Stretched DNA Investigated Using Molecular-Dynamics and Quantum-Mechanical Calculations

Jan Řezáč,[†‡] Pavel Hobza,[†‡] and Sarah A. Harris[§¶*]

[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and [‡]Center for Biomolecules and Complex Molecular Systems, Prague, Czech Republic; and [§]School of Physics and Astronomy, and [¶]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, United Kingdom

ABSTRACT We combined atomistic molecular-dynamics simulations with quantum-mechanical calculations to investigate the sequence dependence of the stretching behavior of duplex DNA. Our combined quantum-mechanical/molecular-mechanical approach demonstrates that molecular-mechanical force fields are able to describe both the backbone and base-base interactions within the highly distorted nucleic acid structures produced by stretching the DNA from the 5′ ends, which include conformations containing disassociated basepairs, just as well as these force fields describe relaxed DNA conformations. The molecular-dynamics simulations indicate that the force-induced melting pathway is sequence-dependent and is influenced by the availability of noncanonical hydrogen-bond interactions that can assist the disassociation of the DNA basepairs. The biological implications of these results are discussed.

INTRODUCTION

The mechanical properties of duplex DNA have been extensively studied using single molecule nanomanipulation techniques such at atomic force microscopy and optical and magnetic tweezers (1–3). These methods have been used to stretch, twist, untwist, and unzip DNA as a function of helix length (4,5), temperature (6,7), solution conditions (8,9), DNA sequence (10–12), and pulling direction (13). They have also been used to probe protein-DNA interactions (14,15) and to investigate the mechanism of action of DNA processing motors such as polymerases (16,17) and topoisomerases (18,19). Single-molecule experiments that stretch long DNA strands of mixed sequence have shown that DNA undergoes a distinctive transition at 65 pN, where there is a plateau in the force-extension curve. This is known as the B-S or overstretching transition and is followed by a melting transition to two separate strands at ~150 pN (depending on the experimental conditions, such as the pulling rate) (20,21). Experimental studies comparing synthetic constructs of pure poly(dG-dC) with poly(dA-dT) have shown that these two sequences have very different mechanical properties (10–12). Both the melting transition and the overstretching transition have been found to occur at lower forces for poly(dA-dT) relative to poly(dG-dC). However, such single-molecule studies can provide no insight into the difference in structural changes that may occur when AT- and GC-rich DNA sequences are stretched in the laboratory.

Molecular modeling studies at the atomic level have proved invaluable for interpreting the results of DNA-stretching experiments (22–27). The most serious limitation of all molecular-dynamics (MD) calculations on biomacromolecules is the short timescales that are accessible, especially when the method is used to model processes that involve large conformational changes, such as stretching. MD simulations of DNA are generally limited to nanosecond timescales, whereas the stretching experiments take place on a timescale that is 3–6 orders of magnitude slower than the calculations. In a previous series of calculations to investigate the stretching behavior of a dodecamer (26), we reported simulation protocols specifically designed to overcome this gap in timescale that we also implement in the study presented here. A conventional steered MD simulation is used to elongate the DNA to over twice its original length. A number of increasingly stretched configurations sampled from this trajectory are then used as the starting point for additional calculations in which the DNA is held in an elongated configuration by external restraints and subjected to long-timescale MD. This discretization of the force-extension profile enables us to focus on a small number of the most interesting conformations and helps narrow the timescale gap between theory and experiment.

In this study we report a series of classical and quantum mechanical (QM) computer simulations in which the stretching behavior of the 30 basepair oligomer $d(GC)_{15}$ is compared with that of $d(AT)_{15}$. Classical MD calculations at the atomic level are still extremely computationally expensive despite the approximations inherent in molecular-mechanical (MM) force fields. Consequently, it is not possible to simulate the kilobase DNA sequences most commonly used in such experiments. Measurements performed by Strunz et al. (4) on a series DNA oligomers of increasing length showed that the overstretching transition is only observed if the duplex contains more than 30 basepairs. Therefore, the 30 basepair duplexes chosen for this study are the smallest sequences that still display mechanical properties similar to those of very long DNA polymers. In

Submitted December 23, 2008, and accepted for publication August 31, 2009.

*Correspondence: s.a.harris@leeds.ac.uk

Editor: Gregory A. Voth.

0006-3495/10/01/0101/10 $2.00

doi: 10.1016/j.bpj.2009.08.062

addition, these experiments identify the extension of the DNA at the transition state as the key rate-independent parameter that can be used to validate simulation results against the experimental data. For a 30 mer duplex of mixed sequence, these experiments indicate that the DNA should start to unbind at extensions of ~28 Å (4).

Atomistic MD is an established technique that can provide extremely good agreement with experimental data from x-ray crystallography and NMR as long as state-of-the-art force fields and simulation techniques are used (28). However, most of the testing and refinement of the force fields has necessarily taken place for canonical DNA structures, as there is very limited experimental information available regarding the structure of highly distorted or denatured DNA duplexes at the atomic level. Therefore, it remains to be established whether these force fields are also able to describe deformed molecular configurations (which may lie far outside of the regime in which the force field was designed to perform well) with sufficient accuracy. Therefore, we also used QM methods to confirm the conformational energies obtained using the empirical MM force field. We used the recently developed density function theory with dispersion (DFT-D) methodology, which combines high-level DFT calculations with an empirical treatment of dispersion, to investigate many molecular structures at reduced computational expense relative to the traditional correlated wavefunction theory methods, which include dispersive effects. The DFT-D method was selected as the benchmark QM method because it yields results with a chemical precision ($<$1 kcal/mol) that is comparable to highly accurate CCSD/(T)/complete basis set limit calculations (29), and it is capable of describing the hydrogen bonding and dispersion interactions that are known to make an important contribution to the stability of duplex DNA (30,31). The computational efficiency of the method therefore makes it an invaluable tool for studying distorted biomolecular conformations. Our QM calculations specifically focus on the ability of MM force fields to reproduce the interaction energy between DNA base-pairs within the distorted duplexes, and consequently use base geometries sampled from the MM simulations without a separate QM optimization of the individual bases (which might well add a separate correction to the MM geometries and consequently the single base energies).

