Rigid Base Biasing in Molecular Dynamics Enables Enhanced Sampling of DNA Conformations

Voorspoels, A.; Vreede, J.; Carlon, E.

DOI
10.1021/acs.jctc.2c00889

Publication date
2023

Document Version
Final published version

Published in
Journal of Chemical Theory and Computation

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

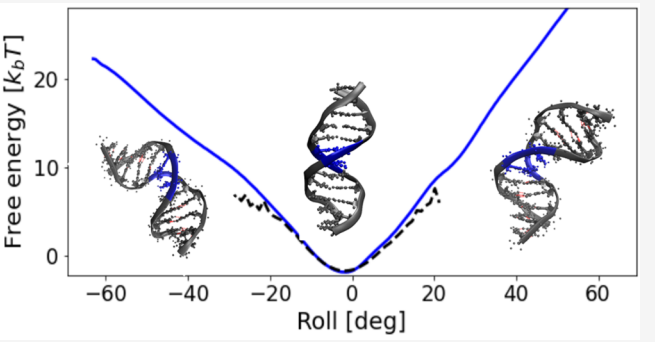
Rigid Base Biasing in Molecular Dynamics Enables Enhanced Sampling of DNA Conformations

Aderik Voorspoels,* Jocelyne Vreede,* and Enrico Carlon*

Cite This: *J. Chem. Theory Comput.* 2023, 19, 902−909 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: All-atom simulations have become increasingly popular to study conformational and dynamical properties of nucleic acids as they are accurate and provide high spatial and time resolutions. This high resolution, however, comes at a heavy computational cost, and, within the time scales of simulations, nucleic acids weakly fluctuate around their ideal structure exploring a limited set of conformations. We introduce the RBB-NA algorithm (available as a package in the Open Source Library PLUMED), which is capable of controlling rigid base parameters in all-atom simulations of nucleic acids. With suitable biasing potentials, this algorithm can "force" a DNA or RNA molecule to assume specific values of the six rotational (tilt, roll, twist, buckle, propeller, opening) and/or the six translational parameters (shift, slide, rise, shear, stretch, stagger). The algorithm enables the use of advanced sampling techniques to probe the structure and dynamics of locally strongly deformed nucleic acids. We illustrate its performance showing some examples in which DNA is strongly twisted, bent, or locally buckled. In these examples, RBB-NA reproduces well the unconstrained simulations data and other known features of DNA mechanics, but it also allows one to explore the anharmonic behavior characterizing the mechanics of nucleic acids in the high deformation regime.



1. INTRODUCTION

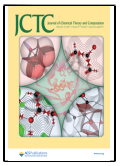
All-atom molecular dynamics (MD) simulations have become standard tools to investigate the mechanical properties of DNA.[1−5] These properties are of high relevance in biological processes, as DNA is often physically deformed by proteins or other biomolecules (see refs 6−8 for recent reviews on DNA mechanics).

DNA force fields have been constantly improved over the years using inputs from experiments and quantum mechanical calculations,[9] and simulations have reached a high degree of sophistication. At present DNA simulations also have some drawbacks. They are CPU-intensive and can thus sample molecular trajectories for only short time intervals (∼1 μs) and short DNA sequences (∼20−30 base-pairs). In these trajectories a DNA molecule weakly fluctuates around its ideal double helical structure. To investigate strongly deformed conformations, one has to resort to advanced sampling techniques. Many enhanced sampling techniques add potentials that constrain or drive the system along a predefined collective variable. This collective variable typically involves several atoms and represents a relevant degree of freedom.

In the past, umbrella sampling was used to induce strong bending by constraining the two ends of a linear DNA molecule,[10] mimicking the bending induced by large protein complexes. Likewise, by applying a force or a torque at the two DNA ends, previous studies have stretched or twisted a DNA molecule.[11,12] In the above studies, stretching or twisting is typically distributed over several base-pairs (as in most DNA−protein-bound structures), which are allowed to locally relax. In this way the DNA conformations induced by an extended nonlocal constraint can be studied.[10−12]

To gain more insights into the local origin of the mechanical deformability, one should apply suitable constraints at the base or base-pair geometry. A biased MD scheme to induce local deformations to perform enhanced sampling simulations of DNA was developed by van der Vaart and collaborators.[13−15] These authors use a simplified representation of DNA step parameters; for instance, they define a *pseudoroll* bending angle,[13] which differs from the canonical variable "roll" as conventionally defined.[16] With the exceptions of a few cases (most notably the rise[14]), the simplified rigid base parameters of van der Vaart and collaborators are found to correlate rather well with the canonical ones.

Received: August 29, 2022
Published: January 25, 2023



In this paper we introduce RBB-NA (rigid base biasing of nucleic acids), an algorithm which, by using suitable local constraints, is capable of deforming a DNA molecule by bending, stretching, or twisting at the single base-pair level. RBB-NA uses the canonical 12 rigid base coordinates tilt, roll, twist, etc., as they are defined in Curves+,[16] as collective variables in advanced sampling techniques. This makes our biasing scheme more complex than those of refs 13−15 but likely more accurate in the high deformation regime. Moreover, RBB-NA is made available as a package in the Open Source Library PLUMED,[17,18] implemented as a collective variable. PLUMED already contains much functionality to allow advanced sampling techniques, such as metadynamics, umbrella sampling, and various path sampling techniques.

Although we focus on a few illustrative examples on DNA, the algorithm also works with RNA. Different constraints can be applied simultaneously (as, e.g., bend and twist) either to a single site or to multiple sites, contiguous or not, of a DNA sequence. As such, RBB-NA is a very flexible tool to investigate the structure, energetics, and dynamics of highly deformed DNA molecules in MD simulations.

2. METHODOLOGY

2.1. The RBB-NA Algorithm. The aim of the algorithm is to compute suitable constraint forces $\vec{F}_i^{(c)}$ which act on a group of atoms (i labels a given atom) and induce biases on rigid base coordinates. The flowchart shown in Figure 1 depicts the operations of RBB-NA. It essentially consists of three different steps (labeled as 1, 2, and 3 in Figure 1) which are schematically described in the rest of this section.

