Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science* **2004**, *305*, 386–389.

(58) Seckfort, D.; Lynch, G. C.; Pettitt, B. M. The lac repressor hinge helix in context: The effect of the DNA binding domain and symmetry. *Biochim. Biophys. Acta, Gen. Subj.* **2020**, *1864*, 129538.

(59) McDermott, M. L.; Vanselous, H.; Corcelli, S. A.; Petersen, P. B. DNA's Chiral Spine of Hydration. *ACS Cent. Sci.* **2017**, *3*, 708–714.

(60) http://lammps.sandia.gov/index.html.

(61) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

(62) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; CRC Press, 2021.

(63) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(64) https://casegroup.rutgers.edu/fd_helix.c.

(65) Zgarbová, M.; Otyepka, M.; Šponer, J.; Lankaš, F.; Jurečka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177–3189.

(66) Ben Imeddourene, A.; Elbahnsi, A.; Guéroult, M.; Oguey, C.; Foloppe, N.; Hartmann, B. Simulations Meet Experiment to Reveal New Insights into DNA Intrinsic Mechanics. *PLoS Comput. Biol.* **2015**, *11*, No. e1004631.

(67) Lu, X.; Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.