The primary aim of this study was to validate the use of MM force fields for the description of distorted DNA structures. However, since the classical calculations are of interest in their own right, rather than simply stretching the DNA using external restraints to generate elongated conformations for a QM analysis, we followed our previous successful strategy for investigating DNA extension by an external force (26). First, the DNA is subjected to a rapid elongation using steered MD to provide the molecular conformations needed for a more detailed analysis. These preparatory simulations also provide an initial estimate of the position of the transition state from the location of the force plateau in the calculated force-extension curves. The transition state indicates the extension at which the DNA becomes thermodynamically unstable. This quantity is of prime importance because it does not depend on the pulling rate (32), and therefore enables the simulation results to be quantitatively compared with the available experimental data. To characterize the stretching pathway in more detail, and to locate the position of the transition state unambiguously, representative conformations at a variety of extensions on either side of force plateau are selected for long timescale (10 ns) restrained or ''pinned'' MD, as described in detail in the Materials and Methods section. These pinned simulations provide a quasistatic picture of the behavior of the DNA as a function of its extension at discrete points along the stretching pathway, and therefore do not suffer from the artifacts associated with the rapid elongation rates used during the steered MD. All subsequent data analysis, including QM calculations, an investigation of the sequence dependence of the stretching behavior of DNA, and validation against the available experimental data are based on these pinned MD calculations; the data from the steered MD were discarded.

METHODS

Classical MD simulations

Simulation setup

MD simulations were performed with the AMBER8 suite of programs (33) and the ff99 (34) and BSC0 force fields (35), using the default ion parameters. The two sequences—$d(AT)_{15}$ and $d(GC)_{15}$—were built using the NUCGEN module available within AMBER. Both systems were electrically neutralized by the addition of 58 potassium counterions. The steered MD simulation stretched the DNA to over twice its original length by pulling it along the helix axis. Therefore, an elongated rectangular periodic box containing 31,310 water molecules for $d(AT)_{15}$ and 33,944 for $d(GC)_{15}$ was required to ensure that the extended structures were surrounded by at least 15 Å of solvent in each orthogonal direction. Although the ff99 version of the AMBER force field has been shown to lead to unstable DNA structures for simulations in excess of 50 ns due to a poor description of backbone torsion angles (35,36), we nevertheless used this force field in our calculations to ensure consistency with previous simulation studies. Since these calculations on stretched DNA structures indicate that the dominant contribution to the destabilization of the duplex occurs through disruption of the base-stacking interactions rather than conformational transitions within the backbone, we are confident that alterations to the backbone torsion parameters within newer versions of the force field will make a negligible contribution to the behavior observed. Nevertheless, we performed a number of additional simulations of stretched DNA using the newer BSC0 force field for comparison.

Preparatory steered MD simulations

For the MD simulations, the fast particle mesh Ewald method was used to calculate long-range electrostatic interactions. Constraint of all covalent bonds to hydrogen with SHAKE allowed an integration timestep of 2 fs to be used in conjunction with the Verlet integration algorithm. MD was performed at a constant temperature (298 K) and pressure (1 atm). The system was thermalized and equilibrated for 1 ns using a standard multistage protocol (37). During the steered MD runs, $d(AT)_{15}$ and $d(GC)_{15}$ were forced to extend from a lengths of ~90 Å and 97 Å to 259 Å and 254 Å, respectively, by the application of strict restraints to the 5′-terminus of each strand of the molecule. These restraints consisted of a parabolic well

with an associated force constant of 100 kcal/Mol/Å^2. The DNA molecule was stretched by increasing the distance between the two restrained atoms by 0.5 Å, followed by an intervening 50 ps relaxation period. The new positions of these atoms were maintained by the high energetic penalties associated with a deviation from this set value. Steered MD was run for a total of 16.9 ns and 15.7 ns for d(AT)$_{15}$ and d(GC)$_{15}$ respectively. The internal energy change on stretching was calculated using the MD implementation of the generalized Born/surface area implicit solvation method using the Tsui and Case parameters (38), and based on trajectories obtained using explicit solvent, as previously described (39). The force-extension curves were obtained by differentiating the seventh-order polynominal fit to the change in energy with respect to the extension. Thus, we were able to obtain an initial estimate for the location of the transition state from the position of the plateau in the force-extension curves.

"Pinned" simulations

Configurations of the DNA sampled from various points along the steered dynamics trajectory were then used as the starting points for five additional MD simulations of d(AT)$_{15}$ (at extensions of 16, 18, 30, 33, and 39 Å) and seven simulations of d(GC)$_{15}$ (at extensions of 16, 23, 29, 32, 38, 47, and 58 Å). These simulations were chosen to span the interesting part of the force-extension pathway on either side of the force plateau. The number of simulations performed was a compromise between maximizing the exploration of the free-energy landscape and the computational expense of the calculations. An additional two simulations were performed for GC because it has a longer force-extension profile. All simulations were solvated and electrically neutralized with potassium counterions. Harmonic restraints with a force constant of 100 kcal/Mol/Å^2 applied to the 5′ termini were used to maintain the molecule in an elongated configuration. These restrained structures were then subjected to standard 1 ns solvent and counterion equilibration protocols before the 10 ns data production phase. A subset of these simulations (at extensions of 18, 33, and 39 Å for d(AT)$_{15}$, and 23, 32, 38, and 58 Å for d(GC)$_{15}$) were then repeated, but only after an additional 1 ns equilibration run was performed before the data production phase in which the hydrogen bonds were restrained to prevent denaturation. This extended equilibration protocol was designed primarily to ensure that the repeat simulations would enter a entirely different region of conformational space, but also in an attempt to remove any artifacts introduced by the rapidity of the steered MD from which the starting conformations were obtained. A second repeat of these simulations was performed using the new parmBSC0 force field for comparison. In addition, a conventional unrestrained 10 ns MD run was performed for both sequences to be used as the thermodynamic reference state for comparison with distorted DNA structures.

QM calculations

A representative set of simulations from the series of MD calculations were selected for a detailed QM analysis; simulations were selected so as to span as much of the forced-unbinding profile (from slighted elongated to highly extended) as was computationally feasible. Duplexes at extensions of 16, 30, and 39 Å were chosen for d(AT)$_{15}$, and extensions of 16, 23, 29, and 58 Å were selected for d(GC)$_{15}$. Molecular conformations were sampled every 1 ns from each trajectory. The DNA bases were extracted from the duplex, and the glycosidic bond was terminated with a hydrogen atom. The interaction energy was then calculated for each interacting pair of bases using both AMBER (using the restrained electrostatic potential-derived charges as recommended) and DFT-D. The interacting bases were selected from each molecular conformation using a cutoff distance of 5.5 Å between atoms. This criterion counted all nearest-neighbor bases as interacting, and also included contacts with any denatured DNA bases present due to the distortion of the helix by the applied force. To calculate interaction energies using MM, a set of partial charges for isolated bases is required (the force field already contains the necessary atom types and bonding parameters). Charges were generated using the restrained electrostatic potential procedure based on QM calculations at the HF/6-31G* level (40). The DFT-D calculations used a combination of the TPSS functional (41) and a triple ζ-basis set (TZVP) (42) with the corresponding set of dispersion parameters (29). The resolution of identity approximation within DFT (43), as implemented within the Turbomole software package (44), was used to further improve computational efficiency.