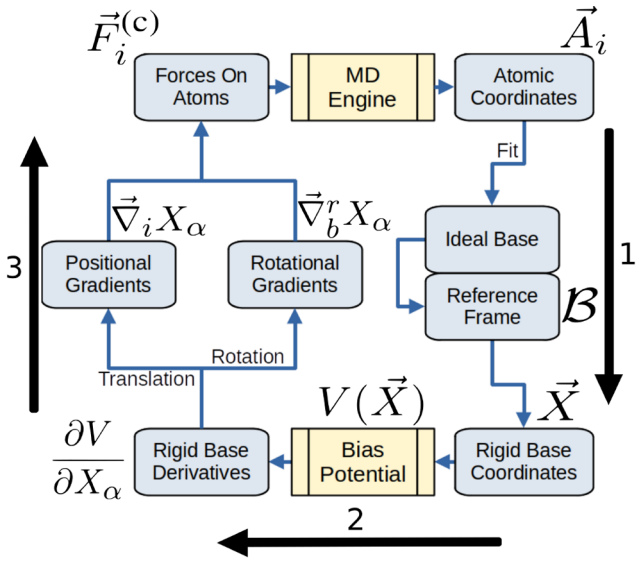


Figure 1. Flowchart showing the operations done by RBB-NA. The numbered arrows correspond to (1) mapping atomic coordinates onto rigid base coordinates, as done in Curves+, (2) introducing constraints at the rigid base level using the potential $V(\vec{X})$, and (3) calculating constraints forces on atoms $\vec{F}_i^{(c)}$. This third step consists of translational and rotational constraints. Total forces used in the MD integration step are the sum of constraints ($\vec{F}_i^{(c)}$) and force fields contributions ($\vec{F}_i^{(ff)}$). Constraints are typically applied to a limited set of atoms in the bases; hence, the running of the algorithm does not significantly slow down the running of the MD code.

2.1.1. From Atomic to Rigid Base Coordinates. This first step performs a mapping from atomic coordinates $\vec{A}_i$ to rigid base coordinates for the bases to which the constraint is applied. Rigid base coordinates have long been a standard description of the double helix down to the base level.[19] Software packages to extract rigid base coordinates, such as Curves+[16] and x3dna,[20] were developed in the past. Usually these packages are used in a postprocessing analysis step of MD trajectories. In our algorithm we have implemented Curves+ to map atomic coordinates to rigid base coordinates during the simulation run. We briefly recall how Curves+ works. To obtain the rigid base coordinates, one first associates to each base a reference frame $\mathcal{B} \equiv \{\mathbf{B}, \vec{r}\}$ consisting of an orthonormal triad of unit vectors $\mathbf{B} = [\hat{e}_1, \hat{e}_2, \hat{e}_3]$ (Figure 2) and their origin $\vec{r}$. This step

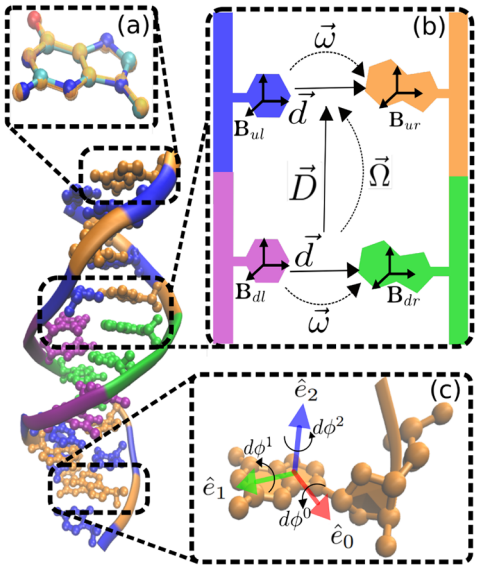


Figure 2. Illustration of different steps of the RBB-NA algorithm. (a) Mapping of atomic coordinates of a base (colored) into an ideal base configuration (orange). (b) From the reference frames $\mathcal{B} = (\mathbf{B}, \vec{r})$ consisting of an orthonormal triad $\mathbf{B}$ and its origin $\vec{r}$, one calculates translational ($\vec{d}$, $\vec{D}$) and rotational ($\vec{\omega}$, $\vec{\Omega}$) rigid base coordinates. ($\vec{d}$, $\vec{\omega}$) are intrabase coordinates, while ($\vec{D}$, $\vec{\Omega}$) are interbase coordinates. (c) Rigid body rotations are used to calculate the constraint forces $\vec{F}^{(c)}$. Steps (a) and (b) are as in the Curves+ algorithm.[16]

conventionally involves fitting a set of "ideal" base coordinates to the actual atomic coordinates measured in the simulation. To parametrize the translations and rotations connecting two complementary base frames on the opposite strands of a DNA molecule, one uses the rigid base coordinates $\vec{d}$ (a vector whose components are referred to as shear, stretch, and stagger in the DNA literature[19]) and $\vec{\omega}$ (buckle, propeller, and opening). The coordinates $\vec{D}$ (shift, slide, and rise) and $\vec{\Omega}$ (tilt, roll, and twist) parametrize translations and rotations between two consecutive base-pairs, see Figure 2. More details are given in the Supporting Information.

2.1.2. The Constraint Potential. For a sequence of N base-pairs, one has N base-pair coordinates ($\vec{d}$, $\vec{\omega}$) and $N - 1$ base-pair steps coordinates ($\vec{D}$, $\vec{\Omega}$). We collect all rigid base coordinates into a single $12N - 6$ dimensional vector

$$\vec{X} \equiv (\vec{\omega}_1, \vec{d}_1, ..., \vec{\omega}_N, \vec{d}_N, \vec{\Omega}_1, \vec{D}_1, ..., \vec{\Omega}_{N-1}, \vec{D}_{N-1}) \quad (1)$$

https://doi.org/10.1021/acs.jctc.2c00889
J. Chem. Theory Comput. 2023, 19, 902−909

In principle, the constraint potential is a generic function $V(\vec{X})$ of all these coordinates. In practice, we fix constraints to a limited number of base-pairs and rigid base coordinates, as shown in the examples discussed in Results and the Supporting Information.

