

Accurate Sequence-Dependent Coarse-Grained Model for Conformational and Elastic Properties of Double-Stranded DNA

Salvatore Assenza and Rubén Pérez*



Cite This: *J. Chem. Theory Comput.* 2022, 18, 3239–3256



Read Online

ACCESS |



Metrics & More

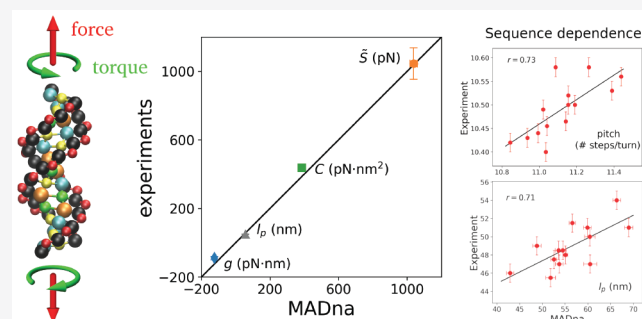


Article Recommendations



Supporting Information

ABSTRACT: We introduce MADna, a sequence-dependent coarse-grained model of double-stranded DNA (dsDNA), where each nucleotide is described by three beads localized at the sugar, at the base moiety, and at the phosphate group, respectively. The sequence dependence is included by considering a step-dependent parametrization of the bonded interactions, which are tuned in order to reproduce the values of key observables obtained from exhaustive atomistic simulations from the literature. The predictions of the model are benchmarked against an independent set of all-atom simulations, showing that it captures with high fidelity the sequence dependence of conformational and elastic features beyond the single step considered in its formulation. A remarkably good agreement with experiments is found for both sequence-averaged and sequence-dependent conformational and elastic features, including the stretching and torsion moduli, the twist–stretch and twist–bend couplings, the persistence length, and the helical pitch. Overall, for the inspected quantities, the model has a precision comparable to atomistic simulations, hence providing a reliable coarse-grained description for the rationalization of single-molecule experiments and the study of cellular processes involving dsDNA. Owing to the simplicity of its formulation, MADna can be straightforwardly included in common simulation engines. Particularly, an implementation of the model in LAMMPS is made available on an online repository to ease its usage within the DNA research community.



INTRODUCTION

Sequence-dependent conformational and elastic properties of DNA are of the utmost importance for its regulation *in vivo*, as they directly affect DNA–protein interactions.¹ The detailed shape of a DNA fragment and its deformability are indeed key determinants of protein recognition and binding.^{1,2} For instance, the unique conformational properties of A-tracts are known to affect nucleosomal organization,³ as well as DNA replication and recombination.⁴ Analogously, TATA-box elements present a strong deviation from the canonical B-DNA conformation, which is exploited to enhance binding by the TATA-box binding protein.¹ Moreover, proteins continuously exert mechanical stress on bound DNA. For example, torsion is applied on DNA by topoisomerases and polymerases^{5,6} and plays a central role in chromatin remodeling.⁷ Also, DNA stretching is relevant *in vivo*, e.g., in the action of recombinases⁸ or for site recognition in nucleosomes.⁹

Together with the practical implications of DNA elasticity in nanotechnological applications,^{10,11} this fundamental interest has prompted a conspicuous amount of experimental efforts devoted to the detailed characterization of DNA conformational ensemble and elasticity. For the study of conformers, classical crystallographic and NMR studies are nowadays being complemented by X-ray interferometry^{12,13} and by single-

molecule imaging with atomic force microscopy (AFM)^{14,15} or cryoelectron microscopy.¹⁶ Single-molecule techniques like AFM, optical tweezers, and magnetic tweezers are employed to assess the elastic properties of long DNA molecules, providing stiffness values for various mechanical perturbation modes such as stretching,^{17–21} twisting,^{20,22–27} bending,^{17–19,21,28–30} and the coupling between them.^{20,31–35} At short length scales, the interpretation of experiments becomes cumbersome due to both theoretical and experimental challenges. Indeed, direct AFM imaging has provided conflicting results on the bendability of short DNA fragments.^{36,37} For other experimental approaches such as cyclization assays,^{38–41} the elastic parameters can be obtained only by interpreting the data within specific theoretical frameworks, and conflicting conclusions have been reported based on different physical assumptions.^{12,42–44}

Received: February 10, 2022

Published: April 8, 2022



Molecular simulations provide a valuable means to complement experimental studies, particularly in view of the discrepancies mentioned above. All-atom molecular dynamics has been successfully employed to capture DNA elastic and conformational features^{45–49} as well as their dependence on sequence.^{35,50–53}

An evident limit of atomistic simulations originates from the associated high computational cost, which puts severe boundaries on the length and time scales that can be simulated. The need to overcome this barrier has fostered the development of coarse-grained approaches, where one selects only the degrees of freedom relevant to the problem at hand. An elegant solution is implemented in the software packages cgDNA⁵⁴ and MC-eNN,⁵⁵ where the nucleotides are represented as rigid frames and the energy of the system is described by means of a stiffness matrix. This approach is extremely efficient from a computational perspective and enables assessing the conformational ensemble of long DNA chains.⁵⁶ However, an important drawback lies in the absence of an explicit description of the backbone and the related difficulty in interfacing this representation with external perturbations present in most systems of interest, e.g., mechanical stress, confinement, or a binding protein. A good compromise between computational efficiency and modeling flexibility is achieved by considering coarse-grained models in which effective particles represent the various moieties along the DNA molecule.^{57–65} These models have been employed to describe a wide palette of systems involving DNA, including, e.g., DNA origami,⁶⁶ nucleosomes,⁶⁷ and cyclization assays.⁴⁴

Despite this success, the coarse-grained models of the latter kind present in the literature do not satisfactorily capture all the main elastic features of double-stranded DNA at once (see, e.g., refs 62 and 68). In this work, we fill this gap by introducing the mechanically-accurate DNA (MADna) model, a sequence-dependent coarse-grained model whose parameters are tuned to reproduce local conformational features of double-stranded DNA and the stiffness of the various mechanical perturbation modes obtained via all-atom molecular dynamics.⁵² We show that the model satisfactorily reproduces the atomistic values also for DNA fragments different than the ones employed for fitting, as well as experimental data from the literature. Particularly, and at variance with existing models, our coarse-grained description captures the negative twist–stretch coupling recently unveiled by single-molecule force spectroscopy^{31,32} as well as the experimentally determined sequence dependence of dsDNA helical pitch and persistence length.²⁸

The full details of MADna and the definitions of the various physical and geometrical quantities are provided in [Methods](#). Nevertheless, they are shortly introduced also in [Results and Discussion](#), making this section somewhat self-standing. In this way, the main findings of this work are presented succinctly, leaving the interested reader to look up the technical details within the [Methods](#) section.

Finally, we remark that an online repository is available (<https://github.com/saassenza/MADna/tree/main/LAMMPS>) where MADna has been implemented in LAMMPS⁶⁹ without the need to introduce additional custom packages. The repository also contains a molecular builder which automatically creates the initial coordinates and topology for simulation in LAMMPS.

METHODS

Coarse-Grained Model. MADna describes a molecule of double-stranded DNA (dsDNA) by considering three effective particles per nucleotide, located in the geometric centers of the phosphate group, the sugar, and the base, as proposed in the past in the 3SPN⁵⁷ and TIS⁶⁵ models. In [Figure 1](#), two

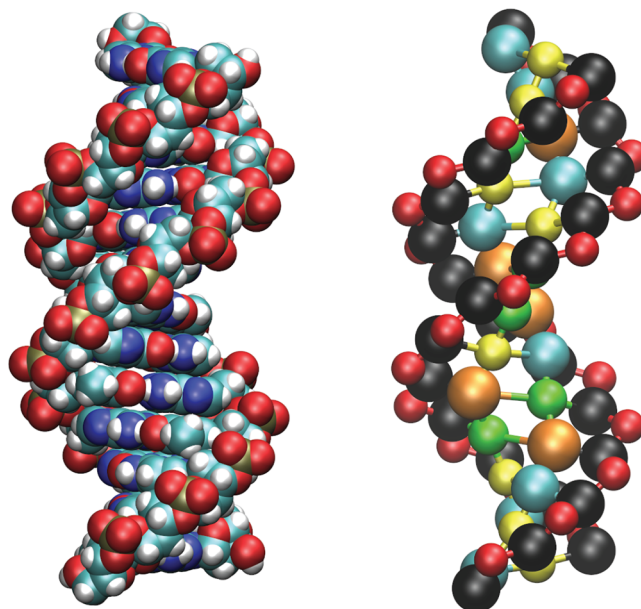


Figure 1. Atomistic (left) and coarse-grained (right) description for a representative dsDNA molecule with leading sequence 5'-CGCTACTTCGAGG-3' in the B-DNA form as obtained by employing the NAB software.⁷⁰ In the coarse-grained cartoon, the color code is the following: sugar ↔ black; phosphate group ↔ red; adenine ↔ green; cytosine ↔ cyan; guanine ↔ yellow; thymine ↔ orange. The size of each bead is proportional to the WCA radius of the corresponding moiety.

cartoons are reported to compare the atomistic and coarse-grained description of a representative sequence. Further details on the coarse-graining procedure are reported in section S1 in the [Supporting Information](#). In spite of the similar level of coarse-graining, MADna diverges from the other models in terms of the choice of bonded interactions both in the way they are built and in their parametrization. Particularly, our choice of bonded interactions enables capturing the structural heterogeneity predicted by atomistic simulations, as will be showed below when we comment on the results in [Figure 2](#) and [Figure 5i](#). On the other hand, we chose to fix the double-stranded structure by means of bonded interactions, which simplifies the implementation of the model (for instance, MADna does not need modulation factors to account for anisotropic interactions) but makes it less general than the other ones as it cannot account for DNA thermodynamics. In this regard, MADna is more akin to the MARTINI representation of double-stranded DNA, where the double helix is maintained through a suitably defined elastic network.⁶⁴ We will add the possibility of strand separation in future versions of MADna. As for the parametrization, MADna extracts information from a large data set provided by atomistic simulations, while both TIS and 3SPN are parametrized by means of experiments. Experimental results are of course the ultimate source of information, but we decided to tune the parameters of MADna from atomistic simulations based on

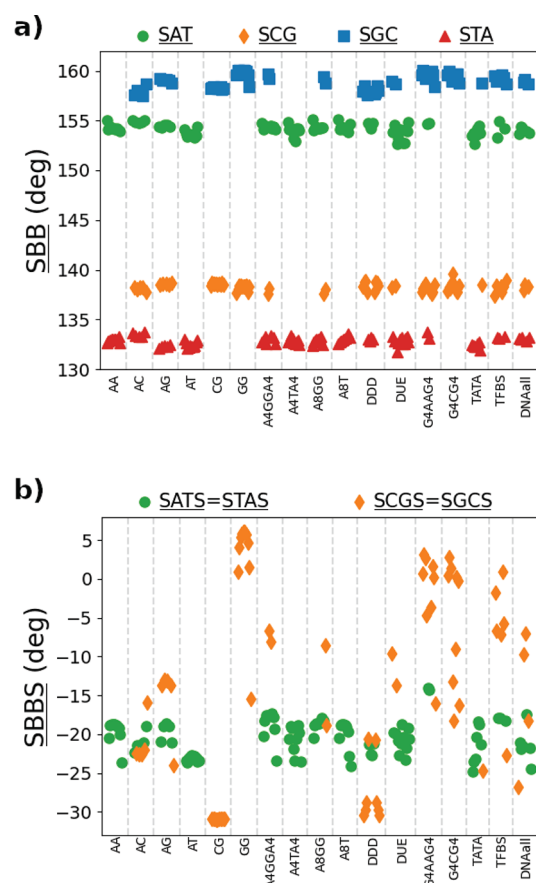


Figure 2. Average values obtained by coarse-graining the atomistic simulations for the angles \overline{SBB} (a) and the dihedrals \overline{SBBs} (b). Vertical lines separate the distinct sequences, which are listed in Section S3.1 in the [Supporting Information](#). In the legends, the various labels correspond to the particular bases involved in the local conformation under consideration. For instance, \overline{SAT} considers a \overline{SBB} angle in which an adenine is bound to the sugar. Note also that the dihedrals are symmetric with respect to the inversion of the involved bases, as it just corresponds to changing the arbitrary reference strand.

both the far richer amount of available information and their demonstrated accuracy in capturing local conformational and elastic features of dsDNA. As we show in the [Results and Discussion section](#), this choice is rewarded with a very good performance of MADna in capturing many distinct experimental observations on the structure and mechanical response of dsDNA.

Bonded Interactions. The double-stranded topology is fixed once and for all and is maintained by introducing several two-, three-, and four-body bonded interactions (from here on, we will refer to them as bonds, angles, and dihedrals, respectively), which connect beads within the same strand as well as providing interstrand links.

Before going into the details of the bonded interactions considered, we first proceed to clarify the nomenclature used in this section. In order to label the bonded interactions, we indicate the sugar, phosphate, and base beads as S, P, and B, respectively. In this way, a bond between, e.g., a sugar and a base is indicated as SB. Whenever the bonded interaction runs along the 5'-3' direction of a strand, we add a corresponding tag at the two ends of the label. For instance, a bond between a sugar and a phosphate in the 5'-3' direction is indicated as 5'-SP-3'. Analogously, we denote as 5'-SPSB-3' a dihedral

involving a sugar, a phosphate, the following sugar, and the base attached to it (ordered according to the 5'-3' direction of the strand). In order to keep in place the double-stranded structure, some bonded interactions involve both strands. Such bonded interactions are marked by underlining their label. For instance, the bonds accounting for Watson–Crick base pairing are indicated as $\overline{BB-WC}$ (in this specific case we also added “WC” in order to clearly distinguish these bonds from 5'-BB-3', which account for stacking interactions). As another example, we denote as 5'- \overline{PSBB} -3' a dihedral involving in the 5'-3' direction a phosphate, a sugar, the attached base, and the base paired to it. We further note that, due to the asymmetry of the 5'-3' direction, the order in which the same beads are considered is important. For instance, the bonds 5'-SP-3' and 5'-PS-3' are different from each other. This can be clearly seen by comparing the corresponding entries in Table S2 in the [Supporting Information](#), where we show the equilibrium values computed by coarse-graining the atomistic simulations used to parametrize MADna (see below). Analogously, the dihedrals 5'-SPSP-3' and 5'-PSPS-3' are different from each other, as can be seen from Table S4 in the [Supporting Information](#).