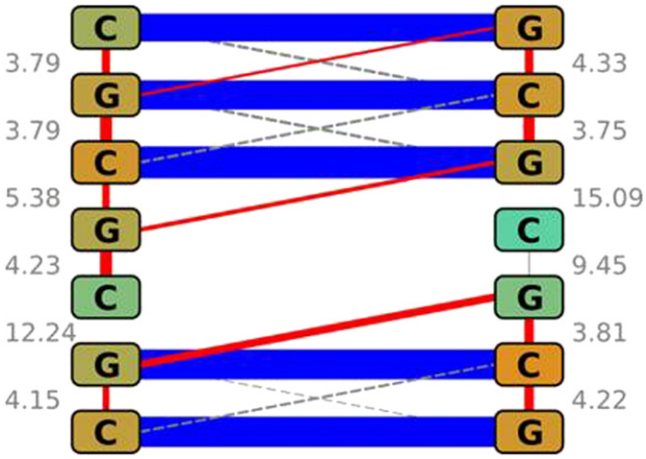

FIGURE 1 Example of the schematic representation of the DNA used to visualize QM interaction energies between bases. The line thicknesses are proportional to the interaction strength, with blue indicating hydrogen bonding, red indicating other attractive interactions (mainly dispersion), and gray showing repulsive interactions. Key interbase distances in Å are also provided (numerical values are given in *gray*). The solvent-accessible surface areas are depicted by the rectangle color, with orange indicating that the base is embedded within the duplex; the rectangle is green if the base is exposed to solvent. The representations available online are interactive and provide additional numerical values for all of the variables presented graphically in this work. Color available online only.

Visualization: MM and QM interaction energies and solvation

We also developed a visualization tool to aid the analysis of this extensive set of MM and QM calculations in a scalable vector graphics format based on a schematic representation of the DNA. The sequence of the two strands is shown alongside all calculated pairwise interactions between the bases, the important interbase distances, and the solvent-accessible surface area. An example is shown in Fig. 1. The solvation of bases plays an important role in DNA denaturation because energetically favorable interactions with water can partially compensate for broken basepairing (45). Although an accurate calculation of the solvation free energies is not computationally feasible for such a large number of molecular conformations, we obtained qualitative information for every base in all of the molecular structures selected for QM analysis by evaluating the difference in the solvent-accessible surface area relative to canonical B-form DNA using the NACCESS program (46). This interactive, schematic visualization tool provides a useful mechanism to summarize the qualitative results of a simulation ''at a glance''; it is provided as Supporting Material.

RESULTS

Validating the AMBER force field for stretched DNA structures using DFT-D

Contribution from interbase interactions

The interaction energies between basepairs calculated using the AMBER force field for both d(GC)$_{15}$ and d(AT)$_{15}$ are

plotted against those calculated using DFT-D in Fig. 2. These energies were partitioned into the contribution from hydrogen bonding (*blue*) and intra- and interstrand stacking (shown in *green* and *red*, respectively). There is an excellent agreement (the average errors are 0.2 kcal/mol and 0.25 kcal/mol) between the MM and QM methods for intra- and interstrand stacking interactions (a linear fit to both sets of data gives a regression line that passes through the origin and has a gradient of 1.04). However, the hydrogen-bond interactions are systematically underestimated by AMBER, giving on average a discrepancy of −1.9 kcal/mol (17%) for an AT basepair and −3.3 kcal/mol (12%) for a GC basepair. We attribute these differences to effects that cannot be included by MM approaches—in particular to polarization, but also to the partial covalent character of the hydrogen bonds between complementary basepairs (31,47). This hypothesis is supported by the observation that the effect is more pronounced for higher basepair interaction energies (a linear fit to the data gives a regression line with a gradient of 0.87). However, given the approximations inherent in MM, these errors are surprisingly small, and explain the successful use of AMBER MM calculations for describing nucleic acids in general. Most significantly, the discrepancy between the MM and QM results is not more pronounced for the noncanonical DNA structures. Therefore, these results, which were obtained using a statistically significant set of structures, indicate that it is possible to use the AMBER force field for a description of stretched DNA with the same confidence as for canonical DNA structures.

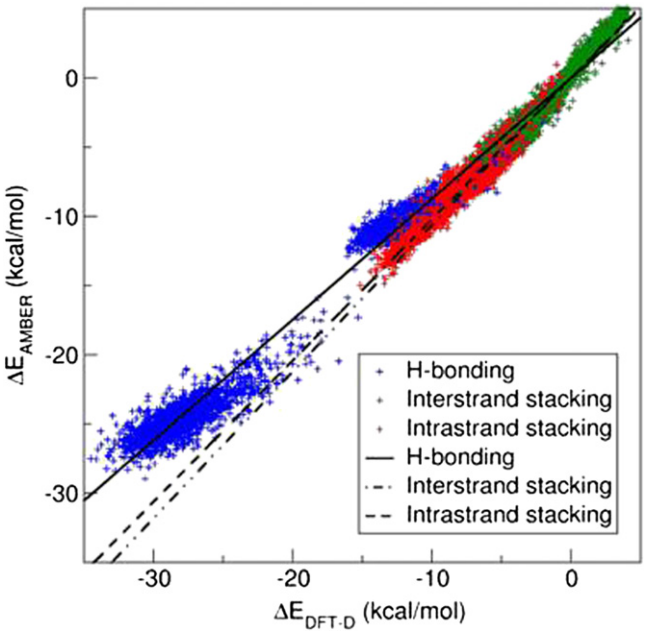

FIGURE 2 Comparison of AMBER energies with those calculated using DFT-D. Color available online only.

Contribution from the DNA backbone

To determine the contribution from the distortion of the sugar-phosphate backbone on the energetics of stretched DNA, we calculated the potential energy of the backbone under varying extensions. This was complicated by the fact that the extension of the duplex was not homogeneous during the stretching process, but rather occurred in well-defined regions while the rest of the double helix remained in the canonical B-form, in agreement with a recent simulation study of DNA denaturation by over- and undertwisting (48). To observe these local perturbations, we separated the backbone potential energy into contributions from each dinucleotide step. We calculated the energies of each backbone fragment relative to canonical B-DNA using both ff99 and the newer BSC0 force fields, which has been shown to improve the description of DNA backbone dynamics. Surprisingly, the backbone energies remained very close to the value for B-form DNA, even for the most elongated DNA structures. The maximum difference in energy detected was only 3.2 kcal/mol with ff99. On average, the BSC0 force field gave backbone stretching energies per base step of ~1 kcal/mol higher than ff99.