2.1.3. Translational and Rotational Constraints. In order to implement the constraints in the MD simulation, one needs to compute the forces acting on each particle. The constraint force on the ith particle is given by

$$\vec{F}_i^{(c)} = -\vec{\nabla}_i V(\vec{X}) = -\sum_\alpha \frac{\partial V}{\partial X_\alpha}\vec{\nabla}_i X_\alpha \tag{2}$$

where the sum runs through all $12N - 6$ components of the vector $\vec{X}$, α labels rigid base coordinates, and $\vec{\nabla}_i$ is the gradient with respect to the Cartesian coordinates of the ith particle. The derivatives $\frac{\partial V}{\partial X_\alpha}$ are trivial to compute as the function $V(\vec{X})$ is analytically simple (see examples in Results). To compute $\vec{\nabla}_i X_\alpha$, one would need to know the function that maps atomic coordinates into rigid base coordinates $\vec{X}$. However, the relation between the two sets of variables is complex, in part because it makes use of fitting procedures. Unfortunately, these fitting methods make the mapping nondifferentiable. Moreover, the mapping is not one-to-one, as there are multiple atomic configurations that map to the same rigid base coordinate set. The approximation made here to resolve these issues is to treat the bases as rigid when calculating and applying constraint forces. To incorporate the entire base into the computation of $\vec{X}$ and to reduce noise, ideal base atomic coordinates are fitted to the measured ones. Afterward, when forces are calculated, the measured coordinates are taken to represent a rigid body.

There are two different types of rigid base coordinates: those associated with translations $(\vec{d}, \vec{D})$ and those associated with rotations $(\vec{\omega}, \vec{\Omega})$. The calculation of $\vec{\nabla}_i X_\alpha$ for translations is trivial. This is just a unit vector parallel to the component of the translation X_α. For rotations the construction of $\vec{\nabla}_i X_\alpha$ is less straightforward. It requires the computation of rotational gradients $\vec{\nabla}_b^r X_\alpha$ with respect to the orientation of the bases as a whole, from which torques can be calculated. Using rigid body dynamics, these torques can then be converted to forces. Details of this procedure are given in the Supporting Information.

Once the constraint forces $\vec{F}_i^{(c)}$ on all the atoms in the constrained bases are computed, we add to these the force fields contributions $\vec{F}_i^{(ff)}$, so the total force on the ith atom is

$$\vec{F}_i^{tot} = \vec{F}_i^{(c)} + \vec{F}_i^{(ff)} \tag{3}$$

This is used in the MD step to update positions and velocities from the current time t to $t + \Delta t$. We note that in our scheme only atoms in bases are constrained, as these are the atoms used by Curves+ to calculate rigid base coordinates. Atoms in the backbone are not constrained.

2.2. Simulations. *2.2.1. System Preparation.* All simulations were done using version 2020.4 of Gromacs,[21] version 2.8.0 of PLUMED,[17] and the Amberff99 parmbsc1 force field (based on the parm99 force field, improved with bsc0 corrections[22] and bsc1 corrections[9]). The RBB-NA algorithm has been implemented to work with Gromacs; however, because PLUMED is a highly portable plug-in, it can be made to work with a multitude of MD engines after some minor adjustments. Water was modeled using the TIP-3P model,[23] nonbonded interactions were cut off at 1.0 nm, and PME Mesh Ewald interactions were used for electrostatics. All simulations started from the structure of the Drew−Dickerson (DD) dodecamer,[24] a self-complementary oligomer with sequence CGCGAATTCGCG which has been extensively investigated in prior experimental and computational studies (see, e.g., ref 25). The DD sequence was placed into a dodecahedral box, leaving 2.0 nm on either side of the molecule, with periodic boundary conditions and solvated in a 150 mM NaCl solution after which the overall charge in the system was neutralized. This structure was energy minimized with a tolerance of 1000 kJ/mol to make sure no overlap remained between solvent molecules and DNA. Subsequently, the molecule was equilibrated in the NVT ensemble for 100 ps where the temperature was kept at 300 K using a velocity rescaling thermostat[26] and then equilibrated for another 100 ps in the NPT ensemble at the same temperature but using a Parrinello−Rahman barostat[27] to fix the pressure at 1.0 bar. These first equilibration simulations were performed using a 2 fs time step in a leapfrog integrator, using LINCS[28] to constrain the covalent bonds involving hydrogen atoms.

2.2.2. Production Runs. A reference unconstrained MD simulation of 100 ns was performed with the same settings as the NPT equilibration. Snapshots were stored every 1 ps, and collective variables were printed with the same frequency. Subsequently four sets of simulations were done using the RBB-NA algorithm. All four were umbrella sampling runs which will be analyzed for different applications in the Results Section. In the runs we used three different types of constraints induced by parabolic potentials

$$V_\alpha = \frac{K_\alpha}{2}(X_\alpha - \bar{X}_\alpha)^2 \tag{4}$$

where X_α is either roll, twist , or buckle (in the Supporting Information we discuss some additional simulations constraining also the opening variable, using an anharmonic bias.) The roll and twist constraints were applied on the central base-pair step (AT) of the DD sequence. The buckle constraint was applied on the second GC pair of the DD sequence. The first three runs were done using this type of constraint for only one parameter at a time; in the fourth run, constraints on roll and twist were applied simultaneously. In all constrained simulations, we set the spring constant to K_α = 1000 kJ/mol.

For the case of the roll simulation run, the system was first pulled from its equilibrium configuration. This was done by running molecular dynamics in consecutive 100 ps pulling windows with the same settings as the NPT equilibration but decreasing the time step to 0.2 fs. The time step was reduced to avoid pulling past LINCS constraints.[28] In the present case we used a conservative small value to avoid any problems with the stability of the algorithms. In future work this time step could be optimized and chosen to be different for different constraints. In these runs the average roll, $\bar{X}_\alpha$ in the potential 4, was increased (or decreased) by 0.01 radians (starting from a zero roll average) every pulling window. Here a maximum of 1.0 radians and a minimum of −1.0 radians were reached. The system resulting from every fifth pulling window (i.e., with $\bar{X}_\alpha$ increasing in steps of 0.05 radians) was then used to start a 10 ns production run. Here again the same settings as in the NPT equilibration were used, now setting the time step back to 2.0 fs. This resulted in a total of 41 windows of 10 ns with the imposed average roll spanning from 1.0 to −1.0 radians.

https://doi.org/10.1021/acs.jctc.2c00889
J. Chem. Theory Comput. 2023, 19, 902−909