For the bonds potential U_{bond} , a harmonic function was considered:

$$U_{\text{bond}}(r) = k_{\text{bond}}(r - r_0)^2 \quad (1)$$

where r is the distance separating the two beads connected by the bond, k_{bond} is the elastic constant specific to the type of bond considered, and r_0 is the corresponding equilibrium distance. An analogous formula was used for the angle potential U_{angle} :

$$U_{\text{angle}}(\theta) = k_{\text{angle}}(\theta - \theta_0)^2 \quad (2)$$

where θ is the angle characterizing the three-body bonded interaction, k_{angle} is the bending constant relative to the type of angle considered, and θ_0 is the equilibrium value of θ . Finally, the following formula was chosen for the dihedral potential U_{dihedral} :

$$U_{\text{dihedral}}(\phi) = k_{\text{dihedral}}[1 + \cos(\phi - \phi_0)] \quad (3)$$

where ϕ is the dihedral angle, k_{dihedral} is the elastic constant of the particular type of dihedral considered, and $\phi_0 = \phi_{\text{min}} - 180^\circ$, with ϕ_{min} being the equilibrium dihedral angle. This choice for the bonded interactions relies on the observation of normal distributions for most of the corresponding observables (see Section S2.2 in the [Supporting Information](#)).

The general guideline in choosing how to build the topology was to obtain a minimal set of bonded interactions which reproduces the structural features obtained in all-atom simulations. With this spirit, we considered only bonded interactions involving up to a single step and implemented the sequence dependence by setting different values of the parameters according to the particular step under consideration. For instance, if along the sequence we have a CT step, the bond 5'-BB-3' accounting for the stacking interaction is implemented by means of eq 1, where k_{bond} and r_0 are equal to the values assigned to a CT step. Analogously, the dihedral 5'- \overline{PSBB} -3' is implemented by means of eq 3, where k_{dihedral} and ϕ_0 are set to the values corresponding to CT. The same is repeated for all the step-dependent bonded interactions considered, so that 16 different values are considered for the corresponding parameters depending on the involved step. Some local features are symmetric with respect to the 5'-3'

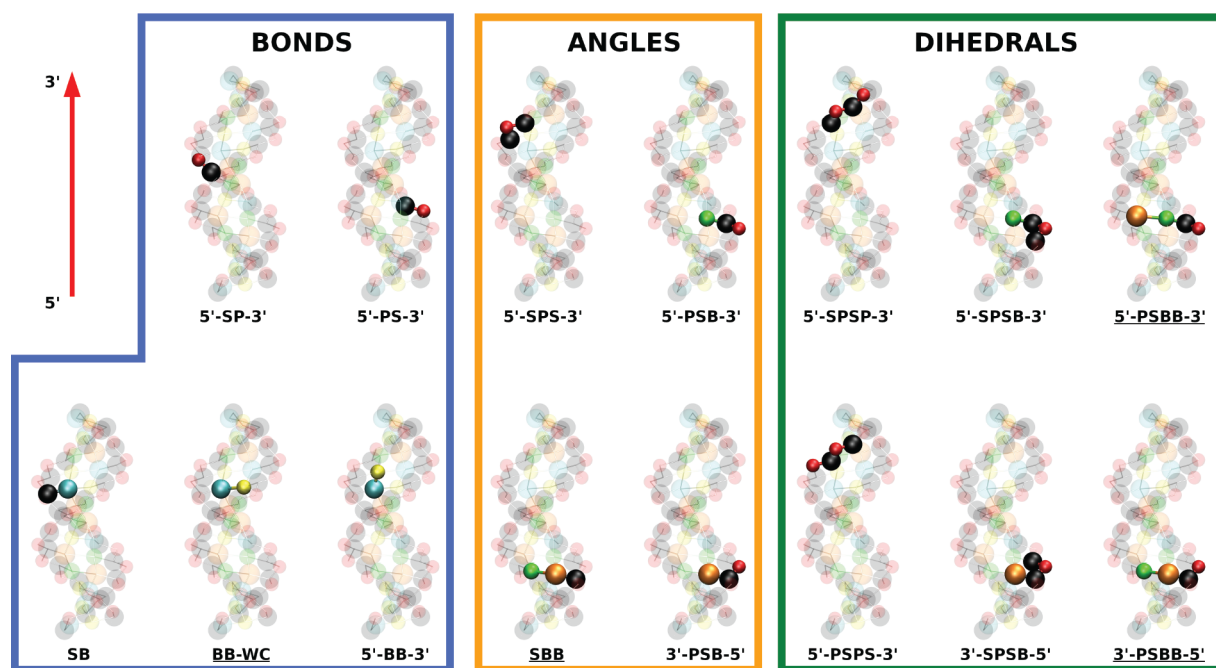


Figure 3. List of the various bonded interactions considered in the model, together with representative examples based on the same molecule as in Figure 1. Step-dependent bonded interactions are indicated by the presence of the tags 5' and 3' in their label. For interstrand interactions, the corresponding label is underlined. The letters present in the labels indicate sugars (S), phosphate groups (P), or generic bases (B). For clarity, all the selected examples involve beads belonging to the same strand, whose 5'-3' direction is indicated by the arrow in the top-left panel.

direction. The corresponding bonded interactions were then implemented as being dependent only on the corresponding base pair. For instance, for the bond SB we considered four possibilities (A, C, G, and T), while for the bond BB-WC there were just two choices (AT or CG). Yet, this assumes that these local features do not depend on the neighboring steps, which is not always the case. As an example, we report in Figure 2 the averages obtained by coarse-graining the atomistic simulations for the angles SBB (Figure 2a) and the dihedrals SBBS (Figure 2b). The angles SBB are obtained by considering a sugar, the base attached to it, and the corresponding Watson–Crick partner. For instance, we indicate as SAT the angle formed by a sugar, the adenine attached to it, and the thymine paired with the adenine. As for the dihedrals SBBS, we similarly consider a Watson–Crick base pair and the corresponding sugars. In Figure 2, the points are color-coded according to a classification based on a single base pair. Hence, four possible choices are present for SBB, depending on the base being attached to the sugar (A, C, G, T), while only two choices are available for SBBS, corresponding to the Watson–Crick base pairs AT and CG. As shown in Figure 2a, the data for SBB nicely cluster around their average values, indicating that this angle is in practice independent of the rest of the sequence of the hosting DNA molecule and that it can thus be considered for bonded interactions depending only on the base pair. In contrast, the dihedrals SBBS show a wide variability, particularly in the case of CG base pairs (Figure 2b). This demonstrates that these dihedrals depend on their neighborhood and are thus not apt for implementing bonded interactions depending only on the base pair. As we show in the Results and Discussion, the heterogeneity of the SBBS dihedrals is quantitatively captured by MADna as an emergent property resulting from the interplay between the other bonded interactions. We also note that this is a distinguishing feature of MADna when compared to the TIS⁶⁵ and 3SPN2⁷¹

models, where instead the base-pair-dependent SBBS interactions are introduced as part of the model, thus preventing the possibility of capturing the heterogeneity suggested by the atomistic simulations.

The choice of the bonded interactions was obtained in practice by trial and error: starting from a far too simplistic description, where the beads were barely kept together by a minimal set of bonds, we incrementally added bonded interactions. At each iteration, we ran simulations with the coarse-grained model and computed the averages obtained for local conformational features that were not yet directly implemented. We then compared these averages with the results obtained by coarse-graining the atomistic trajectories. If the comparison was not good, we then proceeded to add other bonded interactions. The iterative procedure stopped when we reached a satisfactory comparison. We also considered an opposite approach, where we started by considering far too many bonded interactions and deleted some of them until further eliminations resulted in disrupting the structure. The process has some arbitrariness in the properties being monitored, the order in which we add bonded interaction, etc. Yet, the high fidelity of the coarse-grained model in reproducing the atomistic values (as shown in the Results and Discussion) validates the final choice of bonded parameters that we considered. The final set of employed bonds, angles, and dihedrals is reported in Figure 3, together with some representative examples based on the dsDNA molecule sketched in Figure 1. Some of these bonded interactions are directly related to physically meaningful features, such as hydrogen bonds between Watson–Crick pairs (labeled as BB-WC in Figure 3) and stacking interactions (S'-BB-3' in Figure 3). Summing up, there are in total 202 distinct bonded interactions in the model. Further details are reported in Section S2 in the Supporting Information.

Importantly, since all the bonded interactions involve at most two consecutive base pairs, any feature at scales larger than a single step is an emergent property originating from the propagation of the local interactions in combination with the electrostatic repulsion between phosphates (see below).

Excluded Volume. Excluded-volume interactions were implemented by means of a Weeks–Chandler–Andersen (WCA) potential, i.e., by retaining the repulsive part of a Lennard–Jones interaction:

$$U_{\text{excl}}(r) = \begin{cases} \epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - 2 \left(\frac{\sigma}{r} \right)^6 \right] + \epsilon & \text{if } r \leq \sigma \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In the previous formula, r is the distance between the two particles, $\epsilon = 1$ kcal/mol, while σ depends on the bead considered, as reported in Table S1 in the [Supporting Information](#).

Electrostatics. In order to account for the presence of charges along the backbone, a charge $q \cdot e_0$ was assigned to the beads corresponding to phosphate groups, with $e_0 = 1.6 \times 10^{-19}$ C being the elementary charge. Although each phosphate carries a unit negative charge, in order to account for counterions condensation an effective reduced value $q = -0.6$ was considered.^{65,71,72} The salt-induced electrostatic screening was modeled via a Debye–Hückel interaction:⁷³

$$U_{\text{el}} = \frac{q^2 e_0^2}{4\pi \epsilon_0 \epsilon_r r} e^{-r/l_D} \quad (5)$$

In the previous formula, $\epsilon_0 = 8.859 \times 10^{-12}$ F/m is the absolute permittivity; $\epsilon_r = 78.3$ is the relative permittivity of water; and l_D is the Debye length, defined as

$$l_D = \sqrt{\frac{k_B T \epsilon_0 \epsilon_r}{2 N_A e_0^2 I}} \quad (6)$$

where $k_B = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant; T is the temperature of the system in Kelvin; $N_A = 6.022 \times 10^{23}$ mol⁻¹ is Avogadro's constant; and I is the ionic strength of the solution in mM.

Determination of Parameters. The parameters of the various sequence-dependent bonded interactions were tuned in order to reproduce the results of atomistic simulations from ref 52, which were performed in AMBER14⁷⁴ with the parm99 force field⁷⁵ including the bsc0 modifications.⁷⁶ In those simulations, dsDNA molecules with different sequences were pulled under the action of forces ranging from 1 to 20 pN. The main details of the atomistic simulations are reported in Section S2.1 in the [Supporting Information](#). Here, we used a subset of the atomistic simulations covering all possible base steps (we refer to this subset as “training sequences”, see Section S3.1 in the [Supporting Information](#) for the sequences) to determine the bonded parameters, while the rest of sequences were employed as test cases to benchmark the coarse-grained model (“testing sequences”, also reported in Section S3.1 in the [Supporting Information](#)). It has to be noted that, although parm99+bsc0 simulations reproduce well most structural features determined in experiments,⁷⁷ more precise modifications have been recently introduced, namely bsc1⁷⁸ and OL15.⁷⁹ Our reason to choose bsc0 was rooted in the convenience of having the atomistic data already available in our group,^{46,52} together with the reasonable precision of these

modifications.⁷⁷ Future developments of MADna will include the reparameterization of the model via atomistic simulations based on bsc1 or OL15.

A first estimation of the parameters was obtained via Boltzmann inversion of the all-atom simulations performed at 1 pN. In this regard, the atomistic trajectories were first coarse-grained according to the three-beads representation. Then, for each bond, angle, and dihedral the ensemble averages were computed from the coarse-grained trajectories, which enabled fixing the values of r_0 , θ_0 , and ϕ_0 . Analogously, the elastic constants were determined in order to reproduce the size of fluctuations starting from eqs 1, 2, and 3. Further details on this procedure can be found in Section S2.2 in the [Supporting Information](#).

Coarse-grained simulations performed using the obtained force field showed that the set of parameters obtained by Boltzmann inversion provides reasonable values for the elastic constants (see Figure S1 in the [Supporting Information](#)). Nevertheless, some of the parameters were further tuned in order to improve the quantitative agreement with the atomistic simulations, focusing on reproducing the elastic constants of the training sequences (see [Results and Discussion](#) and Section S2.2 in the [Supporting Information](#)).

Molecular Dynamics Simulations. All the simulations were performed in LAMMPS (<http://lammps.sandia.gov>⁶⁹). The temperature $T = 300$ K was maintained through a Langevin thermostat with damping constant $\tau_{\text{damp}} = 20$ ps. The integration step was set to 20 fs.

Benchmark Simulations. In order to benchmark the coarse-grained model, we performed pulling simulations of the training and testing sequences following the same protocol as in refs 46 and 52. A pulling force \mathbf{f} was applied to the center of mass ξ_2 of the sugars belonging to the second base pair. Analogously, an opposite force $-\mathbf{f}$ was applied to the center of mass $\xi_{l_{\text{seq}}-1}$ of the second-to-last base pair (l_{seq} is the length of the sequence). The direction of the forces was taken along the line connecting ξ_2 and $\xi_{l_{\text{seq}}-1}$. Simulations were performed for $f = 1, 5, 10, 15,$ and 20 pN. For each simulation, the system was initialized by considering the average structure created by the molecular builder (defined below), and the equations of motion were integrated for 20 ns. In order to ensure full equilibration, only the last 10 ns were considered for analysis. The convergence of a representative simulation is reported in Figure S8 in the [Supporting Information](#). For each sequence and force, 100 independent simulations were performed. The ionic strength I was set at the same value as in the atomistic simulations from refs 46 and 52 and was computed as $I = N_{\text{ions}}/N_A V$, where N_{ions} is the number of counterions and V is the volume of the simulation box.