To further validate this analysis and the performance of the two force fields, we performed DFT-D calculations on a small model extracted from the simulated system. Two trimer duplexes were extracted from a snapshot from the simulation of $d(GC)_{15}$ at an extension of 30 Å: one in the B-DNA region, and one in a well-defined melted region. In these trimers, the bases were removed and the severed bonds capped with hydrogen atoms. The positions of all hydrogens in the structure were then optimized. The energies of the two strands were calculated separately using both DFT-D (TPSS/TZVP) and the MM force fields. We obtained the energy of stretching of the backbone by subtracting the energy of the B-form from the stretched structure. Finally, the stretching energy was averaged from both strands and expressed as an energy difference for one dinucleotide. The energy difference amounts to 3.3 kcal/mol in the DFT-D calculation. In the ff99 force field, the backbone is slightly softer and the energy difference is 2.2 kcal/mol. The BSC0 force field yields a larger energy difference of 5.5 kcal/mol. However, the energetic contribution from backbone distortions was still far smaller than the energetic cost from disruption of base stacking and hydrogen-bonding interactions. A geometrical analysis of the twist angle of the DNA performed using the CURVES program (49) showed that large separations between stacked bases occurred by local untwisting of the double helix, rather than through stretching of the sugar-phosphate backbone. Therefore, the energetic cost that imposes this backbone distortion is negligible compared to the disruption of interbase interactions. We assume therefore that the interaction between bases and the backbone also makes a negligible contribution to any discrepancy between QM and MM methods for describing highly distorted DNA structures.

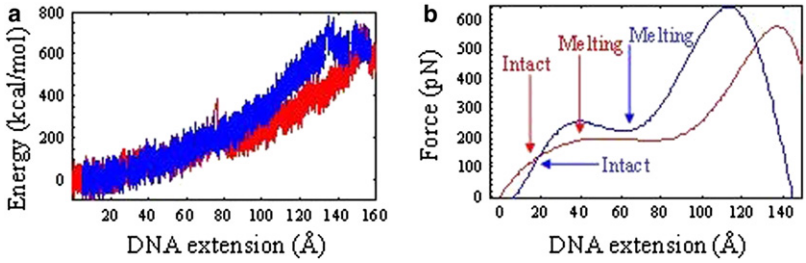

FIGURE 3 Generalized Born/surface area energy as a function of extension (*left*) measured from steered MD simulations of $d(AT)_{15}$ (*red*) and $d(GC)_{15}$ (*blue*). The corresponding force-extension curves (*right*) were calculated by differentiating the seventh-order polynomial fit to Fig. 3 *a* for both $d(AT)_{15}$ (*red*) and $d(GC)_{15}$ (*blue*). Arrows indicate extensions where DNA conformations were observed to be either intact or melted during the ''pinned'' MD calculations. Color available online only.

Sequence-dependent behavior in the preparatory steered MD calculations

To investigate the sequence dependence of the melting process in detail, it is necessary to locate the minimum extension at which melting can be detected using MD for both $d(AT)_{15}$ and $d(GC)_{15}$. In our previous study, conformations that extended into the region of the force-plateau were unstable and formed one or more denatured regions or ''denaturation bubbles'' in which hydrogen-bonding interactions between complementary basepairs broke as the simulation proceeded. The purpose of the steered MD simulations of $d(AT)_{15}$ and $d(GC)_{15}$ was first to locate regions of interest in the stretching profile, such as the position of the force plateau, and then to provide starting structures for a more detailed analysis of the sequence-dependent behavior of stretched DNA. Fig. 3 *a* shows the energy as a function of extension (*left*) and the force-extension curves (Fig. 3 *b*, *right*) obtained from the steered MD simulations of $d(AT)_{15}$ (*red*) and $d(GC)_{15}$ (*blue*). The position of the force plateau is clearly identifiable in Fig. 3 *b* for both sequences. Taken literally, these calculations imply that both sequences unbind at far higher forces than are measured experimentally (because the point of mechanical instability is where there is a maximum in the force-extension curves), as is to be expected from such rapid elongation rates. The agreement for GC-containing sequences (which unbind at ~250–350 pN experimentally, compared with the 650 pN measured by the simulation) is somewhat better than for the AT-containing duplex (which unbinds at 35 pN in the laboratory, compared with the 550 pN calculated by MD). The force values obtained are elevated relative to the experimental values by the rapid elongation rates used in the calculations. DNA stretching in the simulations takes place so rapidly that friction dominates, obscuring the differences between the AT- and GC-rich sequences. A traditional steered MD approach can only provide limited insight into the sequence dependence of the stretching properties of DNA at the pulling rates accessible using current computational resources. Consequently, these steered calculations can only be used to indicate the location of the transition state, and to provide elongated DNA conformations for a more detailed analysis, as described below. Albrecht et al. (13) recently demonstrated analogous kinetic effects experimentally by monitoring the directional dependence of the DNA response to an applied force, and invoked the formation of denaturation bubbles, as we indeed observe in our ''pinned'' simulations (as described below), to explain these new data.

Sequence-dependent behavior revealed by the "pinned" simulations of stretched DNA

Our previous calculations on a DNA dodecamer showed that the force plateau signifies the extension where the DNA becomes unstable and starts to separate into two separate strands. Therefore, we extracted five and seven configurations of $d(AT)_{15}$ and $d(GC)_{15}$, respectively, from the region close the center of the plateau for further analysis, as previously described. These simulations allowed the exploration of far longer timescales and consequently a more detailed investigation of the melting mechanism. We then repeated a subset of these simulations using an alternative equilibration protocol, and performed a second repeat using the newer BSC0 force field. The results of these repeat simulations are provided in the Supporting Material.