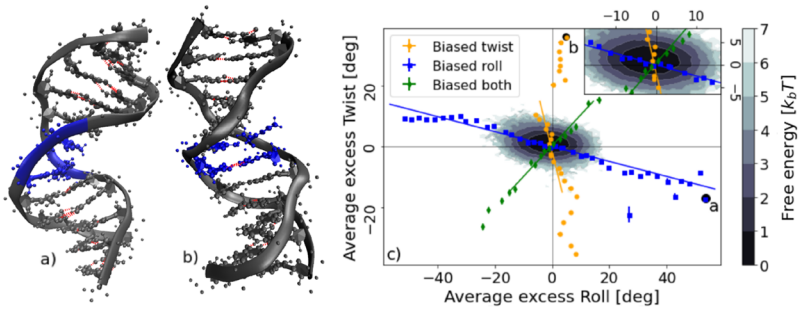


Figure 3. (a, b) Snapshots of simulations using the RBB-NA algorithm in which either (a) roll or (b) twist is constrained, via eq 5 (K_r = 1000 kJ/mol and $\bar{\Omega}_2 = 1.0$ radian (57.3°)) or eq 6 (K_t = 1000 kJ/mol and $\bar{\Omega}_3 = 0.7$ radian (40.1°)), respectively. In both snapshots the consecutive blue-colored base-pairs are those to which the constraint is applied. We note that a bias enforcing positive roll induces an undertwist in the molecule (a), as expected from positive twist−roll coupling. While (a) shows an intact DNA molecule, base-pairing is disrupted in (b) when a strong overtwist is applied. (c) Colored symbols: average excess roll vs average excess twist in simulations biasing the twist via eq 6 (orange circles), the roll via eq 5 (blue squares), and both with $\bar{\Omega}_2 = \bar{\Omega}_3$ (green diamonds). For small excess roll and twist, the relation between $\langle\Omega_2\rangle$ and $\langle\Omega_3\rangle$ is linear, as expected in the harmonic regime (solid lines, see the Supporting Information for details). The contour plot, shown in detail in the inset, is a free energy calculation from unbiased MD simulations. The values of Ω_2 and Ω_3 corresponding to the snapshots (a) and (b) are indicated with black dots in (c).

After the production runs for umbrella sampling on roll, a similar procedure was carried out for twist in the central step. Again the system was pulled from the equilibrium configuration in 100 ps windows (with the same settings as in the NPT equilibration but decreasing the time step to 0.2 fs), using a potential of the shape (eq 4), while increasing and decreasing the imposed average $\bar{X}_\alpha$ in steps of 0.01 radians. Notably for these twist constraining simulations, the intrinsic twist of DNA was taken into account using 0.61 radians as the starting average. The maximum imposed twist reached was 1.31 radians, and the minimum was −0.09 radians. Like for roll every fifth window was used to start a production run of 10 ns, with the same settings as in the NPT equilibration restoring the time step to 2 fs. This resulted in 29 windows spanning an imposed average twist from −0.09 to 1.31 radians.

For constraining the rigid base coordinate buckle of the second GA pair, again the same procedure was used. The system was pulled in 100 ps windows (with the same settings as in the NPT equilibration but decreasing the time step to 0.2 fs) with the buckle changing from 0.0 radians in 0.02 rad steps. The maximum reached here was 1.3 radians, and the minimum was −1.3 radians. The system resulting from every fifth pulling window (i.e., with the imposed average buckle increasing in steps of 0.1 radians) was then used to start a 10 ns production run. This resulted in 17 windows.

Finally, the run combining twist and roll constraints was again done using the same procedure. The imposed excess rolls and twists were set equal to each other. Pulling occurred in 100 ps windows with a time step of 0.2 fs, changing the imposed deformation by 0.01 radians. Production runs again lasted 10 ns with sampling occurring in 20 equidistant windows between −0.5 and 0.5 radians.

2.3. Weighted Histogram Analysis. A detailed mathematical description of umbrella sampling and the weighted histogram analysis method (WHAM) that is used to reconstruct the free energy landscape has been given by J. Kaestner.[29] In short, the method entails constraining the conformation space of a DNA molecule by imposing a steep potential, which is usually quadratic in the desired coordinate, as in the examples in eq 6and eq 5. By changing the location of the minimum of the constraint potential (e.g., by varying $\bar{\Omega}_3$ and $\bar{\Omega}_2$ in eq 6and eq 5), one can then control which region of the conformation space is sampled. If the sampled regions partially overlap, the relative free energy landscape over all regions can be reconstructed afterward using WHAM analysis. This involves an iterative procedure[29] which here was considered to have converged if the integral of the difference between the free energies of the previous and current iterations was below $10^{-7}k_BT$.

3. RESULTS

The RBB-NA algorithm was developed to enable umbrella sampling to characterize the free energy landscape by constraining local rigid base coordinates of DNA. However, the same algorithm can also be used in any other type of enhanced sampling applications, such as metadynamics or steered MD. We conducted three sets of umbrella sampling simulations, one imposing a local twist, one imposing a local roll, and one imposing a local buckle (additional constrained simulations on the opening coordinate are discussed in the Supporting Information). From these simulations we present three different types of analysis which illustrate the functioning of the algorithm. In the first analysis, we explore how imposing a local roll or twist affects other rigid base coordinates. We find that a bias in the twist induces a nonvanishing excess roll and, vice versa, a bias in the roll induces a nonvanishing excess twist. This is a clear demonstration of the effect of twist−roll coupling.[30] In the second analysis, we compute free energy profiles along the imposed twist or roll coordinate. Here our analysis points to anharmonic behavior. Finally, the third analysis illustrates the effect of a local bias on base-pairs flanking the constrained site.

3.1. Twist−Roll Coupling. Figure 3a shows a snapshot of an MD simulation in which an average excess roll of ∼60° is imposed on the central base-pair step (AT) of the DD sequence for which a local constraint potential of parabolic type was used

$$V_{\text{roll}} = \frac{K_r}{2}(\Omega_2 - \bar{\Omega}_2)^2 \tag{5}$$

(we use Ω_1, Ω_2, and Ω_3 to indicate the excess tilt, roll, and twist). As shown in the snapshot of Figure 3a, the bending induces a remarkably strong twisting deformation as well due to the effect of the twist−roll coupling.[30] In homogeneous DNA models such interaction is described by a term of the type $G\Omega_2\Omega_3$, with G the twist−roll coupling. Twist−roll coupling and its influence on the

https://doi.org/10.1021/acs.jctc.2c00889
J. Chem. Theory Comput. 2023, 19, 902−909

mechanical properties of DNA have been discussed in the recent literature.[31−33] Figure 3b shows the snapshot of a conformation generated by a constraint potential of the type

$$V_{\text{twist}} = \frac{K_t}{2}(\Omega_3 - \bar{\Omega}_3)^2 \quad (6)$$

which, depending on the sign of $\bar{\Omega}_3$, locally over- or undertwists DNA. We note that in the snapshot of Figure 3b the overtwisting induces hydrogen bond breaking with a base flipping out of the strand.