Persistence-Length Simulations. For the computation of the sequence-averaged persistence length, we considered 20 random sequences made of 100 base pairs. The sequences are listed in Section S3.2 in the [Supporting Information](#). For each realization, the molecule was initialized by considering the average structure obtained by the molecular builder (see below), and the simulation was run for 500 ns. In order to ensure full equilibration, the first 100 ns were discarded from the analysis. The convergence of a representative trajectory is reported in Figure S9 in the [Supporting Information](#). For each sequence, 10 independent simulations were performed. The ionic strength was set at $I = 150$ mM.

In a second set of simulations, we studied the sequence dependence of persistence length and helical pitch and compared our results with the experiments from ref 28. We considered 14 sequences of 100 base pairs obtained as the central fragments of the corresponding experimental ones.²⁸ The sequences are listed in Section S3.2 in the [Supporting Information](#). The same protocol as for the sequence-averaged study was applied. The ionic strength was set at $I = 1000$ mM to ease the comparison with the predictions of CGDNA.⁵⁶ The convergence of a representative trajectory is reported in Figure S10 in the [Supporting Information](#). Simulations were performed with MADna and, for comparison, with oxDNA2⁸⁰ (using the LAMMPS implementation⁸¹ and considering sequence-dependent stacking interactions) and the sequence-dependent model 3SPN2C.⁸²

Stretch–Torsion Simulations. In order to determine the elastic constants, a simulation setup was considered where a 40bp dsDNA molecule was subjected to the simultaneous action of a pulling force f and a torque τ directed along the z axis and with constant magnitude.

For each simulation, the molecule was initialized via the molecular builder and was aligned with the z axis. The two bottom base pairs and the corresponding sugars and phosphates were tethered to their initial position via harmonic constraints with constant $100 \text{ kcal/mol } \text{\AA}^2$. A constant pulling force f oriented along the positive z direction was exerted on the center of the sugars belonging to the second base pair counting from the top. A harmonic constraint with constant $100 \text{ kcal/mol } \text{\AA}^2$ was also applied to the x and y coordinates of the center to keep the molecule aligned with the z axis and avoid spurious effects originating from the overall orientational entropy.⁸³ Finally, a constant torque τ was imposed on the two top base pairs and the corresponding sugars and phosphates.

Simulations were performed at $\tau = 0$ for values of the force $f = 2, 4, 6, 8, 10, 15, 20, 25, 30, 35,$ and 40 pN and at $f = 2$ pN for values of the torque $\tau = 0, 5, 10, 15, 20, 25,$ and 30 pN·nm. To account for sequence-induced heterogeneity, five random sequences were generated, which we refer to as sequences ST1, ..., ST5 (the sequences are listed in Section S3.3 in the [Supporting Information](#)). Moreover, to check the dependence of the elastic constants on the length of the dsDNA molecule, we also performed simulations on shorter sequences containing 20 base pairs, which we refer to as sequences ST1-short, ..., ST5-short (the sequences are listed in Section S3.3 in the [Supporting Information](#)). For each combination of sequence, force, and torque, 10 independent simulations of length 250 ns were run. The first 50 ns were not considered for analysis in order to let the system equilibrate. The convergence of a representative trajectory is reported in Figure S11 in the [Supporting Information](#). Simulations were performed with MADna, oxDNA2, and 3SPN2C. In all cases, the ionic strength was set at 150 mM. The four base pairs at each end were discarded from the analysis.

Finally, in order to study the effect of sequence on the elasticity, we considered another set of simulations with phased A-tracts, which are listed in Section S3.3 in the [Supporting Information](#). The same protocol as for the rest of the simulations was applied.

Twist–Bend Coupling Simulations. In order to determine the softer response to torque due to the coupling between twist and bending, we performed simulations similar to those of the previous section. Here, 150bp long sequences were placed under the action of a pulling force directed along the z axis and

acting on one end of the molecule. The other end was tethered with the same protocol as in the previous section. Note that the pulled end is here free to move in the x and y directions in order to let the molecule bend.

Simulations were performed for values of the force f in steps of 0.05 pN within the range $0.05 \text{ pN} \leq f \leq 0.50 \text{ pN}$ and in steps of 0.25 pN within the range $0.50 \text{ pN} \leq f \leq 2.50 \text{ pN}$. Three different sequences were considered, which we refer to as TB1, TB2, and TB3 (the sequences are listed in Section S3.4 in the [Supporting Information](#)). For each combination of sequence and force, three independent simulations of length 2000 ns were run. The first 1000 ns were not considered for analysis in order to let the system equilibrate. The convergence of a representative trajectory is reported in Figure S12 in the [Supporting Information](#). Simulations were performed with MADna, 3SPN2C, and oxDNA2.

Definitions and Protocols. Basic Definitions. We will indicate as $\mathbf{P}_{s,i}$, $\mathbf{S}_{s,i}$, and $\mathbf{B}_{s,i}$ the position vectors of the beads corresponding to the i th phosphate group, sugar, and base, respectively, and located on strand $s = 1, 2$. The index i is counted in the 5′–3′ direction of the arbitrarily chosen strand 1. Note that $i = 1, \dots, n$ for sugar and base beads, while $i = 2, \dots, n$ for the phosphate groups, where n is the total number of base pairs (compare Figure 1). Hence, there are in total $N = 2[(n - 1) + n + n] = 6n - 2$ beads, where the factor of 2 is included to account for the presence of the two strands. Handles were always discarded from the analysis, so the various indexes introduced here refer to the subfragment under consideration.

Crookedness. The crookedness β accounts for the deviation of a dsDNA molecule from a straight line⁵² (Figure 4b). For short fragments (such as the training and testing sequences considered here), thermally induced bending is negligible; hence, in this case β effectively quantifies the spontaneous bending of the molecule. The definition of β is here modified with respect to its original formulation to adapt it to the coarse-grained context. The contour length L of the line connecting the base-pair centers along the molecule is $L = \sum_{i=1}^{n-1} |\mathbf{\Gamma}_{i+1} - \mathbf{\Gamma}_i|$, where $\mathbf{\Gamma}_i \equiv (\mathbf{B}_{1,i} + \mathbf{B}_{2,i})/2$. The crookedness β is defined as

$$\beta = \arccos\left(\frac{|\mathbf{\Gamma}_n - \mathbf{\Gamma}_1|}{L}\right) \quad (7)$$

If the line connecting the centers is perfectly straight, then $\beta = 0$. The more curved the line, the larger the corresponding β . As a reference, we computed the crookedness for the conformations obtained via NAB⁷⁰ for the training and testing sequences, obtaining an average value $\beta \approx 0.17$.

Helical Parameters. Due to the limited information on local coordinates, the definition of the helical parameters has to be adapted to the coarse-grained framework. Particularly, the representation of the bases as single beads prevents defining the local reference frames, which are crucial for the standard definitions of helical parameters.^{84,85} Taking advantage of the short length of the training and testing sequences, the helical axis $\hat{\mathbf{h}}$ in this case was defined globally as the axis of the cylinder best fitting the position of phosphates in the double helix (Figure 4a; see Section S4.1 in the [Supporting Information](#) for the technical details), which was also employed to estimate the diameter of the DNA. In order to define the h-twist, we first introduce the sugar separation vector $\hat{\boldsymbol{\zeta}}_i \equiv \mathbf{S}_{2,i} - \mathbf{S}_{1,i}$ (Figure 4d). The h-twist for the step $i, i + 1$ was defined by $\cos(\text{h-twist}_{i,i+1}) \equiv \hat{\boldsymbol{\zeta}}_i \cdot \hat{\boldsymbol{\zeta}}_{i+1}$, where $\hat{\boldsymbol{\zeta}}_i \equiv \boldsymbol{\zeta}_i / |\boldsymbol{\zeta}_i|$ and

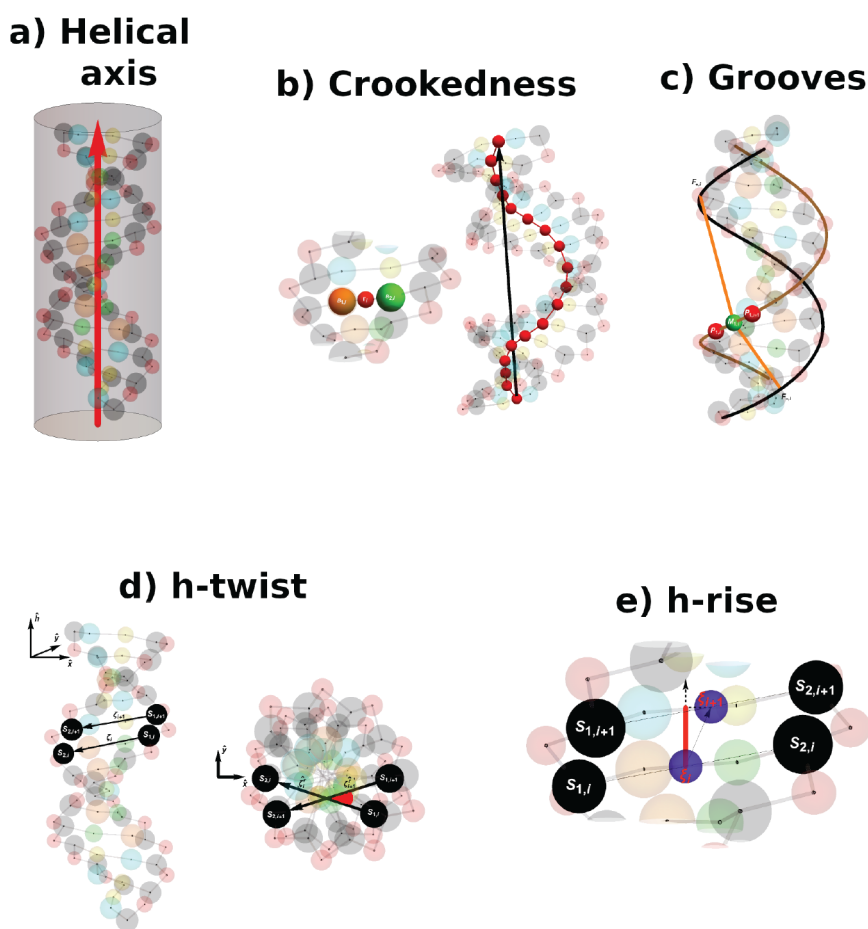


Figure 4. Sketches showing pictorially the definitions to characterize the geometry of DNA. (a) The helical axis is obtained as the axis of the best-fitting cylinder. (b) The crookedness is obtained by computing the arccosine of the ratio between the end-to-end distance (black arrow) and the contour-length of DNA (red line). These quantities are obtained along the line formed by the points Γ_i , which are determined as the centers between bases belonging to the same pair (inset). (c) Grooves are defined by considering the lines interpolating the phosphate beads (brown and black lines). For any couple of phosphates on the first strand ($P_{1,i}$ and $P_{1,i+1}$ in this case), we define the midpoint (M_i). From the midpoint, we find the closest points on the second strand. The groove widths are obtained as the corresponding minimum distances (orange segments), suitably shifted to account for the excluded volume of the backbone. (d) The h-twist is defined by considering the vectors $\xi_i \equiv S_{2,i} - S_{1,i}$ joining the two sugars within each base. The vectors ξ_i are projected onto the plane perpendicular to the helical axis, thus obtaining ξ'_i . The h-twist is then defined as the angle depicted in red, corresponding to $\cos \text{h-twist} = \xi'_i \cdot \xi'_{i+1} / |\xi'_i| |\xi'_{i+1}|$. The h-rise is defined by considering the geometrical centers of the sugars $\xi_i \equiv (S_{1,i} + S_{2,i})/2$ and projecting the vector separating two consecutive centers onto the helical axis, thus obtaining $\text{h-rise} = (\xi_{i+1} - \xi_i) \cdot \hat{h}$, corresponding to the red segment in the figure.

$\xi'_i \equiv \xi_i - (\xi_i \cdot \hat{h}) \hat{h}$ is the projection of ξ_i onto the plane perpendicular to the helical axis. Following the usual convention,⁸⁴ the sign of the h-twist was determined according to the sign of the product $(\xi'_i \times \xi'_{i+1}) \cdot \hat{h}$. The employment of the positions of the sugars in the definition of the h-twist was motivated by the high correlation found with the standard definition in 3DNA⁸⁴ (see Figure S2 in the [Supporting Information](#)). Finally, the h-rise for the step i , $i + 1$ was computed as $\text{h-rise}_{i,i+1} = (\xi_{i+1} - \xi_i) \cdot \hat{h}$, where $\xi_i \equiv (S_{1,i} + S_{2,i})/2$ is the center of the sugars of base pair i (Figure 4e).