In agreement with our previous study, both $d(AT)_{15}$ and $d(GC)_{15}$ started to melt at extensions corresponding to the force plateau, as long as sufficient time was allowed for denatured regions of the duplex to start to appear. Fig. 4 shows hydrogen-bonding interactions as a function of time for each extension of $d(AT)_{15}$ (*left*) and $d(GC)_{15}$ (*right*). The final structure from each 10 ns simulation is shown alongside. Since fraying of single basepairs at the ends of the duplex is observed regularly in simulations of unstretched DNA, denaturation is considered to occur either when more than one basepair interaction is broken, or when this occurs in the center of the duplex. Fig. 4 and Fig. S1 of the Supporting Material show that two complementary basepair interactions at the end of the duplex consistently break in $d(AT)_{15}$ at an extension of 33 Å (note that a repeat simulation was not performed at an extension of 30 Å, so this corresponds to an upper limit for the position of the transition state of AT containing 30 mers). Broken basepairs do not consistently occur in all $d(GC)_{15}$ simulations until the far larger extension of 58 Å, indicating that the AT-containing sequence is significantly softer than that containing GC basepairs, as would be expected. Even though the data in Fig. 4 suggest that the basepairing is intact at 38 Å, there is nevertheless significant distortion to the base stacking, as discussed below and shown in detail in Fig. 5. A thermodynamic analysis of the stabilities of the duplexes (provided in the Supporting Material) provides an estimate of the position of the transition

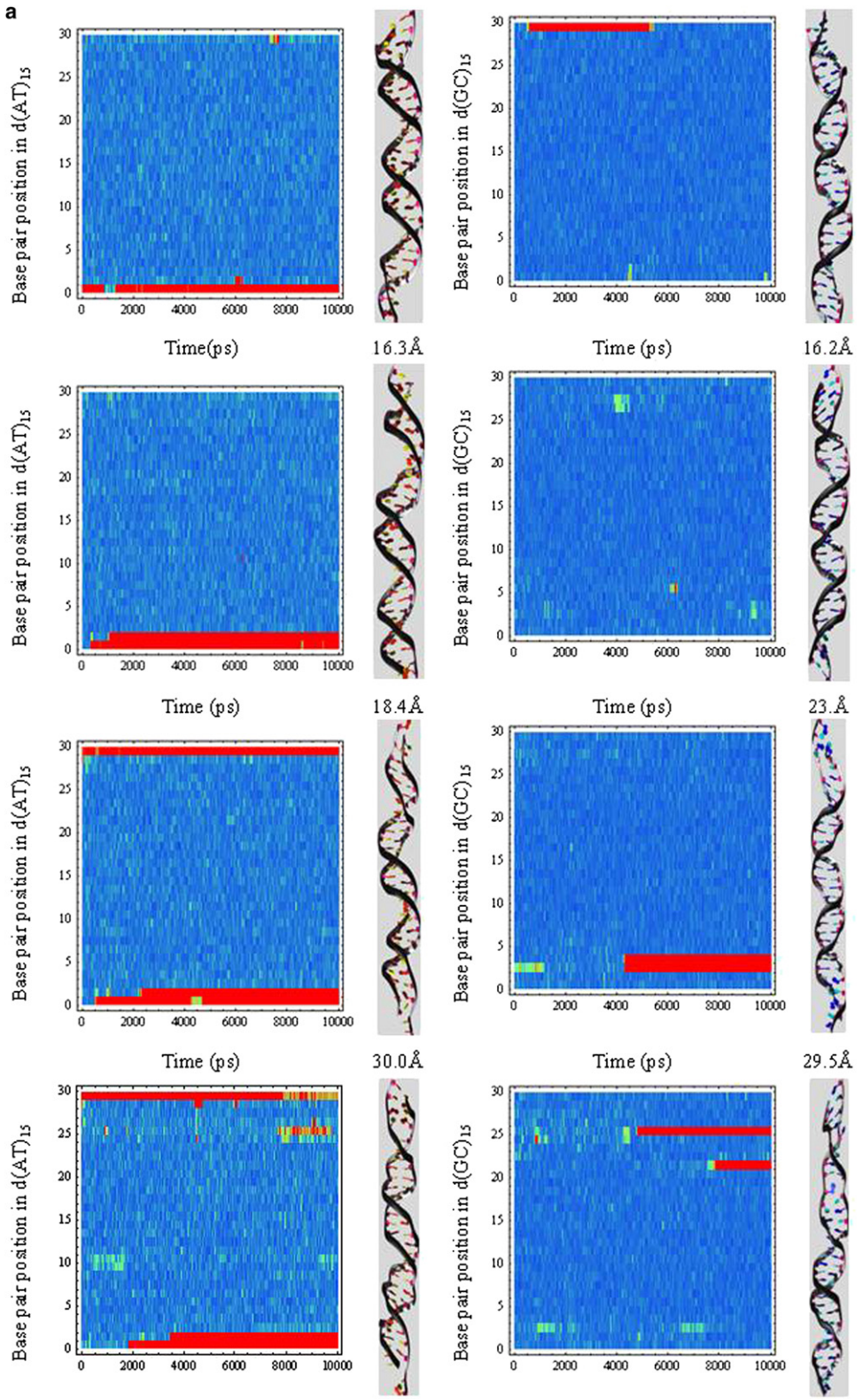

FIGURE 4 Hydrogen-bond distances and molecular conformations obtained from the "pinned" simulations of stretched DNA structures at various extensions close to the force plateau for $d(AT)_{15}$ (*left*) and $d(GC)_{15}$ (*right*). Color available online only.