Figure 3c plots the average excess twist vs average excess roll for a roll bias (eq 5), for a twist bias (eq 6), and for simultaneous twist and roll bias. In the latter case we imposed a bias which is the sum of eqs 5 and 6 with $\bar{\Omega}_2 = \bar{\Omega}_3$. Note that, for weak deformations, average excess twist and roll, $\langle\Omega_2\rangle$ and $\langle\Omega_3\rangle$, induced by the biasing potentials are linearly correlated. Moreover, they follow lines with different slopes (orange, blue, and green lines in Figure 3c), a behavior which can be explained by a simple harmonic model calculation, see the Supporting Information. For pure twist or roll bias, induced excess twist and roll also have opposite signs (orange and blue lines in Figure 3c), as expected for a positive twist−roll coupling $G > 0$. The contour lines in Figure 3b show the free energy calculated by Boltzmann inversion ($F = -k_B T \ln(P)$, with P the equilibrium probability distribution) from the sampling of unconstrained MD simulations. Overall these data show the characteristic behavior observed in protein−DNA crystal structures[34] as well as in other simulation studies. The RBB-NA algorithm generates DNA conformations with roll and twist angles which greatly exceed those generated by unbiased simulations. We note that in the high deformation regime the excess twist and roll are more weakly coupled to each other than in the harmonic regime. For instance, biasing the excess twist above 20° does not produce a significant roll, and a roll bias below −30° does not lead to an increase of the twist, see Figure 3c. Hydrogen bond breakage and base flipping produce a jump in the twist vs roll data of Figure 3c (visible at an overtwist of 20°, see the orange set). RBB-NA keeps working also for average excess twist beyond the discontinuity at 20°, in the sense that it is capable of inducing higher twist angles when $\bar{\Omega}_3$ is increased. (Here the "twist" variable is mathematically defined as in Curves +. With such definition we can define twist also for noncanonical conformations where bases are unpaired.) While we show in Figure 3c the effect of biasing twist and/or roll, the RBB-NA can bias any of the 12 base step parameters. A bias on twist and on the "opening" parameter (describing the opening angle between complementary base-pairs) is discussed in the Supporting Information.

While the mechanics of DNA in the weak deformations regime has been intensively studied,[3,4] the strong deformation regime is still largely unexplored. This regime is within the reach of the RBB-NA algorithm.

3.2. Free Energies from Umbrella Sampling. The tight control over rotation angles of the RBB-NA algorithm can be used for umbrella sampling to obtain free energies. Figure 4 shows a comparison between free energies obtained by unconstrained simulations (dashed lines) and with the RBB-NA algorithm via umbrella sampling (solid lines). The latter are calculated using the constraints 6 and 5. A much wider range of twist and roll angles is available via umbrella sampling. For roll, Figure 4a, , beyond a narrow interval around the free energy minimum, the landscape no longer follows a parabolic shape.

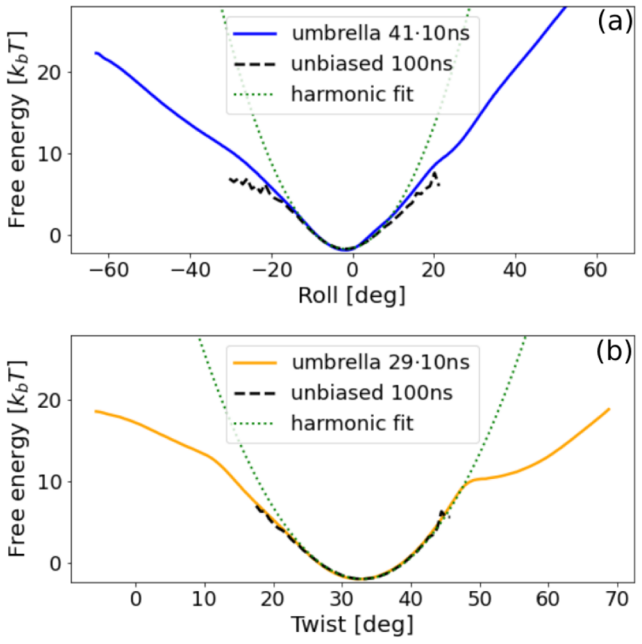


Figure 4. Free energy of roll (a) and twist (b) deformation of the central (AT) step of the DD dodecamer. Solid orange line: umbrella sampling with the RBB-NA algorithm. Dashed lines: sampling from unconstrained simulations. The latter can generate only small deformations around the minimal free energy state. The umbrella sampling reproduces very well the unconstrained simulations data. Dotted lines: parabolic fits to the unconstrained simulation. Both examples show deviations from the "ideal" parabolic behavior indicating that highly bent or highly twisted deformations are energetically less costly than a quadratic model would predict. Both data show asymmetries with respect to positive or negative excess roll and twist.

The roll data instead suggest a behavior reminiscent of the linear subelastic chain (LSEC) model, which was introduced to explain anomalous high bending of DNA as observed in atomic force microscopy (AFM) experiments[35] (we note that AFM data analysis has led to some debate and to different conclusions about the existence of an actual anomalous bending behavior[36]). The LSEC model posits that the bending free energy of DNA is quadratic for low bending angles, while it is linear beyond a given threshold.[35] This implies that large bend angles are less costly than the quadratic model would predict. A free energy consistent with the LSEC model was also observed in prior all-atom simulations,[10] where a short DNA sequence was forced to bend by applying a constraint at its two ends. Such constraint induces bending which is distributed over several nucleotides, and it is perhaps more appropriate for a direct comparison with AFM data, where DNA bending is likely delocalized over a few base-pairs. The RBB-NA localized constraint of Figure 4 acts only on two adjacent base-pairs, and it is useful for gaining more insight on the microscopic origin of the deviations from the quadratic behavior. The bending free energies from AFM experiments[35] and the RBB-NA data of Figure 4a are qualitatively similar, but there are quantitative differences. In AFM data,[35] as well as in globally constrained simulations,[10] the deviation from parabolic behavior appears at rather high bending angles (≈60°) as compared to those seen in Figure 4a, which occur for roll angles of ≈20°. This is consistent with the idea that bending deformations in free DNA on surfaces, as studied in AFM experiments, are delocalized over several base-pairs. We note

https://doi.org/10.1021/acs.jctc.2c00889
J. Chem. Theory Comput. 2023, 19, 902−909

that the data in Figure 4a are in quite good agreement with umbrella sampling simulations constraining a pseudoroll angle.[13] (The quasi-linear behavior found in simulations[13] at high roll was attributed to ionic screening due to counterions aggregation in the concave side of a strongly bent conformation.)