Grooves. For a visual description of the definition of groove geometry, see Figure 4c. The quantification of the geometry of grooves was assessed in a similar way as in Curves+.⁸⁵ The positions of the phosphate groups in each strand were interpolated via centripetal Catmull-Rom splines.⁸⁶ Particularly, the helical fragment connecting $P_{s,i}$ and $P_{s,i+1}$ was interpolated by considering $P_{s,i-1}$, $P_{s,i}$, $P_{s,i+1}$, and $P_{s,i+2}$ as control points. For each of the terminal phosphates, the missing external control point was built by extrapolation of the h-rise and h-twist of the last step. In order to determine the groove

geometry corresponding to base pair i , we first considered the midpoint $M_{1,i}$ between $P_{1,i}$ and $P_{1,i+1}$ along the interpolated curve. Then, starting from the analogous midpoint $M_{2,i}$ on the second strand, we followed the corresponding interpolating curve in both directions while computing the distance from $M_{1,i}$ until a local minimum in the distance was first encountered at two points denoted as $F_{+,i}$ and $F_{-,i}$. From the two identified minimal distances, we subtracted 0.58 nm to account for the van der Waals radius of the backbone⁸⁵ and assigned the resulting values to the width of the major and minor groove. In order to compute the depth of each groove, we considered the midpoints on the width vectors, i.e., $0.5(M_{1,i} + F_{+,i})$ and $0.5(M_{1,i} + F_{-,i})$ and computed their distance from the base-pair center Γ_i . Again, in order to account for the van der Waals radii, we estimated the final groove depths by subtracting 0.35 nm from the calculated distance.⁸⁵

Determination of Elastic Constants. In order to interpret single-molecule experiments, dsDNA is often modeled as an elastic rod. When a pulling force f and a

torque τ are exerted along the direction of the rod, the elastic energy U reads^{22,31}

$$U = \frac{1}{2} \frac{S}{L_0} (\Delta L)^2 + \frac{1}{2} \frac{C}{L_0} (\Delta\theta)^2 + \frac{g}{L_0} \Delta L \Delta\theta - f \Delta L - \tau \Delta\theta \quad (8)$$

where S is the stretching modulus; C is the twist modulus; L is the extension; θ is the twist in radians; $\Delta L \equiv L - L_0$; and $\Delta\theta \equiv \theta - \theta_0$, with L_0 and θ_0 being the equilibrium extension and twist at zero force and torque. On the right-hand side of eq 8, the first and second terms are stretching and twisting energy, respectively; the third term is the stretch–twist coupling energy; the fourth and fifth terms are the work performed by the external force and torque, respectively.

Simultaneous minimization with respect to extension and force ($\partial U/\partial \Delta L = 0$ and $\partial U/\partial \Delta\theta = 0$) leads to the following equations for the elastic equilibrium:

$$\begin{cases} \Delta L = L_0 \left(\frac{f}{\tilde{S}} - \frac{g}{C\tilde{S}} \tau \right) \\ \Delta\theta = L_0 \left(-\frac{g}{C\tilde{S}} f + \frac{S}{C\tilde{S}} \tau \right) \end{cases} \quad (9)$$

where $\tilde{S} \equiv S - g^2/C$ is the effective stretching modulus. From the previous formulas, the sign of g can be immediately ascertained by looking at the behavior of ΔL as a function of τ at constant force, with positive and negative correlation implying $g < 0$ and $g > 0$, respectively (an analogous response is obtained at constant τ for $\Delta\theta$ as a function of f). For a quantitative assessment of the elastic constants, three equations are needed. Different strategies were pursued for the two simulation protocols considered in this work.

Benchmark Simulations. In this case, the torsion angle is estimated as $\theta = \sum_{i=1}^{n-1} \text{h-twist}_{i,i+1}$. The extension is computed as $L = |\mathbf{\Gamma}_n - \mathbf{\Gamma}_1|$. Moreover, no torque is applied, $\tau = 0$. From eq 9, one has $\Delta L = A_1 f$ and $\Delta\theta = A_2 f$, with $A_1 \equiv L_0/\tilde{S}$ and $A_2 \equiv -gL_0/C\tilde{S}$. The third equation is provided by applying the equipartition theorem to eq 8 and neglecting the twist–stretch coupling term.³¹ This is not a problem for the present purposes since it affects equally the analysis of both atomistic and coarse-grained results, thus not jeopardizing the quality of their comparison, although this detail should be kept in mind for the physical interpretation of the results. Application of the equipartition theorem gives $\text{var}(\Delta\theta) = \langle (\Delta\theta)^2 \rangle - \langle \Delta\theta \rangle^2 = k_B T L_0 / C$. Inverting these formulas, one finds

$$\begin{cases} \tilde{S} = \frac{L_0}{A_1} \\ C = \frac{k_B T L_0}{\text{var}(\Delta\theta)} \\ g = -\frac{A_2 C \tilde{S}}{L_0} \end{cases} \quad (10)$$

Fitting the linear dependence of L and θ on the pulling force f , we computed the constants L_0 , θ_0 , A_1 , and A_2 . $\text{var}(\Delta\theta)$ was instead calculated directly from the simulations performed at $f = 1$ pN. Plugging the values of these quantities into eq 10, we finally obtained the values of \tilde{S} , C , and g reported in Figure 6.

As for the crookedness constant k_β , following ref 52 we define it via the relation $\cos \beta = \cos \beta_0 (1 + f/k_\beta)$, where β_0 is

the crookedness at zero force. k_β is then suitably extracted from the slope of $\cos \beta$ versus f .

Stretch–Torsion Simulations. Taking advantage of the directionality imposed by the external force, the torsion angle in this case is computed as $\theta = \sum_{i=1}^{n-1} \psi_{i,i+1}$, where the angle ψ is calculated as the h-twist but considering the rejection of ξ_i from the direction of the force but instead of the helical axis. In the same way as done above, for $\tau = 0$ one has $\Delta L = A_1 f$, with $A_1 \equiv L_0/\tilde{S}$. At variance with the benchmark simulations, the presence of an imposed torque enables deriving all the equations without resorting to the analysis of the fluctuations. Indeed, for a constant pulling force $f = f_0$, $\Delta L = \text{constant} + A_3 \tau$ and $\Delta\theta = \text{constant} + A_4 \tau$, with $A_3 = -gL_0/C\tilde{S}$ and $A_4 = L_0 S/C\tilde{S} = L_0/C(1 + g^2/C\tilde{S})$. From the knowledge of the constants A_1 , A_3 , and A_4 one can obtain the elastic constants as

$$\begin{cases} \tilde{S} = \frac{L_0}{A_1} \\ C = \frac{A_1 L_0}{A_4 A_1 - A_3^2} \\ g = -\frac{A_3 L_0}{A_4 A_1 - A_3^2} \end{cases} \quad (11)$$

Fitting the linear dependence of ΔL on the pulling force f at $\tau = 0$, we computed the constants L_0 and A_1 . Analogously, fitting of ΔL and $\Delta\theta$ versus τ at $f = 2$ pN enabled the computation of A_3 and A_4 . Plugging the values of these quantities into eq 11, we finally obtained the values of \tilde{S} , C , and g reported in Figure 7.

Computation of Persistence Length. In order to compute the persistence length, for each base pair i we considered the geometrical center $\xi_i = (\mathbf{S}_{1,i} + \mathbf{S}_{2,i})/2$ of the sugar beads. By introducing the displacement vector $\mathbf{R}_{ij} \equiv \xi_j - \xi_i$, the i th tangent vector is defined as $\hat{\mathbf{t}}_i \equiv \mathbf{R}_{i-5, i+5}/|\mathbf{R}_{i-5, i+5}|$. The contour length l separating the points of application of two consecutive tangent vectors is computed as the average modulus of the displacement vector. The overall correlation function c_p is defined as

$$c_p(s) \equiv \langle \hat{\mathbf{t}}_i \cdot \hat{\mathbf{t}}_{i+s}/l \rangle_{\text{th},i} \quad (12)$$

where s is the contour length, while $\langle \dots \rangle_{\text{th},i}$ denotes the average over both the conformational ensemble and the value chosen for i . Therefore, c_p computes the decay of the orientation of the tangent vectors by taking thermal fluctuations into account.

Analogously, the static correlation function c_s is defined as

$$c_s(s) \equiv \langle \hat{\mathbf{t}}_i^{\min} \cdot \hat{\mathbf{t}}_{i+s}^{\min}/l \rangle_i \quad (13)$$

where $\hat{\mathbf{t}}_i^{\min}$ is the i th tangent vector computed on the minimum-energy structure and $\langle \dots \rangle_i$ denotes the average over the value chosen for i . The static correlation function quantifies the decay of the orientation of tangent vectors along the conformation corresponding to the ground state. For each sequence, the minimum-energy conformation was estimated by simulated annealing: starting from the average structure obtained via the molecular builder, we ran 100 short simulations at 300 K, after which the temperature was quenched at 0 K, while maintaining the Debye length corresponding to 300 K (this ensured that the minimum-energy configuration corresponded to the electrostatics at the temperature of interest). The ground state was then estimated

as the final conformation of the quenched simulation attaining the smallest value of the total energy.

Finally, following ref 56, the dynamic correlation function is defined as

$$c_d(s) \equiv \frac{c_p(s)}{c_s(s)} \quad (14)$$

This is in fact equivalent to impose, for each sequence, that the harmonic relation $1/l_p = 1/l_k + 1/l_d$ proposed by Trifonov and co-workers holds exactly.^{56,87}

Errors. Indeterminacies on the computed quantities were estimated according to the procedures below. Note that error bars are not present in the figures whenever they are smaller than the size of symbols.

Observables. For all coarse-grained simulations, several independent realizations were performed for each case, whose number depends on the particular set of simulations considered (see above). For each computed observable, the indeterminacy on the average was obtained as the standard error of the mean computed across the independent realizations. As for $\text{var}(\Delta\theta)$, the standard error of the variance was considered. In the case of the atomistic trajectories, the time series of each observable was decorrelated by performing a block analysis, from which the error was then estimated.⁸⁸

Elastic Constants. In order to account for the errors on the elastic constants, the following procedure was employed. In all cases, the constants were obtained starting from a linear relation $y = ax + y_0$, where x is either the force or the torque, while y is a certain observable (e.g., the extension). Let $x_i, y_i, i = 1, \dots, N$ be the N average values of the variables, and let dy_i be the error associated with y_i (x_i does not have an associated error). Assuming the values of y to be independent and distributed normally, we iteratively considered sets of variables $\bar{y}_{i,\alpha}$ extracted from Gaussians with average y_i and standard deviation dy_i , where $\alpha = 1, \dots, M$ indicates the particular realization considered. For each such realization, we obtained the fitting parameters a_α and $y_{0,\alpha}$ by linear regression. The mean slope was finally obtained as the average $\sum_{\alpha=1}^M a_\alpha / M$, while the associated error was estimated as the corresponding standard deviation. We checked numerically that a satisfying convergence was obtained for $M = 5000$.

Molecular Builder. The results obtained for the training sequences were employed to devise a molecular builder, which provides sequence-dependent average structures to be used as initial configurations. For each step XY, with both X and Y chosen among the four bases A, C, G, and T, we considered the simulations run for the training sequence which contains it. For instance, the sequence PolyAA was considered for the steps AA and TT. The corresponding trajectories were aligned in order to minimize the root-mean-square deviation (rmsd) of the central step XY, computed according to the Kabsch algorithm.⁸⁹ After this operation, the average coordinates of the 10 beads belonging to the step (four sugars, four bases, and two phosphates) were computed. These coordinates can be employed to build the average structure of a molecule, where the junctions between adjacent steps are aligned in order to minimize the rmsd of the overlapping base pair. As an example, let us consider the molecule with sequence 5'-ACT-3'. Two steps are present: AC and CT. The molecule is initiated by considering the average coordinates of the step AC. At this point, we note that the four beads (two sugars and two bases)

belonging to the second base pair of the step AC are the same as the four beads belonging to the first base pair of the next step, CT. The average coordinates of the latter are then translated and rotated in order to minimize the rmsd between the two sets of coordinates for the overlapping beads. The final coordinates for the four overlapping beads are computed as the mean of the two aligned sets. This procedure can be iterated for molecules of arbitrary length. As a last step, the final structure is translated so that the origin corresponds to the first sugar bead and is subsequently aligned to the z axis. The molecular builder is provided at <https://github.com/saassenza/MADna/tree/main/LAMMPS> and, using as input only the sequence, creates both the initial coordinates and the double-stranded topology in LAMMPS format.

RESULTS AND DISCUSSION

Summary of the Coarse-Grained Model. In line with previous work,^{57,65} the coarse grain considered in MADna describes each nucleotide by means of three beads, each centered on the sugar, phosphate group, and base, respectively (Figure 1). The beads interact with each other via steric interactions and, in the case of phosphate groups, electrostatic repulsion. The double-stranded molecular structure is maintained by introducing bonded interactions (eqs 1, 2, and 3 and Figure 3), with parameters depending on the sequence up to the level of a single step. The parameters were determined by Boltzmann inversion of atomistic trajectories from the literature,⁵² whose sequences encompass all the possible steps (hereby referred to as training sequences, see section S3.1 in the Supporting Information for details). Further tuning was performed in order to reproduce the elastic response to an external pulling force computed for the same set of sequences. Full details of the model and the coarse-graining procedure are found in the Methods and in Sections S1 and S2 in the Supporting Information.

MADna Reproduces Conformational Features from Atomistic Simulations. In order to benchmark the optimized simulation setup, we proceeded with a systematic comparison between coarse-grained predictions and results from atomistic simulations. We performed coarse-grained simulations of dsDNA molecules under the action of a pulling force f ranging in the interval 1–20 pN, following the same protocol as in refs 46 and 52 (see Methods for details on the implementation). Apart from the training sequences, we considered a second, independent set of molecules (testing sequences), which has also been studied in refs 46 and 52. The testing sequences are reported in Section S3.1 in the Supporting Information and include biologically relevant structures as well as synthetic A-tracts fragments. Both conformational and elastic properties were considered for the analysis.