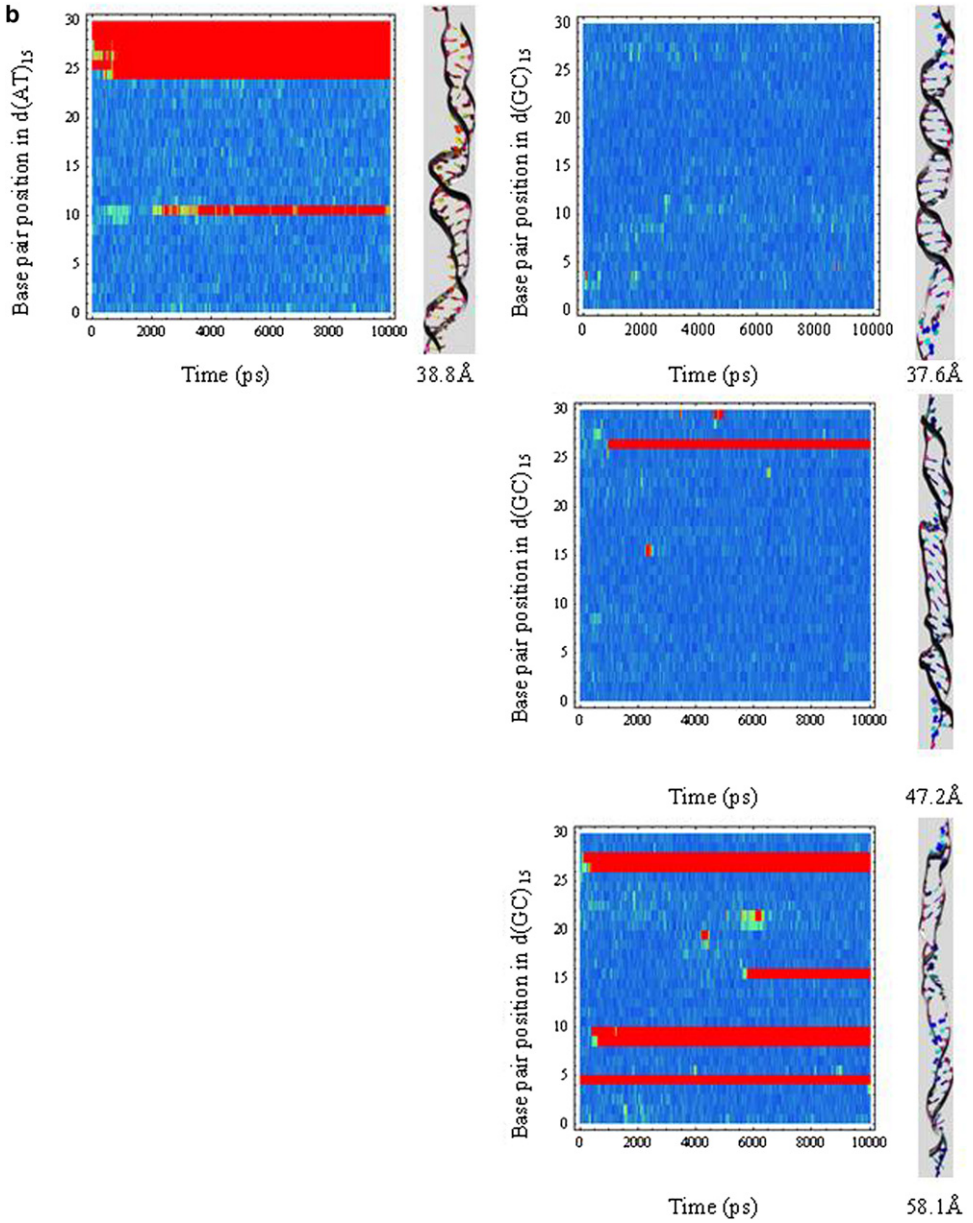

FIGURE 4 (*continued*).

state of 18–33 Å for $d(AT)_{15}$ and 32–38 Å for $d(GC)_{15}$, in good agreement with the 28 Å estimated from the experimental data for mixed sequences.

The simulations show that denaturation of $d(AT)_{15}$ is more often initiated from the ends of the duplex, whereas the melting of $d(GC)_{15}$ can start from the center of the DNA, in agreement with previous calculations of sequence-dependent mechanical properties by Piana (27). This is despite the fact that the melting process is stochastic, as illustrated by the repeat simulations presented in the Supporting Material. Of particular interest is the simulation of $d(GC)_{15}$ restrained at an extension of 38 Å. Although no complementary hydrogen-bonding interactions are broken during this simulation, the DNA elongates by breaking stacking interactions between consecutive basepairs, as shown in detail in Fig. 5. Similar behavior was also observed for $d(GC)_{15}$ at an extension of 32 Å using the BSC0 force field, but it was never observed for $d(AT)_{15}$. This suggests that the stacking interactions in GC-rich DNA are weaker than hydrogen bonding, even when the positive change in solvation free energy on liberating a basepair form within the duplex is taken into account, as was previously reported (45). This is not the case, however, for AT-containing duplexes, which are held together by weaker hydrogen bonds. We hypothesize that the contrasting melting preferences of these two sequences is due to this difference in the importance of hydrogen bonding relative to stacking interactions. AT-containing sequences have a higher propensity to melt from the

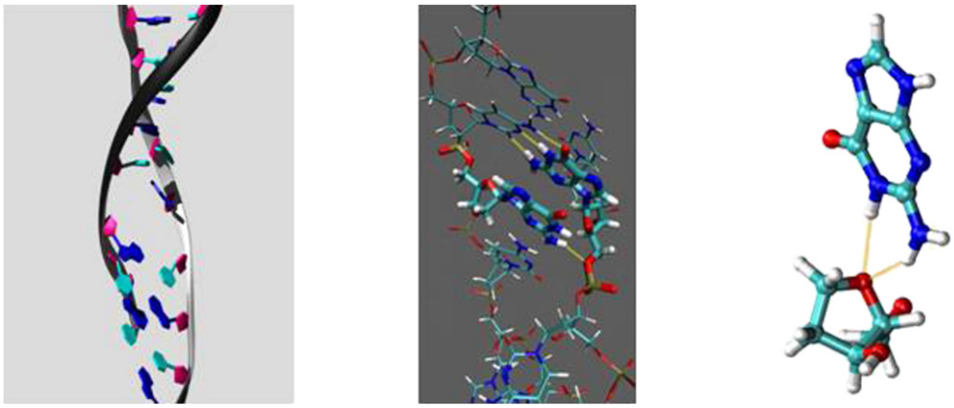

FIGURE 5 Intermediate conformations observed in the force-induced melting of $d(GC)_{15}$. At an extension of 38 Å, the base stacking interactions are disrupted, whereas hydrogen-bonding interactions between complementary basepairs remain intact (*left*). At an extension of 47 Å, a transient hydrogen bond is observed between the amine group on the guanine base and the 5′ oxygen of the sugar phosphate backbone (*middle*). In addition, transient hydrogen bonds can be formed between the sugar ring oxygen and the 1-NH and C-2 exocyclic amino group of guanine. Color available online only.

ends of the duplex because the stacking interactions play a more significant role relative to hydrogen bonding in maintaining the stability of the DNA, and therefore these sequences are more sensitive to end effects than the corresponding GC-containing duplexes.

Insight into the melting mechanism provided by "pinned" simulations of stretched DNA

We observe distinct differences in the melting pathway of GC and AT in the simulations. The AT-containing sequences melt predominantly through the ''peeling'' of successive basepairs from the ends of the structure, and the denatured regions grow from the ends of the helix as the extension of the biomolecule is increased, as shown in Fig. 4 and Fig. S1. The GC-containing sequences, however, can initiate melting in the center of the duplex. This result may somewhat surprising, since the energy barrier for melting a basepair is expected to be higher in the center of the helix than at the ends of the DNA. However, we observed a number of interesting intermediate conformations just before basepair separation in $d(GC)_{15}$, in which transient hydrogen-bond interactions between the bases (mostly guanines) and the sugar-phosphate backbone assisted the melting process by lowering the energy barrier between basepaired and unpaired states. An example is shown in Fig. 5. A full QM treatment shows that the hydrogen bond between the amine nitrogen on the guanine base and the 5′ oxygen in the backbone is strong, having an energy of 12 kcal/mol (where the average hydrogen-bond energy is ~7.5 kcal/mol in DNA). The abundance of potential noncanonical interactions between the melted bases themselves and the DNA backbone can therefore assist the melting process. The $d(GC)_{15}$ sequence explores a complex series of reaction pathways as the basepairs separate, and the process is highly stochastic. Fig. 4 shows that there is no clear preference for a given basepair to separate. By contrast, $d(AT)_{15}$ consistently follows the same path to the denatured state by nucleating unbound bases from the two ends of the helix.