The free energy landscape for twist is shown in Figure 4b. Again we note a good overlap between unconstrained and umbrella sampling simulations. Compared to the roll data, the twist follows a harmonic behavior for a wider range of angles. For high over- and undertwisting we observe deviations from the parabolic free energy profile, which indicate that highly deformed twisted conformations are more likely than a quadratic model would predict. The deviations from the harmonic behavior are however different than in roll, showing more abrupt transitions. The overtwisted regime closely follows the quadratic shape predicted by the TWLC until a total twist of ≈48° is reached. There, the landscape shows a sharp transition which, upon close inspection of the trajectories, can be linked to a structural change involving partial breaking of the hydrogen bonds (see the snapshot of Figure 3b). In the undertwisted regime one can also note an abrupt change of behavior in the free energy around a total twist of 15°, although less sharp than in the overtwisted case. We note that here we have sampled free energies for roll and twist as induced by biasing via eqs 6 and 5. This analysis provides just a "one-dimensional" projection of the two-dimensional free energy $f(\Omega_2, \Omega_3)$. The full calculation of this free energy will be presented elsewhere.

3.3. Effect of a Local Bias on Neighboring Sites. Several studies have observed correlations between rigid base coordinates separated by a few nucleotides along the sequence, indicating the existence of couplings between distal sites.[1,2,37] A recent analysis quantified these nonlocal couplings for tilt, roll, and twist for both DNA[38] and dsRNA.[39] Very similar couplings were found in these two molecules.[39] Here we analyze structural deformations along the whole sequence of a DD dodecamer induced by local perturbations at a specific site.

Figure 5a shows the average excess twist in the DD sequence upon imposing an excess twist via a potential of the type 6 on the central AT site, with the different lines referring to different strengths of the perturbation. We note that the perturbation produces an oscillating decay of the twist. This is in line with the predictions of nonlocal models of DNA and dsRNA elasticity.[38,39] This behavior can be understood by a minimal (homogeneous) model of twist free energy E_{twist} of the following type

$$E_{twist} = \frac{1}{2}\sum_n [\tilde{C}\Omega_3^2(n) + \tilde{C}'\Omega_3(n)\Omega_3(n+1)] \quad (7)$$

where the sum runs through all the sites and $\tilde{C}$ and $\tilde{C}'$ are the on-site and next-neighbor twist−twist couplings and $\Omega_3(n)$ is the excess twist at site n. For simplicity we have limited the off-site interactions to a coupling between consecutive sites n and $n \pm 1$, but this argument can be easily extended to interactions coupling more distant sites, as n and $n + 2$, etc.[38] Prior work showed that both DNA and dsRNA are characterized by $\tilde{C}' > 0$,[39] while stability requires that $\tilde{C} > 0$. Let us consider a perturbation of type 6 enforcing $\langle\Omega_3(k)\rangle > 0$ at a given site k. As $\tilde{C}' > 0$ energy minimization induces $\langle\Omega_3(k \pm 1)\rangle < 0$ at the two flanking sites. This ultimately produces an alternating decaying profile of over- and undertwist, which is indeed observed in Figure 5a. The fact that a localized twist perturbation reproduces the expected pattern of twist decay is an indication that the RBB algorithm performs correctly.

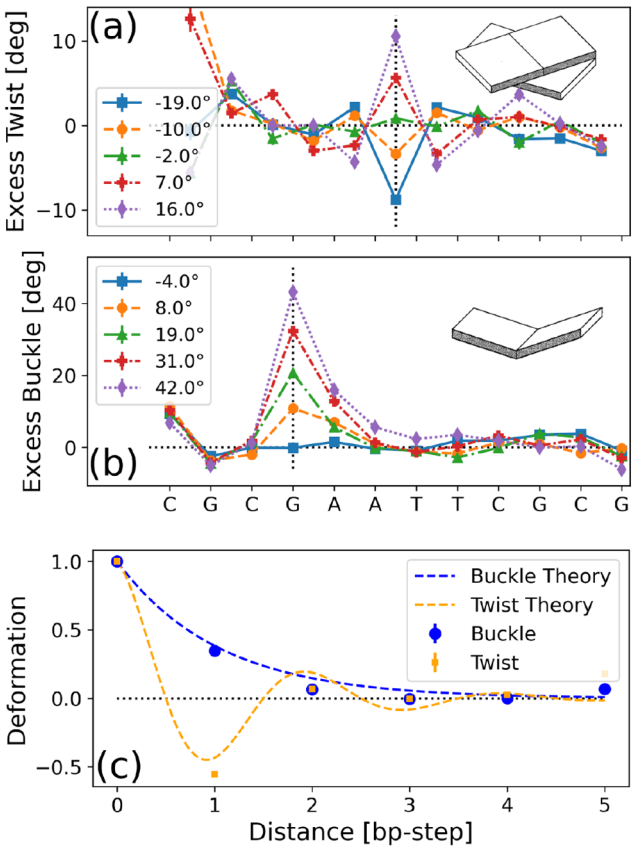


Figure 5. (a) Average excess twist induced along the molecule by a twist deformation imposed by a constraint of type 6 in the central AT step of the DD dodecamer. The twist has an oscillatory decay as predicted by a twist free energy of type 7 with $\tilde{C}' > 0$. Note that Curves+ reports an anomalous high twist at the left end of the sequence, presumably due to some end effects. (b) Average excess buckle induced by a constraint of type 8 applied to the second G nucleotide from the left. (c) Average decays of twist and buckle from MD simulations (symbols) and predictions from the continuum nonlocal twistable wormlike chain model.[38,39] We have explicitly added to the data points the error on the mean, defined as $\sigma/\sqrt{M}$, with σ the standard deviation and M the number of independent samples. For most of the data, this is smaller than symbol sizes.