The study of various conformational quantities is reported in Figure 5. For each observable, a scatter plot is depicted to compare coarse-grained and atomistic results at the smallest force considered, $f = 1$ pN. Training and testing sequences are indicated by red circles and green diamonds, respectively, while the black line indicates the bisector of the first and third quadrant. Overall, an excellent agreement is found, as indicated by the localization of the plotted points in the vicinity of the bisector and by the large values attained by the Pearson coefficient for all the considered quantities. In Figure 5a, we plot the crookedness β ,⁵² a dimensionless parameter accounting for the global deviation of the helical axis with

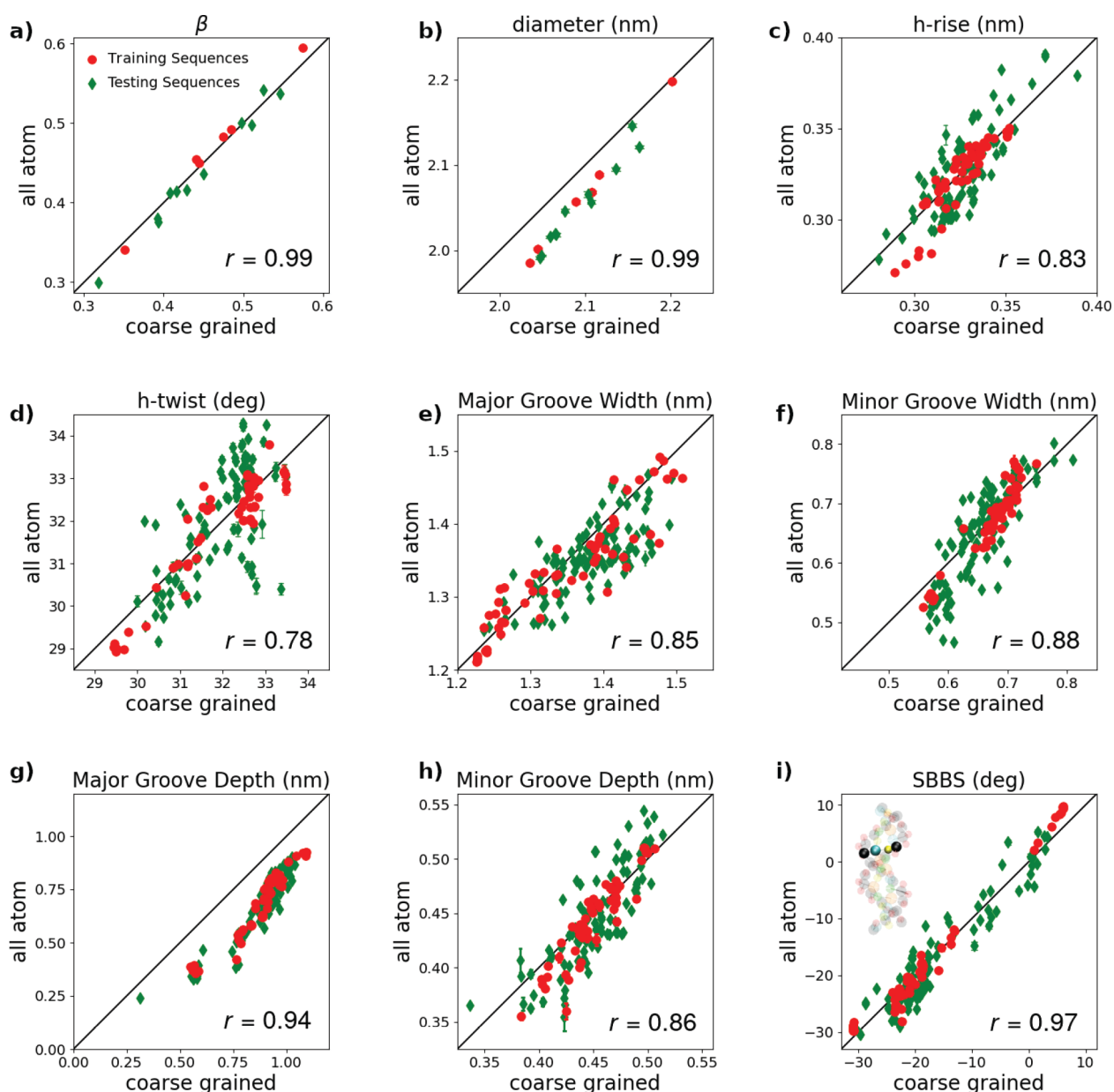


Figure 5. Scatter plot comparing atomistic and coarse-grained results at $f = 1$ pN for various structural features: crookedness β (panel a); helical diameter (b); helical rise (c); helical twist (d); width of major (e) and minor (f) groove; depth of major (g) and minor (h) groove; SBBS dihedrals (i). Training and testing sequences are denoted by red circles and green diamonds, respectively. Black lines indicate the bisector of the first and third quadrant. For each panel, the Pearson coefficient indicating the linear correlation between the two data sets is reported. The atomistic results were obtained by coarse-graining the trajectories obtained from all-atom simulations and performing the analysis reported in the [Methods](#).

respect to a straight line, i.e., the presence of a spontaneous curvature. Larger values of β correspond to more curved structures, and it was found that A-tracts show the straightest conformations.⁵² In [Figure 5b](#), we analyze the helical diameter. Also in this case, the coarse-grained model faithfully captures the dependence on sequence within a range spanning about 0.2 nm, although a slight overestimation of the atomistic values can be appreciated. In [Figure 5c–h](#) we plot the comparison of various helical features, namely h-rise ([Figure 5c](#)), h-twist ([Figure 5d](#)), and groove geometry ([Figure 5e–h](#)). The coarse-grained predictions closely follow the sequence dependence of the atomistic values, although a systematic overestimation of about 0.2 nm is present in the case of the depth of the major groove ([Figure 5g](#)). Finally, in [Figure 5i](#) we compare the values for the local dihedral SBBS, i.e., the dihedral formed by the

sugars and bases within each base pair (inset in [Figure 5i](#)). Again, MADna reproduces with high precision the large variability induced by sequence heterogeneity, showing its ability to capture features emerging from the interaction between neighboring base pairs.

The close agreement between coarse grain and atomistic simulations indicates the high accuracy of MADna in describing the conformational properties of dsDNA. Several striking features are worth mentioning. First, as mentioned above, the testing sequences were not employed to build the model, so that in this case the coarse-grained results are pure predictions. Second, none of the quantities reported in [Figure 5](#) were directly used to build the model; thus, a quantitative agreement is not trivial also in the case of the training sequences. Third, these observables are related to different

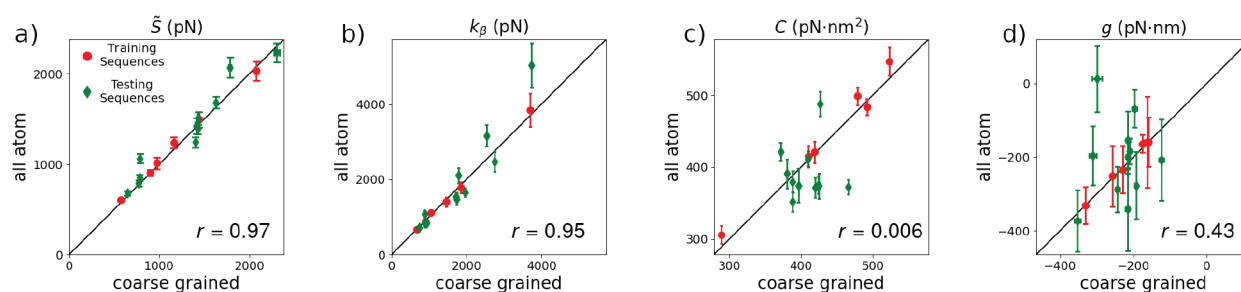


Figure 6. Scatter plot comparing atomistic and coarse-grained results for the effective stretching modulus \tilde{S} (panel a), the crookedness rigidity k_β (b), the torsion modulus C (c), and the twist–stretch coupling constant g (d). Training and testing sequences are denoted by red circles and green diamonds, respectively. Black lines indicate the bisector of the first quadrant. For each panel, the Pearson coefficient indicating the linear correlation between the data sets corresponding to the testing sequences is reported.

scales, encompassing a single base pair (SBBS), a base step (h-twist and h-rise), multiple-step geometry (grooves), and the molecule as a whole (β and diameter). Fourth, the selected quantities have a stark dependence on sequence, as shown by their wide range of variability, indicating that the model captures also this intrinsic heterogeneity. Particularly striking is the case of the dihedral SBBS: despite the only two possible choices for the bases (the Watson–Crick pairs AT and CG), this quantity displays a strong variability, ranging from -30 to 10 deg (Figure S5). This is evidently an emergent behavior induced by the interaction of these base pairs with their neighbors and is perfectly reproduced by the coarse-grained model.

MADna Reproduces Elastic Constants Obtained from Atomistic Simulations. Next, we focused on the elastic properties of training and testing sequences (Figure 6). The effective stretching modulus \tilde{S} (Figure 6a) and the crookedness elastic constant k_β (Figure 6b) are related to the elastic response of extension and crookedness to the external force, respectively, and are obtained as the slopes of the corresponding observables as a function of f (see Methods and Figure S3 in the Supporting Information). The torsional modulus C (Figure 6c) accounts for the change of the h-twist upon application of an external torque. For the present setup, C was computed by analyzing the fluctuations of the h-twist for $f = 1$ pN. Finally, the twist–stretch coupling constant g (Figure 6d) quantifies the torsional response to the external force. The negative sign displayed by g (Figure 6d) implies that the molecule overwinds when stretched, in agreement with experimental observations.^{31,32} Precise definitions of the elastic constants and their computation are reported in the Methods.

As discussed in the Methods and in Section S2 in the Supporting Information, the values of \tilde{S} , k_β , C , and g for the training sequences were used to refine the parameters of the bonded interactions obtained by the Boltzmann inversion of the atomistic trajectories. Due to the large number of bonded parameters, we opted for a pragmatic approach and adjusted a minimal subset of parameters. Particularly, we empirically observed that a major impact on the elastic constants was obtained by tuning the rigidities of bond $5'$ -BB- $3'$, angle $5'$ -PSB- $3'$, and dihedral $5'$ -SPSP- $3'$. Interestingly, these three bonded interactions have a clear physical meaning (compare Figure 3). Indeed, the bond $5'$ -BB- $3'$ accounts for stacking interactions, which are likely to affect the stretching stiffness⁵² as well as the coupling between twist and stretch deformations. Moreover, the range of values encompassed by the angle $5'$ -PSB- $3'$ is related to the flexibility of the sugar pucker, which has been shown to play a key role in dsDNA elasticity.⁴⁶

Finally, the $5'$ -SPSP- $3'$ dihedral is likely to affect the backbone response and was found to be the quantity most affecting the crookedness rigidity. In this context, it is also worth mentioning that the BB-WC elastic constants for the pairs AT and CG approximately follow the ratio 2:3 (see Table S2 in the Supporting Information). The BB-WC bonds account for the interactions between Watson–Crick base pairs; hence, this ratio nicely reflects the presence of two and three hydrogen bonds for AT and CG, respectively.

In Figure 6, we show the scatter plots comparing \tilde{S} , k_β , C , and g for both the training and the testing sequences, while the corresponding statistical indicators are reported in Table 1.

Table 1. Comparison between the Average Elastic Constants Obtained by MADna and Atomistic Simulations (Compare Figure 6)^a

		training sequences	testing sequences
\tilde{S} (pN)	all atom	1214 (475)	1358 (481)
	coarse-grained	1188 (477)	1307 (490)
k_β (pN)	all atom	1617 (1060)	1901 (1226)
	coarse-grained	1609 (1014)	1790 (894)
C (pN·nm ²)	all atom	445 (77)	392 (36)
	coarse-grained	435 (77)	409 (26)
g (pN·nm)	all atom	-216 (63)	-229 (87)
	coarse-grained	-219 (62)	-228 (60)

^aIn parentheses the standard deviations are reported.

MADna closely captures the average and sequence-induced standard deviation of all the elastic quantities (Table 1). While this is expected for the training sequences, which were used to refine the model, the excellent agreement found for the testing sequences (average within 6% and standard deviation within 30% of the atomistic values) indicates the good performance of our model. The capacity of MADna to describe the detailed sequence dependence can be assessed by looking at Figure 6. MADna precisely accounts for \tilde{S} and k_β (Figure 6a and Figure 6b, respectively), as indicated by the large values obtained for the Pearson coefficient (calculated considering only the testing sequences). In contrast, despite reproducing the average and standard deviation of C , MADna does not capture its detailed sequence dependence, as indicated by the low value of the Pearson coefficient. While this can be due to an intrinsic limit of our model, we note that also the input data set might be at fault: in contrast with \tilde{S} and k_β , which are obtained by considering average values of the corresponding observables, C is determined by computing the fluctuations of the cumulated twist. Fluctuations are knowingly slower to converge when

compared to averages, suggesting that the lack of accuracy of MADna might be due to sampling limitations of the atomistic simulations. Since to our knowledge there are currently no experiments addressing the sequence dependence of the twist modulus, and given the excellent performance of MADna in capturing the average value of C (see Figure 7c below), we decided to leave a thorough assessment of the matter to future investigation. Finally, the twist–stretch coupling g displays a good linear correlation. The modest value of the Pearson coefficient is an expected side effect of the large indeterminacy of the atomistic results, for which the error bars have sizes comparable to the dispersion of the average values. It has to be noted that some outliers are present; in particular for sequence A8T the coarse-grained model predicts a twist–stretch coupling markedly different from the atomistic result $g \simeq 0$. Nevertheless, as discussed in Section S5.1 in the Supporting Information, this disagreement is likely to originate from a possible lack of convergence of the atomistic simulations at large forces for this specific sequence. Based on this observation, we excluded this point for the quantitative evaluation of the agreement between the two data sets by means of the Pearson coefficient reported in Figure 6d and for the corresponding statistical indicators in Table 1.