CONCLUSIONS

In this study we performed a series of simulations to compare the stretching behavior of $d(GC)_{15}$ with $d(AT)_{15}$ at the MM and QM levels. The transition state was calculated to lie between 18 and 33 Å for $d(AT)_{15}$, and between 32 and 38 Å for $d(GC)_{15}$, in good agreement with the 28 Å estimated from the experimental data for mixed sequences (4). Because the position of the transition state does not depend on the pulling rate, it is the principal measurable quantity that enables the simulations to be compared with the experiments (32). To date, no stretching experiments have been performed to compare the mechanical properties of short AT- and GC-containing sequences. However, it seems reasonable to assume that the mechanical strength of mixed DNA oligomers should be intermediate between these two extremes. The simulations show that AT-rich DNA is significantly softer than the GC-containing duplex, in agreement with previous simulation studies (27) and with experiment (10–12). By performing a QM analysis on structures sampled from the MD trajectories, we were able to demonstrate that the AMBER force field provides a good description of denatured DNA structures, as well as canonical DNA. Thanks to the continued improvement of computational resources and methodologies, and the ability of simulators to access increasingly ambitious timescales and larger systems, atomistic calculations of DNA denaturation through processes such as twisting (48,50,51), heating (53), bending (54,55), and even translocation through synthetic pores (56) have been reported. Consequently, the performance of MD force fields for noncanonical DNA structures is of increasing interest to the simulation community.

The simulations provide what we believe is new insight into the sequence-dependent melting mechanism for DNA denaturation by stretching. The energetics of the stretching process is dominated by base stacking and hydrogen-bonding interactions, and changes in energy due to conformational changes within the backbone play only a minor role. The AT-containing sequences almost always melt from the ends of the helix, whereas the GC sequences are able to initiate melting in the central region of the duplex. Basepair separation is always local to the initiation site, with intact B-form regions of the duplex separated by locally melted regions of DNA. The melting of $d(GC)_{15}$ is more highly stochastic relative to $d(AT)_{15}$, and is assisted by non-canonical interactions with other melted bases and with the sugar-phosphate backbone.

The sequence-dependent nature of the melting pathway in these simple alternating DNA sequences implies that similarly complex behavior will also occur for double-stranded nucleic acid structures with mixed sequence. The availability of favorable noncanonical interactions along the forced disassociation pathway, as observed in the simulations, lowers the free-energy barrier between bound and unbound states, and consequently increases the unbinding rate. Moreover, this complexity of the free-energy landscape spanning bound and unbound states will also play an important role in determining the thermal denaturation pathway of duplexes, and the folding and unfolding kinetics of DNA and RNA loops. A recent single-molecule fluorescence resonance energy transfer study comparing the unfolding and refolding of a simple RNA-stem loop of the RNA bacteriophage MS2 showed that a single U-to-C replacement within the loop results in dramatic differences in the pathway between folded and unfolded states, even though the thermodynamic stabilities of the loops are similar (57). This behavior has important biological implications for the relative effectiveness of the native and mutated sequences as a molecular switch for controlling virus coat protein assembly. Many biological molecules undergo large structural changes in the course of their function that require the breaking and/or formation of hydrogen bonds. In these instances, the details of the pathway between these two states can be as important biologically as the free-energy difference between them.

SUPPORTING MATERIAL

Results and analysis, two figures, and references are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01562-8.

The simulations reported in this study were performed with the help of the UK National Grid Service. We thank Charlie Laughton for use of the code to calculate the configurational entropies and assess the α/γ-backbone torsion angles.

J.R. and P.H. were supported by research project 74 055 0506 and grant LC512 from the Ministry of Education of the Czech Republic. P.H. also received support from the Praemium Academiae of the Academy of Sciences of the Czech Republic.

REFERENCES

1. Lavery, R., A. Leburn, …, V. Croquette. 2002. Structure and mechanics of single biomolecules: experiment and simulation. *J. Phys. Condens. Matter*. 14:R383–R414.
2. Strick, T. R., M.-N. Dessinges, …, V. Croquette. 2003. Stretching of macromolecules and proteins. *Rep. Prog. Phys.* 66:1–45.
3. Williams, M. C., and I. Rouzina. 2002. Force spectroscopy of single DNA and RNA molecules. *Curr. Opin. Struct. Biol.* 12:330–336.
4. Strunz, T., K. Oroszlan, …, H. J. Güntherodt. 1999. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. USA*. 96:11277–11282.
5. Pope, L. H., M. C. Davies, …, P. M. Williams. 2001. Force-induced melting of a short DNA double helix. *Eur. Biophys. J.* 30:53–62.
6. Schumakovitch, I., W. Grange, …, M. Hegner. 2002. Temperature dependence of unbinding forces between complementary DNA strands. *Biophys. J.* 82:517–521.
7. Williams, M. C., J. R. Wenner, …, V. A. Bloomfield. 2001. Entropy and heat capacity of DNA melting from temperature dependence of single molecule stretching. *Biophys. J.* 80:1932–1939.
8. Rouzina, I., and V. A. Bloomfield. 2001. Force-induced melting of the DNA double helix 1. Thermodynamic analysis. *Biophys. J.* 80: 882–893.
9. Rouzina, I., and V. A. Bloomfield. 2001. Force-induced melting of the DNA double helix. 2. Effect of solution conditions. *Biophys. J.* 80: 894–900.
10. Rief, M., H. Clausen-Schaumann, and H. E. Gaub. 1999. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* 6:346–349.
11. Clausen-Schaumann, H., M. Rief, and H. E. Gaub. 1999. Sequence dependent mechanics of single DNA molecules. *Biophys. J.* 76:A151.
12. Clausen-Schaumann, H., M. Rief, …, H. E. Gaub. 2000. Mechanical stability of single DNA molecules. *Biophys. J.* 78:1997–2007.
13. Albrecht, C. H., G. Neuert, …, H. E. Gaub. 2008. Molecular force balance measurements reveal that double-stranded DNA unbinds under force in rate-dependent pathways. *Biophys. J.* 94:4766–4774.
14. Hegner, M., S. B. Smith, and C. Bustamante. 1999. Polymerization and mechanical properties of single RecA-DNA filaments. *Proc. Natl. Acad. Sci. USA*. 96:10109–10114.
15. Williams, M. C., I. Rouzina, …, V. A. Bloomfield. 2001. Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc. Natl. Acad. Sci. USA*. 98:6121–6126.
16. Wang, M. D., M. J. Schnitzer, …, S. M. Block. 1998. Force and velocity measured for single molecules of RNA polymerase. *Science*. 282: 902–907.
17. Maier, B., D. Bensimon, and V. Croquette. 2000. Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proc. Natl. Acad. Sci. USA*. 97:12002–12007.
18. Strick, T. R., V. Croquette, and D. Bensimon. 2000. Single-molecule analysis of DNA uncoiling by a type II topoisomerase. *Nature*. 404:901–904.
19. Koster, D. A., V. Croquette, …, N. H. Dekker. 2005. Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature*. 434:671–674.
20. Smith, S. B., Y. J. Cui, and C. Bustamante. 1996. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*. 271:795–799.
21. Cluzel, P., A. Lebrun, …, F. Caron. 1996. DNA: an extensible molecule. *Science*. 271:792–794.
22. Lebrun, A., and R. Lavery. 1996. Modelling extreme stretching of DNA. *Nucleic Acids Res*. 24:2260–2267.
23. Konrad, M. W., and J. I. Bolonick. 1996. Molecular dynamics simulation of DNA stretching is consistent with the tension observed for