Finally, we used RBB-NA to induce a buckle (ω_1) which is an intrabase-pair parameter describing the mutual orientation of two complementary bases in the same pair. Buckle is the angle formed by a rotation around the pair short axis. A nonzero buckle was induced using the following constraint

$$V_{buckle} = \frac{K}{2}(\omega_1 - \bar{\omega}_1)^2 \quad (8)$$

Instead of an alternating pattern (as seen in twist), buckle shows a monotonic decay from the perturbation site, see Figure 5b. This type of behavior can be generated by a buckle free energy of the same form as eq 7, but with a negative off-site coupling (which would be given by $\tilde{C}' < 0$ in the twist case). We note that the decay of the buckle is asymmetric at the two sides of the perturbation, which may be due to sequence or end effects (the biasing potential is applied to the fourth base-pair of the dodecamer).

https://doi.org/10.1021/acs.jctc.2c00889
J. Chem. Theory Comput. 2023, 19, 902−909

Figure 5c compares the results of simulations directly to the theory of nonlocal DNA elasticity developed in ref 38 . For twist, good agreement is found between the observed oscillatory decay and theoretical predictions (dashed lines). Buckle, which was not included in previous studies, also decays exponentially along the chain as a result of a local perturbation, a clear signature of an off-site (nonlocal) coupling. Parameters needed to determine the theoretical decay profile, local stiffness and off-site couplings, were determined using the same reference simulation used in the umbrella sampling simulations.

4. DISCUSSION

We have developed an algorithm (RBB-NA) which allows one to control rigid base parameters in DNA. In RBB-NA suitable bias potentials are used to impose specific values to one (or more than one) of the 12 rigid base parameters tilt, roll, twist, buckle, propeller, opening, shift, slide, rise, shear, stretch, and stagger. The algorithm uses rigid body dynamics to translate potentials on complex collective variables into constraining forces which act on atoms during simulations. RBB-NA can thus generate highly deformed DNA in a controlled manner at the local scale, e.g., at the base-pair level. We believe that the algorithm offers an interesting new tool to explore the dynamics of rare events and to map the full free energy landscape by using advanced sampling techniques, well-beyond what unconstrained simulations can reach.

After giving a schematic overview of the working of RBB-NA (see the Supporting Information for more mathematical details), we have discussed three examples. The first one shows how a bias in the twist actually induces an excess roll as well, and vice versa, a bias in the roll induces a nonzero excess twist. This is an illustration of the twist−roll coupling. Although this coupling is well-known in DNA[30] and also studied at length in the recent literature,[31−33] the fact that we can observe its effect in RBB-NA is an indication that the algorithm works correctly. We find indeed that a positive excess twist bias induces a negative roll and vice versa, as expected in the case of a positive twist−roll coupling constant $G > 0$, in agreement with prior studies.[19] Apparent from these results is also the opportunity RBB-NA provides to study the structure of highly deformed conformations of DNA, such as the snapshots shown in Figure 3a and b. Correspondingly, one can see from Figure 3c, comparing the area accessible in unconstrained sampling to the reach of the simulations using RBB-NA, the breadth of conformations which can now be accessed in a highly controlled manner.

The second example shows free energy calculations where RBB-NA was used to perform umbrella sampling simulations. We considered again roll and twist deformations as illustrative. The results of these simulations were analyzed using the weighted histogram analysis method. Notable here is that this relatively straightforward application of the algorithm already reveals some interesting features in the presented free energy landscape. Especially interesting is the LSEC behavior of roll. This result is in line with previous experimental[35] and computational[10,13] findings. RBB-NA can be used in umbrella sampling studies to identify other relevant and interesting features in the free energy landscape of dsDNA like was done for twist. Here, only "one-dimensional" projections of the entire free energy landscape have been discussed. RBB-NA might well be used to explore and map the full 12 dimensional free energy landscape or other interesting subsets thereof. This complete free energy landscape is likely to hold features (barriers, dips, local minima, ...) which are important for the biological functioning of DNA.

Finally, in a third example we investigated how a local perturbation propagates to neighboring sites. We showed two different types of spatial propagations for twist and buckle: twist decays through damped oscillations, while buckle shows a monotonic decay. This behavior can be attributed to the different signs of the nonlocal coupling terms, as discussed in ref 38.

Summarizing, we have presented a few examples which show how the RBB-NA algorithm is capable of locally controlling rigid base parameters. The local biasing induces different types of deformations of nucleic acids and provides quantitative insights on structural properties, which are out of the reach of standard unbiased simulations. While our main interest is in canonical DNA structures, characterized by Watson−Crick base-pairing, we found that strong constraints may induce different types of noncanonical structures as well. We discussed two of these and calculated their free energies in the Supporting Information. So far, we have restricted our analysis to a few examples on the DD dodecamer, while sequence-dependent effects in the large deformation limit remain to be explored. This analysis will be performed in future studies.

The biasing of RBB-NA is limited to the 12 rigid base coordinates. A possible extension of the algorithm may involve the biasing of other collective variables such as the groove width, the helical twist, the helical rise, and backbone-related conformations associated with sugar puckering. The biasing introduced in RBB-NA relies on rigid body rotations, applied to an idealized rigid base structure. For some simpler applications, rigid base biasing may not be needed, and simpler constraints involving global bending, twisting, or stretching are sufficient.[10−12] To conclude, we believe RBB-NA will serve as a useful tool in all-atom simulations aiming to characterize nucleic acids in terms of rigid body parameters.

■ ASSOCIATED CONTENT

Data Availability Statement

The RBB-NA code is made available through github: https://github.com/AderikVoorspoels/RBB-NA.git. Documentation concerning the algorithm and example scripts are available in the github repository.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00889.

A more detailed mathematical description of the RBB-NA algorithm; a theoretical investigation into the implication of biasing twist and roll DNA; an additional example showing the use of RBB-NA to prevent hydrogen bond breaking while overtwisting (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Aderik Voorspoels − *Soft Matter and Biophysics, Department of Physics and Astronomy, KU Leuven, 3000 Leuven, Belgium;* orcid.org/0000-0003-1955-2254; Email: aderik.voorspoels@kuleuven.be

Jocelyne Vreede − *Van't Hoff Institute for Molecular Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands;* orcid.org/0000-0002-6977-6603; Email: j.vreede@uva.nl

Enrico Carlon − *Soft Matter and Biophysics, Department of Physics and Astronomy, KU Leuven, 3000 Leuven, Belgium;* orcid.org/0000-0001-8266-1096; Email: enrico.carlon@kuleuven.be

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00889

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge interesting discussions with Midas Segers and Enrico Skoruppa.