Comparison with Experiments: Sequence-Averaged Persistence Length. The persistence length l_p quantifies the bending rigidity of a polymer and is possibly the most characterized mechanical property of dsDNA.^{16–19,21,23,29,37,49} In its classical formulation from polymer theory, l_p corresponds to the length of the fragment to be considered in order to observe significant thermally induced bending effects.⁹⁰ More technically, l_p is defined as the decay length of the thermally averaged tangent-vector correlation function c_p , that is, $c_p = \exp(-s/l_p)$, where s is the contour length of the polymer fragment under consideration.⁹⁰ This definition relies on the assumption that the minimum-energy conformation for the polymer is that of a straight rod. In the case of dsDNA, according to the sequence a spontaneous curvature may be present. In order to account separately for intrinsic bending and thermal fluctuations, it is customary to distinguish the static (l_s) and dynamic (l_d) contributions to the persistence length, which characterize the decay length of the suitably defined correlation functions $c_s = \exp(-s/l_s)$ (computed from the minimum-energy structure) and $c_d = \exp(-s/l_d)$.⁵⁶ The three lengths are approximately related by a harmonic sum, $1/l_p \simeq 1/l_s + 1/l_d$.⁸⁷

In order to compute the persistence length and its contributions, we performed simulations for 20 random sequences of length 100 base pairs. The minimum-energy structures (see Methods) and the equilibrated trajectories were then employed to compute the correlation functions c_p , c_s , and c_d according to eqs 12, 13, and 14, respectively. The results are reported in Figure S5 in the Supporting Information for various definitions of the tangent vector. Here, we focus on the results of one of such definitions, according to which the tangent vectors are obtained by joining the geometrical centers of sugars of base pairs separated by 10 steps (see Methods). Fitting the correlation function c_p with an exponential decay resulted in computed values of the persistence length l_p ranging between 46 and 64 nm, with an average equal to $l_p = 56 \pm 1$ nm. This is in good agreement with experimental values on random sequences and standard ionic conditions (45–55 nm^{16–19,21,23,29,37,49}), particularly since l_p was not employed in the parametrization of the model. Other coarse-grained models

give predictions for l_p within the experimental range,^{60,62,65,68,71} although in most of these works the persistence length of double-^{60,68,71} or single-stranded⁶⁵ DNA was employed as a target quantity in the construction of the force field. In the present case, the slight overestimation of l_p is in line with previous results of atomistic simulations, where an average $l_p = 57 \pm 3$ was found.⁴⁹

As expected,⁵⁶ for each sequence the static correlation function c_s is found to decrease more slowly than c_p (Figure S5 in the Supporting Information) since it lacks the disordering action of thermal fluctuations. The corresponding static persistence length l_s was computed by fitting c_s via an exponential decay, in analogy to c_p . The obtained values vary widely, ranging from 199 to 796 nm, with an average equal to $l_s = 485 \pm 42$ nm. This is somewhat smaller than previous estimations ($l_s = 576 \pm 191$ nm⁴⁹), although we note that this value varies wildly by considering different sequences and, even for the same sequences, by employing different definitions (see below). As discussed in ref 49, its large heterogeneity may rationalize the markedly different results reported in experiments, where values of l_s as different as 130 nm and >1000 nm have been estimated.^{16,91}

Fitting c_d via an exponential decay yields values for the dynamic persistence length l_d ranging between 56 and 77 nm, with an average equal to $l_d = 64 \pm 1$ nm. This prediction is in line with a previous report based on atomistic simulations⁴⁹ ($l_d = 64.7 \pm 1.4$ nm) and lies between the results obtained from cryoelectron microscopy¹⁶ ($l_d = 82 \pm 15$ nm) and cyclization experiments⁹¹ ($l_d = 50 \pm 1$ nm). The difference between the two experimental values might be ascribed to the different techniques employed, for which future simulations with the present model may shed some light.

As a final consideration, it has to be noted that the estimation of the persistence length may vary according to the definition of the tangent vectors.⁵⁶ Alternative choices are considered in Section S6 in the Supporting Information, yielding average values in the ranges $l_p = 56$ –63 nm, $l_s = 485$ –1641 nm, and $l_d = 64$ –69 nm. This indicates the robustness of the values of l_p and l_d with respect to the definition of the tangent vectors. The average value of l_s is sensitive to the definition, although the observed variation can be partly ascribed to the intrinsic large variability characterizing this quantity.

Comparison with Experiments: Sequence-Averaged Elastic Constants. A further set of simulations was devoted to determine the sequence-averaged elastic constants. These quantities were already estimated for the training and testing sequences (Figure 6), but their quantitative determination is likely to depend on the details of the microscopic definitions, particularly for the h-twist. In order to enable a quantitative comparison with experimental values, we designed a simulation framework which avoids relying on such definitions and which is more akin to single-molecule experimental setups.

As shown schematically in Figure 7a, in this set of simulations a dsDNA molecule is subjected to a constant force f and torque τ acting along the z axis. This enables defining unequivocally the twist angle θ starting from the projection of the base pairs onto the xy plane, in analogy with single-molecule experiments based on rotor beads.³¹ The linear response of the extension L as a function of f provides access to the effective stretch modulus \tilde{S} . Analogously, the torsion modulus C can be computed from the dependence of θ on τ . Finally, the twist–stretch coupling g can be obtained by

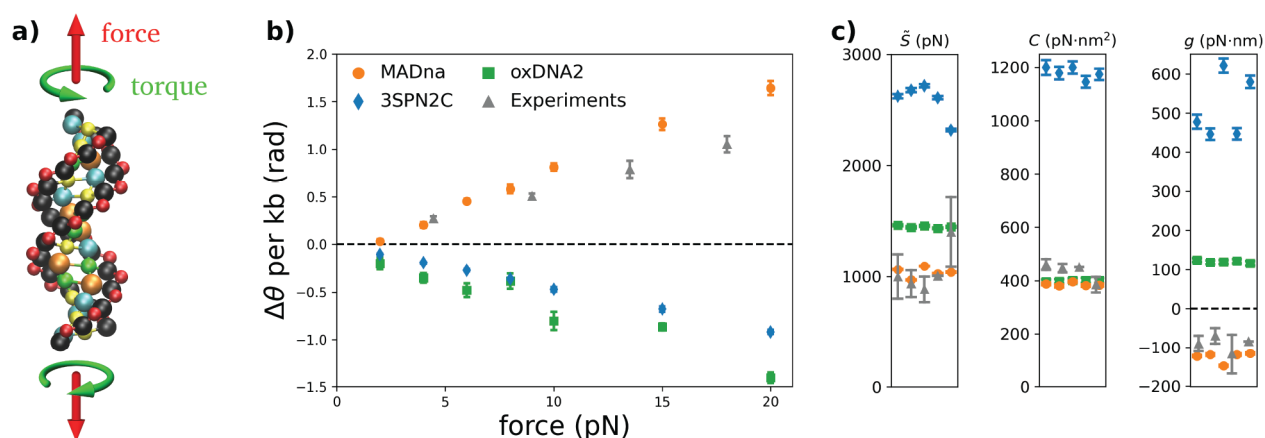


Figure 7. (a) Schematic description of the simulation setup for the stretch–torsion simulations, prescribing a constant force and torque applied along a fixed direction. (b) Twist response to the external force in the absence of imposed torque for MADna (orange circles), oxDNA2 (green squares), 3SPN2C (blue diamonds), and rotor-bead experiments³¹ (gray triangles). The black dashed line indicates $\Delta\theta = 0$; therefore, overwinding and unwinding responses are characterized by points lying above and below the line, respectively. The simulation data correspond to the average of the five sequences considered. (c) Effective stretching modulus \tilde{S} , torsion modulus C , and twist–stretch coupling constant g obtained by the three models for the five sequences reported in Section S3.3 in the Supporting Information. Symbols are the same as in panel b. The gray triangles correspond to experimental measures obtained for unrelated sequences in refs 17–20 for \tilde{S} , refs 20, 27, 31, and 92 for C , and refs 20, 31, 32, and 93 for g . In the case of g , the dashed black line denotes $g = 0$, thus separating two regimes characterized by qualitatively different twist–stretch coupling.

looking at the cross-dependence, i.e., the dependence of L on τ or the response of θ to changes in f . The quantitative details can be found in the Methods.

We performed simulations for five randomly generated sequences of length equal to 40 base pairs at several values of f and τ (see Methods for full details and Section S3.3 in the Supporting Information for the sequences, which are named ST1, ..., ST5). With this choice of sequence length, the molecules are long enough so that several turns of the double helix are present but short enough to neglect the effect of bending. For comparison, we also run simulations for the same set of sequences and parameters by employing the two most widely used coarse-grained models from the literature, namely oxDNA2^{80,81} and the sequence-dependent model 3SPN2C.⁸² From a qualitative perspective, a marked difference between MADna and the other two models becomes evident when analyzing the change in twist $\Delta\theta$ as a function of the force (Figure 7b). Indeed, the present model prescribes that dsDNA overwinds when stretched ($\Delta\theta > 0$), which is in agreement with experimental observations.^{31,32} In contrast, neither oxDNA2 nor 3SPN2C capture this feature, predicting unwinding upon pulling (in the case of oxDNA this fact was already observed in the original publication⁶⁸). To our knowledge, no other coarse-grained model available in the literature has been shown to predict this unintuitive behavior of dsDNA.

MADna shows better agreement with experiments also from a quantitative standpoint. As shown in Figure 7c and Table 2, it predicts values for \tilde{S} , C , and g in good quantitative agreement with the experiments. As for the other models, oxDNA2 shows a similar performance for C , while it tends to overestimate \tilde{S} . Coherently with the results shown in Figure 7b, the wrong sign for g is found for both oxDNA2 and 3SPN2C, with the latter showing in general a significant overestimation of the elastic constants. The results reported in Figure 7c and Table 2 correspond to forces larger than 10 pN, which was chosen as a reasonable threshold to avoid bending effects. Nevertheless, performing the analysis with the full set of forces resulted in a

Table 2. Elastic Constants Obtained by the Various Models and Their Comparison with Experiments (see Figure 7c)^a

	\tilde{S} (pN)	C (pN·nm ²)	g (pN·nm)
experiment	1045 ± 92	436 ± 17	−90 ± 10
MADna	1038 ± 21	386 ± 3	−125 ± 6
oxDNA2	1448 ± 5	399 ± 1	+120 ± 1
3SPN2C	2589 ± 71	1180 ± 10	+514 ± 36

^aExperimental values are obtained as averages of the results reported in refs 17–20 for \tilde{S} , refs 20, 27, 31, and 92 for C , and refs 20, 31, 32, and 93 for g , while the corresponding indeterminacies are computed as the standard error of the mean.

small change in \tilde{S} (roughly 10%) and in virtually no change in C and g . It is also worth mentioning that the range considered for the torque (0–30 pN·nm) corresponds to supercoiling densities below the threshold value $\sigma \approx 0.05$ usually associated with torque-induced denaturation,²² hence making the results obtained for MADna relevant for real dsDNA molecules within the whole range of applied torques. Particularly, we computed the supercoiling density as $\sigma = \Delta\theta/\theta_0$, since the lack of bending results in the absence of relevant writhe. For the largest torque applied ($\tau = 30$ pN·nm), we found $\sigma \approx 0.046$ for MADna, $\sigma \approx 0.043$ for oxDNA2, and $\sigma \approx 0.014$ for 3SPN2C, which is lower than for the other two models due to the larger value of the twist modulus C .

The values of g obtained for MADna in Table 1 and Table 2 differ by a factor of 2 with respect to each other. This striking disagreement appears to be too large to be simply ascribed to the different sequences considered or the different definitions employed for the twist angle. In this regard, based on atomistic simulations it has been suggested that the stretching modulus depends on the size of the fragment under consideration.⁹⁴ Since in the present case we are comparing sequences of size 20 and 40 base pairs, it might be that a similar effect is present for g . In order to check this hypothesis, we performed another set of simulations following the same protocol as in the present section but considering sequences made of 20 base pairs,

which are listed as ST1-short, ..., ST5-short in Section S3.3 in the Supporting Information. The resulting elastic constants are reported in Figure S6 in the Supporting Information. Particularly, we found that for the short sequences $g = -258 \pm 16$ pN·nm, which is indeed in line with the results reported in Figure 6. This confirms the presence of size effects in MADna for the prediction of elastic constants. To our knowledge, this is the first instance in which a length dependence for g has been predicted. This feature might explain the systematic overestimation of g in atomistic simulations that are restricted to short sequences.⁴⁶

Another interesting effect observed in experiments is the coupling between twist and bending, which is relevant at pulling forces up to a few piconewtons.^{33,34} As a consequence of this coupling, the twist of the DNA molecule appears to be softer; i.e., the effective twist modulus is $C_{\text{eff}} < C$. In order to study this feature, we considered a set of simulations involving DNA molecules of 150 base pairs, which is comparable to the persistence length, thus enabling the presence of bending fluctuations. The simulation protocol was similar to the case just analyzed, although here no torque was applied. We considered three independent sequences which are reported in Section S3.4 in the Supporting Information. For each sequence, simulations were performed for MADna, α xDNA2, and 3SPN2C at pulling forces ranging between 0.05 and 2.5 pN. Further simulation details can be found in the Methods. In Figure 8 we report the results obtained and compare them with

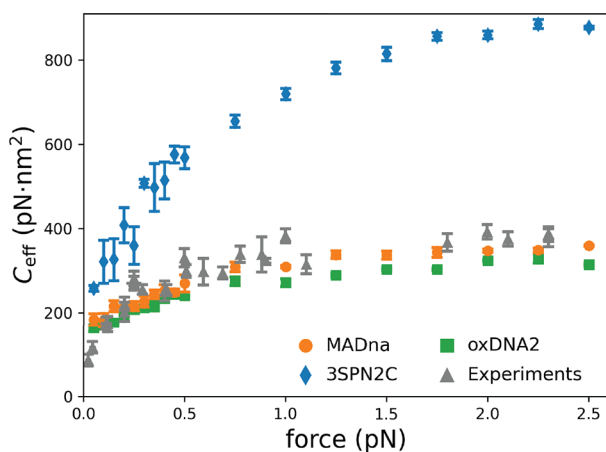


Figure 8. Effective torsion modulus C_{eff} as a function of force for MADna (orange circles), α xDNA2 (green squares), 3SPN2C (blue diamonds), and experiments (gray triangles). Experimental data were extracted from refs 33 and 34.

the experimental values from refs 33 and 34. In line with the trend observed in Figure 7, MADna and α xDNA2 follow quite closely the experimental curve, while 3SPN2C systematically overestimates the value of C_{eff} .