extension and strand separation and predicts a novel ladder structure. *J. Am. Chem. Soc.* 118:10989–10994.

24. Lavery, R., and A. Lebrun. 1999. Modelling DNA stretching for physics and biology. *Genetica.* 106:75–84.
25. MacKerell, Jr., A. D., and G. U. Lee. 1999. Structure, force, and energy of a double-stranded DNA oligonucleotide under tensile loads. *Eur. Biophys. J.* 28:415–426.
26. Harris, S. A., Z. A. Sands, and C. A. Laughton. 2005. Molecular dynamics simulations of duplex stretching reveal the importance of entropy in determining the biomechanical properties of DNA. *Biophys. J.* 88:1684–1691.
27. Piana, S. 2005. Structure and energy of a DNA dodecamer under tensile load. *Nucleic Acids Res.* 33:7029–7038.
28. Rueda, M., C. Ferrer-Costa, …, M. Orozco. 2007. A consensus view of protein dynamics. *Proc. Natl. Acad. Sci. USA.* 104:796–801.
29. Jurecka, P., J. Cerný, …, D. R. Salahub. 2007. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *J. Comput. Chem.* 28:555–569.
30. Jurecka, P., and P. Hobza. 2003. True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine…cytosine, adenine…thymine, and their 9- and 1-methyl derivatives: complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. *J. Am. Chem. Soc.* 125:15608–15613.
31. Sponer, J., P. Jurecka, and P. Hobza. 2004. Accurate interaction energies of hydrogen-bonded nucleic acid basepairs. *J. Am. Chem. Soc.* 126:10142–10151.
32. Strunz, T., K. Oroszlan, …, M. Hegner. 2000. Model energy landscapes and the force-induced dissociation of ligand-receptor bonds. *Biophys. J.* 79:1206–1212.
33. Case, D. A., T. A. Darden, III, …, P. A. Kollman. 2004. AMBER. University of California, San Francisco, CA.
34. Cheatham III, T. E., P. Cieplak, and P. A. Kollman. 1999. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* 16:845–862.
35. Pérez, A., I. Marchán, …, M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* 92:3817–3829.
36. Várnai, P., D. Djuranovic, …, B. Hartmann. 2002. α/γ Transitions in the B-DNA backbone. *Nucleic Acids Res.* 30:5398–5406.
37. Shields, G. C., C. A. Laughton, and M. Orozco. 1997. Molecular dynamics simulations of the d(T.A.T) triple helix. *J. Am. Chem. Soc.* 119:7463–7469.
38. Tsui, V., and D. A. Case. 2000. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.* 122:2489–2498.
39. Harris, S. A., E. Gavathiotis, …, C. A. Laughton. 2001. Cooperativity in drug-DNA recognition: a molecular dynamics study. *J. Am. Chem. Soc.* 123:12658–12663.
40. Bayly, C. I., P. Cieplak, …, P. A. Kollman. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the RESP model. *J. Phys. Chem.* 97:10269–10280.
41. Tao, J., J. P. Perdew, …, G. E. Scuseria. 2003. Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* 91:146401.
42. Schafer, A., C. Huber, and R. Ahlrichs. 1994. Fully optimized contracted Gaussian-basis sets of triple ζ valence quality for atoms Li to Kr. *J. Chem. Phys.* 100:5829–5835.
43. Feyereisen, M., G. Fitzgerald, and A. Komornicki. 1993. Use of approximate integrals in abinitio theory—an application in Mp2 energy calculations. *Chem. Phys. Lett.* 208:359–363.
44. Ahlrichs, R., M. Bar, …, C. Kolmel. 1989. Electronic-structure calculations on workstation computers—the program system Turbomole. *Chem. Phys. Lett.* 162:165–169.
45. Rezac, J., and P. Hobza. 2007. On the nature of DNA-duplex stability. *Chemistry.* 13:2983–2989.
46. Hubbard, S. J., and J. M. Thornton. 1993. "NACCESS" computer program. University College, London.
47. Jurecka, P., J. Sponer, …, P. Hobza. 2006. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA basepairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* 8:1985–1993.
48. Randall, G., L. Zechiedrich, and B. Pettitt. 2009. In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res.* 37:5568–5577.
49. Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.
50. Wereszczynski, J., and I. Andricioaei. 2006. On structural transitions, thermodynamic equilibrium, and the phase diagram of DNA and RNA duplexes under torque and tension. *Proc. Natl. Acad. Sci. USA.* 103:16200–16205.
51. Harris, S. A., C. A. Laughton, and T. B. Liverpool. 2008. Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucleic Acids Res.* 36:21–29.
52. Reference deleted in proof.
53. Piana, S. 2007. Atomistic simulation of the DNA helix-coil transition. *J. Phys. Chem. A.* 111:12349–12354.
54. Lankas, F., R. Lavery, and J. H. Maddocks. 2006. Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure.* 14:1527–1534.
55. Curuksu, J., M. Zacharias, …, K. Zakrzewska. 2009. Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res.* 37:3766–3773.
56. Comer, J., V. Dimitrov, …, A. Aksimentiev. 2009. Microscopic mechanics of hairpin DNA translocation through synthetic nanopores. *Biophys. J.* 96:593–608.
57. Gell, C., T. Sabir, …, P. G. Stockley. 2008. Single-molecule fluorescence resonance energy transfer assays reveal heterogeneous folding ensembles in a simple RNA stem-loop. *J. Mol. Biol.* 384:264–278.