■ REFERENCES

(1) Lankaš, F.; Šponer, J.; Langowski, J.; Cheatham, T. E. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* **2003**, *85*, 2872−2883.
(2) Noy, A.; Golestanian, R. Length Scale Dependence of DNA Mechanical Properties. *Phys. Rev. Lett.* **2012**, *109*, 228101.
(3) Pasi, M.; Maddocks, J. H.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C.; Mitchell, J.; Osman, R.; Orozco, M.; Pérez, A.; Petkevičiute, D.; Spackova, N.; Sponer, J.; Zakrzewska, K.; Lavery, R. μABC: A systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 12272−12283.
(4) Velasco-Berrelleza, V.; Burman, M.; Shepherd, J. W.; Leake, M. C.; Golestanian, R.; Noy, A. SerraNA: a program to determine nucleic acids elasticity from simulation data. *Phys. Chem. Chem. Phys.* **2020**, *22*, 19254.
(5) Walther, J.; Dans, P. D.; Balaceanu, A.; Hospital, A.; Bayarri, G.; Orozco, M. A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level. *Nucleic Acids Res.* **2020**, *48*, e29−e29.
(6) Aggarwal, A.; Naskar, S.; Sahoo, A. K.; Mogurampelly, S.; Garai, A.; Maiti, P. K. What do we know about DNA mechanics so far? *Curr. Op. Struct. Biol.* **2020**, *64*, 42−50.
(7) Dohnalová, H.; Lankaš, F. Deciphering the mechanical properties of B-DNA duplex. *WIREs Comput. Mol. Sci.* **2022**, *12* (3), e1575.
(8) Marin-Gonzalez, A.; Vilhena, J.; Perez, R.; Moreno-Herrero, F. A molecular view of DNA flexibility. *Q. Rev. Biophys.* **2021**, *54*, e8.
(9) Ivani, I.; Dans, P.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J.; González, C.; Vendruscolo, M.; Laughton, C.; Harris, S.; Case, D.; Orozco, M. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55−58.
(10) Curuksu, J.; Zacharias, M.; Lavery, R.; Zakrzewska, K. Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res.* **2009**, *37*, 3766−3773.
(11) Marin-Gonzalez, A.; Vilhena, J.; Perez, R.; Moreno-Herrero, F. Understanding the mechanical response of double-stranded DNA and RNA under constant stretching forces using all-atom molecular dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7049−7054.
(12) Shepherd, J. W.; Leake, M. C. *Chromosome Architecture*; Springer: New York, 2022; pp 249−262.
(13) Spiriti, J.; Kamberaj, H.; De Graff, A. M.; Thorpe, M.; van der Vaart, A. DNA bending through large angles is aided by ionic screening. *J. Chem. Theory Comput.* **2012**, *8*, 2145−2156.
(14) Karolak, A.; van der Vaart, A. Enhanced sampling simulations of DNA step parameters. *J. Comput. Chem.* **2014**, *35*, 2297−2304.
(15) Peguero-Tejada, A.; van der Vaart, A. Biasing simulations of DNA base pair parameters with application to propellor twisting in AT/AT, AA/TT, and AC/GT steps and their uracil analogs. *J. Chem. Inf. Model.* **2017**, *57*, 85−92.
(16) Lavery, R.; Moakher, M.; Maddocks, J.; Petkeviciute, D.; Zakrzewska, D. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917−5929.
(17) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer physics communications* **2014**, *185*, 604−613.
(18) The PLUMED consortium.. The PLUMED consortium, Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **2019**, *16*, 670−673.
(19) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **2001**, *313*, 229−237.
(20) Li, S.; Olson, W.; Lu, X.-J. wed 3DNA 2.0 for the analysis, visualisation, and modelling of 3d nucleic acid structures. *Nucleic Acids Res.* **2019**, *47*, W26−W34.
(21) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.; Hess, B.; Lindahl, E. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.
(22) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophysical journal* **2007**, *92*, 3817−3829.
(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(24) Dickerson, R. Definitions and nomenclature of nucleic acid structure components. *Nucleic acids research* **1989**, *17*, 1797−1803.
(25) Drsata, T.; Pérez, A.; Orozco, M.; Morozov, A. V.; Sponer, J.; Lankas, F. Structure, stiffness and substates of the Dickerson-Drew dodecamer. *J. Chem. Theor. Comp.* **2013**, *9*, 707−721.
(26) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(27) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(28) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(29) Kaestner, J. Umbrella sampling. *Wiley interdisciplinary reviews. Computational molecular science* **2011**, *1*, 932−942.
(30) Marko, J.; Siggia, E. Bending and twisting elasticity of DNA. *Macromolecules* **1994**, *27*, 981−988.
(31) Skoruppa, E.; Nomidis, S.; Marko, J. F.; Carlon, E. Bend-Induced Twist Waves and the Structure of Nucleosomal DNA. *Phys. Rev. Lett.* **2018**, *121*, 088101.
(32) Caraglio, M.; Skoruppa, E.; Carlon, E. Overtwisting induces polygonal shapes in bent DNA. *J. Chem. Phys.* **2019**, *150*, 135101.
(33) Nomidis, S. K.; Caraglio, M.; Laleman, M.; Phillips, K.; Skoruppa, E.; Carlon, E. Twist-bend coupling, twist waves, and the shape of DNA loops. *Phys. Rev. E* **2019**, *100*, 022402.
(34) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein−DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163−11168.
(35) Wiggins, P. A.; van der Heijden, T.; Moreno-herrero, F.; Spakowitz, A.; Phillips, R.; Widom, J.; Dekker, C.; Nelson, P. C. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.* **2006**, *1*, 137−141.
(36) Mazur, A. K.; Maaloum, M. Atomic force microscopy study of DNA flexibility on short length scales: smooth bending versus kinking. *Nucleic Acids Res.* **2014**, *42*, 14006−14012.
(37) Eslami-Mossallam, B.; Ejtehadi, M. Contribution of nonlocal interactions to DNA elasticity. *J. Chem. Phys.* **2011**, *134*, 125106.
(38) Skoruppa, E.; Voorspoels, A.; Vreede, J.; Carlon, E. Length-scale-dependent elasticity in DNA from coarse-grained and all-atom models. *Phys. Rev. E* **2021**, *103*, 042408.
(39) Segers, M.; Voorspoels, A.; Sakaue, T.; Carlon, E. Mechanical properties of nucleic acids and the non-local twistable wormlike chain model. *J. Chem. Phys.* **2022**, *156*, 234105.