Comparison with Experiments: Sequence-Dependent Conformation and Elasticity. Having characterized and benchmarked the sequence-averaged mechanical properties of MADna, we now turn to sequence-dependent features. In this regard, we considered the sequences studied experimentally in ref 28, where the authors characterized the sequence-dependent persistence length and helical repeat by means of cyclization experiments.

We performed simulations for DNA molecules of 100 base pairs obtained by taking the central parts of 14 different

experimental sequences as listed in Section S3.2 in the Supporting Information. For each simulated sequence, we computed the helical pitch by dividing the cumulated helical twist by 2π , while for the persistence length we employed the same approach as above. In Figure 9 we report the comparison

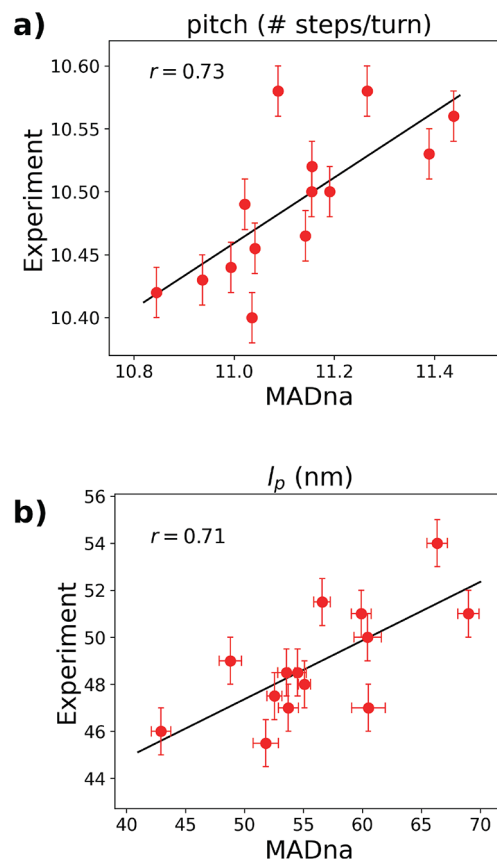


Figure 9. Comparison between experimental values and MADna predictions for the sequence dependence of the helical pitch (a) and the persistence length l_p (b). The lines correspond to the linear fits of the scatter plots and are included as a guide to the eye. The value of the Pearson coefficient is reported in each plot.

between experiments and MADna predictions. As the plots show, there is a high correlation between the two data sets, thus indicating that MADna satisfactorily captures the sequence dependence of two key features of DNA conformation (helical pitch) and elasticity (persistence length). From a quantitative perspective, we see from Figure 9a that the simulated values for the pitch are larger than the experimental ones. Nonetheless, a quantitative comparison has to be performed with care. Indeed, the experimental values were obtained indirectly by analyzing cyclization data by means of a wormlike chain with twist.²⁸ Although the relative differences observed in the experimental results for the different sequences are robust, the values depend on the validity of this model down to the scale of single steps and on the value assigned to the twist modulus when fitting the data. Moreover, it is likely that the details of the definition of the h-twist in the simulations affect the quantitative determination of the helical pitch. To our knowledge, this is the first instance in which these structural data have been compared to predictions from simulations. As for the persistence length, the strong correlation found has to be mitigated by the slight quantitative

overestimation of l_p in simulations, in line with the results reported above for the sequence-averaged persistence length. Moreover, the values of l_p from simulations appear to be more heterogeneous than in experiments. While this can be ascribed to intrinsic limitations of MADna, it has to be noted that the smaller dispersion of values determined from experiments might originate from some assumptions made in their analysis. Indeed, experimental persistence lengths were obtained by fitting j -factors from cyclization experiments, where the theoretical formulas depend on l_p , the helical pitch, the twist modulus C , and the contour length L .²⁸ The data reported in ref 28 were obtained by assuming sequence-independent C and L , which neglects two sources of sequence-driven heterogeneity, suggesting that the values obtained in experiments might underestimate the actual range of variability of l_p .

For comparison, we also performed simulations for oxDNA2 and 3SPN2C, for which the results are reported in Figure S7 in the Supporting Information. We found that oxDNA2 predictions do not correlate with experiments for either the helical pitch or the persistence length. This is expected, since in this model the sequence dependence is implemented by tuning the base-pairing and stacking interactions to account for thermodynamic data but not for the elasticity of DNA. In the case of 3SPN2C, we found a weak correlation for the helical pitch but a strong correlation for the persistence length. This high correlation was also expected, since the persistence-length data from ref 28 were used to parametrize the model.⁸² For completeness, we mention that also CGDNA has been used to reproduce these experimental persistence-length data, for which a similar correlation as for MADna was found ($r = 0.73$).⁵⁶

As a further test, we performed stretching simulations for phased A-tracts, i.e., dsDNA molecules obtained by alternating fragments of consecutive adenines and random sequences, with each fragment having a length of 5–10 base pairs. It was experimentally found that the stretching modulus for such molecules is roughly 50% larger than for random sequences.²¹ We run some stretching simulations for five phased A-tracts of 40 base pairs, whose sequences were extracted as fragments of the experimental ones²¹ and are reported as A-tract-1, ..., A-tract-5 in Section S3.3 in the Supporting Information. Simulation protocols and analysis were the same as in the case of random sequences studied above. The results from the three models were $\bar{S} = (1402 \pm 48)$ pN for MADna, $\bar{S} = (1444 \pm 2)$ pN for oxDNA2, and $\bar{S} = (2793 \pm 67)$ pN for 3SPN2C. By comparing these values with the results reported in Table 2, we thus conclude that MADna predicts an increase of roughly 35%, oxDNA2 does not predict any increase, and 3SPN2C predicts a 8% increase. As expected, the absence of elasticity-oriented sequence dependence for oxDNA2 prevents it from capturing this feature. The prediction of MADna is the one most closely resembling the significant increase observed experimentally, although still underestimating its extent.

Strengths and Limits. The comparison between MADna, oxDNA2, 3SPN2C, and experiments can be summarized as follows. MADna performs in general better at capturing both sequence-averaged and sequence-dependent conformational and elastic features. As for the other two models, oxDNA2 reproduces well the sequence-averaged properties, but it cannot account for sequence dependence. In contrast, 3SPN2C captures well the sequence dependence of elasticity in the case of the persistence length, while its accuracy for the change in stretch modulus or conformation is more limited.

Moreover, it tends in general to overestimate the magnitude of the elastic constants.

The accuracy of MADna in addressing the sequence dependence of conformational and elastic properties of dsDNA appoints it as the ideal choice to interpret the outcome of single-molecule experiments, as well as to study the conformational changes induced by mechanic stress, which are relevant for many systems in vivo. Moreover, it can be employed as a virtual laboratory to test analytical theories on DNA elasticity.^{95,96} Nonetheless, at the present stage MADna cannot account for breaking events such as the formation of kinks or local melting. Hence, it is important to assess the relevance of such mechanisms for the system under study before drawing conclusions based on simulations performed with the present model. We are currently working to surpass such limitations, in order to further extend the palette of possible systems which can be analyzed through the lens of MADna.

Further limitations are introduced by the use of an implicit solvent, which prevents studying DNA solvation. Similarly, the use of reduced charges interacting via the Debye–Hückel potential does not account for ion–ion correlations, which are particularly relevant for multivalent ions.⁹⁷ We note that MADna shares these limitations with all the coarse-grained models based on a similar philosophy, including, e.g., oxDNA2 and 3SPN2C. In view of future modeling of the interaction of DNA with proteins by means of MADna, the potential obstacle provided by these limitations can however be overcome by considering a distance-dependent dielectric constant for the electrostatic interaction between charged sites on DNA and proteins.⁹⁸

LAMMPS Implementation and Availability. The standard potentials employed in MADna enable a straightforward implementation of the model in all common Molecular Dynamics simulators. The molecular builder, the main scripts used for the analysis, and some sample scripts for simulation in LAMMPS are provided at <https://github.com/saassenza/MADna/tree/main/LAMMPS>. The molecular builder takes as input the sequence and provides both the topology and the initial coordinates in LAMMPS format. In order to run MADna in LAMMPS (at the time of writing, the stable version is 29Sep2021), the latter has to be built with the packages MOLECULE, EXTRA-MOLECULE, and EXTRA-PAIR.

CONCLUSION

We have introduced MADna, a novel coarse-grained model for the simulation of double-stranded DNA. MADna captures the sequence dependence of conformational and elastic properties with accuracy comparable to that of atomistic simulations. Key conformational features which closely follow atomistic results include the main helical parameters, the groove geometry, the diameter of the double helix, and the spontaneous curvature as quantified by the crookedness. The model also predicts sequence-averaged and sequence-dependent elasticity and conformation in agreement with experimental results for a wide set of features, namely the stretching and torsion moduli, the counterintuitive negative twist–stretch coupling, the twist–bend coupling, the persistence length, and the helical pitch. At the present stage, the model does not account for breaking events, which are being addressed in ongoing work. The implementation of MADna in Molecular Dynamics software is straightforward due to the common potentials

employed. Sample scripts for LAMMPS and a molecular builder are openly provided on GitHub.

Due to both its accuracy and its simplicity of use, we believe that MADna provides a significant addition to the toolbox of coarse-grained simulations and will enable precise theoretical studies of a wide set of large-scale DNA phenomena.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.2c00138>.

Full list of the parameters of the model, additional analysis, and additional methods (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Rubén Pérez – Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, E-28049 Madrid, Spain; orcid.org/0000-0001-5896-541X; Email: ruben.perez@uam.es

Author

Salvatore Assenza – Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, E-28049 Madrid, Spain; orcid.org/0000-0001-9983-8927

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.2c00138>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are indebted to Alberto Marín-González and Alessandro Barducci for insightful discussions. We are thankful for the financial support from the Spanish MINECO project MAT2017-83273-R. We acknowledge support from the Ministerio de Ciencia e Innovación (MICINN) through the project PID2020-115864RB-I00 and the “María de Maeztu” Programme for Units of Excellence in R&D (grant No. CEX2018-000805-M). The project that gave rise to these results received the support of a fellowship from “la Caixa” Foundation (ID 100010434) and from the European Union’s Horizon research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 847648. The fellowship code is LCF/BQ/PI20/11760019.

■ REFERENCES

- (1) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233–269.
- (2) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* **2009**, *461*, 1248–1253.
- (3) Segal, E.; Widom, J. Poly (dA: dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **2009**, *19*, 65–71.
- (4) Haran, T. E.; Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.* **2009**, *42*, 41–81.
- (5) Seol, Y.; Hardin, A. H.; Strub, M.-P.; Charvin, G.; Neuman, K. C. Comparison of DNA decatenation by *Escherichia coli* topoisomerase IV and topoisomerase III: implications for non-equilibrium topology simplification. *Nucleic Acids Res.* **2013**, *41*, 4640–4649.

- (6) Dillon, S. C.; Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **2010**, *8*, 185–195.

- (7) Saha, A.; Wittmeyer, J.; Cairns, B. R. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 437–447.

- (8) Lee, J. Y.; Terakawa, T.; Qi, Z.; Steinfeld, J. B.; Redding, S.; Kwon, Y.; Gaines, W. A.; Zhao, W.; Sung, P.; Greene, E. C. Base triplet stepping by the Rad51/RecA family of recombinases. *Science* **2015**, *349*, 977–981.

- (9) Ong, M. S.; Richmond, T. J.; Davey, C. A. DNA stretching and extreme kinking in the nucleosome core. *J. Mol. Biol.* **2007**, *368*, 1067–1074.

- (10) Dikic, J.; Seidel, R. Anticooperative binding governs the mechanics of ethidium-complexed DNA. *Biophys. J.* **2019**, *116*, 1394–1405.

- (11) Chhabra, H.; Mishra, G.; Cao, Y.; Prešern, D.; Skoruppa, E.; Tortora, M. M.; Doye, J. P. Computing the Elastic Mechanical Properties of Rodlike DNA Nanostructures. *J. Chem. Theory Comput.* **2020**, *16*, 7748–7763.

- (12) Mathew-Fenn, R. S.; Das, R.; Harbury, P. A. Remeasuring the double helix. *Science* **2008**, *322*, 446–449.

- (13) Shi, X.; Herschlag, D.; Harbury, P. A. Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E1444–E1451.

- (14) Ido, S.; Kimura, K.; Oyabu, N.; Kobayashi, K.; Tsukada, M.; Matsushige, K.; Yamada, H. Beyond the helix pitch: direct visualization of native DNA in aqueous solution. *ACS Nano* **2013**, *7*, 1817–1822.

- (15) Kuchuk, K.; Katrivas, L.; Kotlyar, A.; Sivan, U. Sequence-dependent deviations of constrained DNA from canonical B-form. *Nano Lett.* **2019**, *19*, 6600–6603.

- (16) Bednar, J.; Furrer, P.; Katritch, V.; Stasiak, A.; Dubochet, J.; Stasiak, A. Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J. Mol. Biol.* **1995**, *254*, 579–594.

- (17) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185–6190.

- (18) Wenner, J. R.; Williams, M. C.; Rouzina, I.; Bloomfield, V. A. Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophys. J.* **2002**, *82*, 3160–3169.

- (19) Herrero-Galán, E.; Fuentes-Perez, M. E.; Carrasco, C.; Valpuesta, J. M.; Carrascosa, J. L.; Moreno-Herrero, F.; Arias-Gonzalez, J. R. Mechanical identities of RNA and DNA double helices unveiled at the single-molecule level. *J. Am. Chem. Soc.* **2013**, *135*, 122–131.

- (20) Lipfert, J.; Skinner, G. M.; Keegstra, J. M.; Hensgens, T.; Jager, T.; Dulin, D.; Köber, M.; Yu, Z.; Donkers, S. P.; Chou, F.-C.; et al. Double-stranded RNA under force and torque: similarities to and striking differences from double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 15408–15413.

- (21) Marín-González, A.; Pastrana, C. L.; Bocanegra, R.; Martín-González, A.; Vilhena, J.; Pérez, R.; Ibarra, B.; Aicart-Ramos, C.; Moreno-Herrero, F. Understanding the paradoxical mechanical response of in-phase A-tracts at different force regimes. *Nucleic Acids Res.* **2020**, *48*, 5024–5036.

- (22) Bryant, Z.; Stone, M. D.; Gore, J.; Smith, S. B.; Cozzarelli, N. R.; Bustamante, C. Structural transitions and elasticity from torque measurements on DNA. *Nature* **2003**, *424*, 338–341.

- (23) Lipfert, J.; Kerssemakers, J. W.; Jager, T.; Dekker, N. H. Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. *Nat. Methods* **2010**, *7*, 977–980.

- (24) Lipfert, J.; Wiggin, M.; Kerssemakers, J. W.; Pedaci, F.; Dekker, N. H. Freely orbiting magnetic tweezers to directly monitor changes in the twist of nucleic acids. *Nat. Commun.* **2011**, *2*, 1–10.

- (25) Kriegel, F.; Ermann, N.; Forbes, R.; Dulin, D.; Dekker, N. H.; Lipfert, J. Probing the salt dependence of the torsional stiffness of

- DNA by multiplexed magnetic torque tweezers. *Nucleic Acids Res.* **2017**, *45*, 5920–5929.
- (26) Vanderlinden, W.; Kolbeck, P. J.; Kriegel, F.; Walker, P. U.; Lipfert, J. A benchmark data set for the mechanical properties of double-stranded DNA and RNA under torsional constraint. *Data in Brief* **2020**, *30*, 105404.
- (27) Mosconi, F.; Allemand, J. F.; Bensimon, D.; Croquette, V. Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers. *Phys. Rev. Lett.* **2009**, *102*, 078301.
- (28) Geggier, S.; Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15421–15426.
- (29) Guilbaud, S.; Salomé, L.; Destainville, N.; Manghi, M.; Tardin, C. Dependence of DNA persistence length on ionic strength and ion type. *Phys. Rev. Lett.* **2019**, *122*, 028102.
- (30) Heenan, P. R.; Perkins, T. T. Imaging DNA equilibrated onto mica in liquid using biochemically relevant deposition conditions. *ACS Nano* **2019**, *13*, 4220–4229.
- (31) Gore, J.; Bryant, Z.; Nöllmann, M.; Le, M. U.; Cozzarelli, N. R.; Bustamante, C. DNA overwinds when stretched. *Nature* **2006**, *442*, 836–839.
- (32) Gross, P.; Laurens, N.; Oddershede, L. B.; Bockelmann, U.; Peterman, E. J.; Wuite, G. J. Quantifying how DNA stretches, melts and changes twist under tension. *Nat. Phys.* **2011**, *7*, 731–736.
- (33) Nomidis, S. K.; Kriegel, F.; Vanderlinden, W.; Lipfert, J.; Carlon, E. Twist-bend coupling and the torsional response of double-stranded DNA. *Phys. Rev. Lett.* **2017**, *118*, 217801.
- (34) Gao, X.; Hong, Y.; Ye, F.; Inman, J. T.; Wang, M. D. Torsional stiffness of extended and plectonemic DNA. *Phys. Rev. Lett.* **2021**, *127*, 028101.
- (35) Lionnet, T.; Joubaud, S.; Lavery, R.; Bensimon, D.; Croquette, V. Wringing out DNA. *Phys. Rev. Lett.* **2006**, *96*, 178102.
- (36) 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.
- (37) 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.
- (38) Vafabakhsh, R.; Ha, T. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* **2012**, *337*, 1097–1101.
- (39) Le, T. T.; Kim, H. D. Probing the elastic limit of DNA bending. *Nucleic Acids Res.* **2014**, *42*, 10786–10794.
- (40) Jeong, J.; Kim, H. D. Determinants of cyclization-decyclization kinetics of short DNA with sticky ends. *Nucleic Acids Res.* **2020**, *48*, 5147–5156.
- (41) Basu, A.; Bobrovnikov, D. G.; Qureshi, Z.; Kaykicioglu, T.; Ngo, T. T.; Ranjan, A.; Eustermann, S.; Cieza, B.; Morgan, M. T.; Hejna, M.; Rube, H. T.; Hopfner, K.-P.; Wolberger, C.; Song, J. S.; Ha, T. Measuring DNA mechanics on the genome scale. *Nature* **2021**, *589*, 462–467.
- (42) Becker, N. B.; Everaers, R. Comment on “Remeasuring the double helix. *Science* **2009**, *325*, 538–538.
- (43) Mathew-Fenn, R. S.; Das, R.; Fenn, T. D.; Schneiders, M.; Harbury, P. A. Response to comment on “Remeasuring the double helix. *Science* **2009**, *325*, 538–538.
- (44) Harrison, R. M.; Romano, F.; Ouldridge, T. E.; Louis, A. A.; Doye, J. P. Identifying physical causes of apparent enhanced cyclization of short DNA molecules with a coarse-grained model. *J. Chem. Theory Comput.* **2019**, *15*, 4660–4672.
- (45) Pasi, M.; Lavery, R. Structure and dynamics of DNA loops on nucleosomes studied with atomistic, microsecond-scale molecular dynamics. *Nucleic Acids Res.* **2016**, *44*, 5450–5456.
- (46) 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.
- (47) Bao, L.; Zhang, X.; Shi, Y.-Z.; Wu, Y.-Y.; Tan, Z.-J. Understanding the relative flexibility of RNA and DNA duplexes: stretching and twist-stretch coupling. *Biophys. J.* **2017**, *112*, 1094–1104.
- (48) Kriegel, F.; Matek, C.; Dršata, T.; Kulenkampff, K.; Tschirpke, S.; Zacharias, M.; Lankaš, F.; Lipfert, J. The temperature dependence of the helical twist of DNA. *Nucleic Acids Res.* **2018**, *46*, 7998–8009.
- (49) 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*, 19254–19266.
- (50) Pasi, M.; Maddocks, J. H.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., III; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. μ ABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 12272–12283.
- (51) Zgarbová, M.; Jurecka, P.; Lankas, F.; Cheatham, T. E., III; Spöner, J.; Otyepka, M. Influence of BII backbone substates on DNA twist: a unified view and comparison of simulation and experiment for all 136 distinct tetranucleotide sequences. *J. Chem. Inf. Model.* **2017**, *57*, 275–287.
- (52) Marin-Gonzalez, A.; Vilhena, J.; Moreno-Herrero, F.; Perez, R. DNA crookedness regulates DNA mechanical properties at short length scales. *Phys. Rev. Lett.* **2019**, *122*, 048102.
- (53) Shepherd, J. W.; Greenall, R. J.; Probert, M. I. J.; Noy, A.; Leake, M. C. The emergence of sequence-dependent structural motifs in stretched, torsionally constrained DNA. *Nucleic Acids Res.* **2020**, *48*, 1748–1763.
- (54) Petkevičiūtė, D.; Pasi, M.; Gonzalez, O.; Maddocks, J. cgDNA: a software package for the prediction of sequence-dependent coarse-grain free energies of B-form DNA. *Nucleic Acids Res.* **2014**, *42*, e153–e153.
- (55) 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.
- (56) Mitchell, J. S.; Glowacki, J.; Grandchamp, A. E.; Manning, R. S.; Maddocks, J. H. Sequence-dependent persistence lengths of DNA. *J. Chem. Theory Comput.* **2017**, *13*, 1539–1555.
- (57) Knotts, T. A., IV; Rathore, N.; Schwartz, D. C.; De Pablo, J. J. A coarse grain model for DNA. *J. Chem. Phys.* **2007**, *126*, 084901.
- (58) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. DNA nanotweezers studied with a coarse-grained model of DNA. *Phys. Rev. Lett.* **2010**, *104*, 178101.
- (59) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. *J. Chem. Phys.* **2010**, *132*, 035105.
- (60) Savelyev, A.; Papoian, G. A. Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20340–20345.
- (61) Cragnolini, T.; Derreumaux, P.; Pasquali, S. Coarse-grained simulations of RNA and DNA duplexes. *J. Phys. Chem. B* **2013**, *117*, 8047–8060.
- (62) Korolev, N.; Luo, D.; Lyubartsev, A. P.; Nordenskiöld, L. A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo. *Polymers* **2014**, *6*, 1655–1675.
- (63) Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. DNA duplex formation with a coarse-grained model. *J. Chem. Theory Comput.* **2014**, *10*, 5020–5035.
- (64) Uusitalo, J. J.; Ingólfsson, H. I.; Akhshi, P.; Tieleman, D. P.; Marrink, S. J. Martini coarse-grained force field: extension to DNA. *J. Chem. Theory Comput.* **2015**, *11*, 3932–3945.
- (65) Chakraborty, D.; Hori, N.; Thirumalai, D. Sequence-dependent three interaction site model for single- and double-stranded DNA. *J. Chem. Theory Comput.* **2018**, *14*, 3763–3779.
- (66) Snodin, B. E.; Schreck, J. S.; Romano, F.; Louis, A. A.; Doye, J. P. Coarse-grained modelling of the structural properties of DNA origami. *Nucleic Acids Res.* **2019**, *47*, 1585–1597.

- (67) Brandani, G. B.; Niina, T.; Tan, C.; Takada, S. DNA sliding in nucleosomes via twist defect propagation revealed by molecular simulations. *Nucleic Acids Res.* **2018**, *46*, 2788–2801.
- (68) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, 08S101.
- (69) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.
- (70) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (71) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; De Pablo, J. J. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **2013**, *139*, 144903.
- (72) Olson, W. K.; Manning, G. S. A configurational interpretation of the axial phosphate spacing in polynucleotide helices and random coils. *Biopolymers* **1976**, *15*, 2391–2405.
- (73) Assenza, S.; Mezzenga, R. Soft condensed matter physics of foods and macronutrients. *Nat. Rev. Phys.* **2019**, *1*, 551–566.
- (74) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H. e. a. *AMBER 14*; University of California Press, 2014.
- (75) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (76) Pérez, A.; Marchán, I.; Svozil, D.; Spöner, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* **2007**, *92*, 3817–3829.
- (77) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Spöner, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E., III Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127.
- (78) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walthert, J.; Andrić, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55–58.
- (79) Zgarbová, M.; Spöner, J.; Otyepka, M.; Cheatham, T. E., III; Galindo-Murillo, R.; Jurecka, P. Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.
- (80) Snodin, B. E.; Randisi, F.; Mosayebi, M.; Šulc, P.; Schreck, J. S.; Romano, F.; Ouldridge, T. E.; Tsukanov, R.; Nir, E.; Louis, A. A.; Doye, J. P. Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.* **2015**, *142*, 234901.
- (81) Henrich, O.; Gutierrez-Fosado, Y. A.; Curk, T.; Ouldridge, T. E. Coarse-grained simulation of DNA using LAMMPS. *Eur. Phys. J. E* **2018**, *41*, 57.
- (82) Freeman, G. S.; Hinckley, D. M.; Lequieu, J. P.; Whitmer, J. K.; De Pablo, J. J. Coarse-grained modeling of DNA curvature. *J. Chem. Phys.* **2014**, *141*, 165103.
- (83) Sassi, A. S.; Assenza, S.; De Los Rios, P. Shape of a stretched polymer. *Phys. Rev. Lett.* **2017**, *119*, 037801.
- (84) Lu, X.-J.; Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
- (85) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917–5929.
- (86) Yuksel, C.; Schaefer, S.; Keyser, J. Parameterization and applications of Catmull-Rom curves. *Comput.-Aided Des.* **2011**, *43*, 747–755.
- (87) Trifonov, E. N.; Tan, R. K.-Z.; Harvey, S. Static persistence length of DNA. *Structure and Expression: Proceedings of the Fifth Conversation in the Discipline Biomolecular Stereodynamics*, State University of New York at Albany, June 2–6, 1987; Sarma, M. H., Sarma, R. H., Eds.; 1988.
- (88) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Elsevier: Oxford, U.K., 2001; Vol. 1.
- (89) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **1976**, *32*, 922–923.
- (90) Rubinstein, M.; Colby, R. H.; et al. *Polymer physics*; Oxford University Press: New York, 2003; Vol. 23.
- (91) Vologodskaya, M.; Vologodskii, A. Contribution of the intrinsic curvature to measured DNA persistence length. *J. Mol. Biol.* **2002**, *317*, 205–213.
- (92) Moroz, J. D.; Nelson, P. Entropic elasticity of twist-storing polymers. *Macromolecules* **1998**, *31*, 6333–6347.
- (93) Sheinin, M. Y.; Wang, M. D. Twist-stretch coupling and phase transition during DNA supercoiling. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4800–4803.
- (94) Noy, A.; Golestanian, R. Length scale dependence of DNA mechanical properties. *Phys. Rev. Lett.* **2012**, *109*, 228101.
- (95) Skoruppa, E.; Nomidis, S. K.; Marko, J. F.; Carlon, E. Bend-induced twist waves and the structure of nucleosomal DNA. *Phys. Rev. Lett.* **2018**, *121*, 088101.
- (96) Nomidis, S. K.; Skoruppa, E.; Carlon, E.; Marko, J. F. Twist-bend coupling and the statistical mechanics of the twistable wormlike-chain model of DNA: Perturbation theory and beyond. *Phys. Rev. E* **2019**, *99*, 032414.
- (97) Grosberg, A. Y.; Nguyen, T.; Shklovskii, B. Colloquium: The physics of charge inversion in chemical and biological systems. *Rev. Mod. Phys.* **2002**, *74*, 329.
- (98) Freeman, G. S.; Lequieu, J. P.; Hinckley, D. M.; Whitmer, J. K.; de Pablo, J. J. DNA shape dominates sequence affinity in nucleosome formation. *Phys. Rev. Lett.* **2014**, *113*, 168101.