Molecular Origins of DNA Flexibility: Sequence Effects on Conformational and Mechanical Properties

Vanessa Ortiz and Juan J. de Pablo*

University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

(Received 22 September 2010; published 10 June 2011)

A central question in biophysics is whether DNA sequence affects its mechanical properties, which are thought to influence nucleosome positioning and gene expression. Previous attempts to answer this question have been hindered by an inability to resolve DNA structure and dynamics at the base-pair level. Here we use a model to measure the effects of sequence on the stability of DNA under bending. Sequence is shown to influence DNA's flexibility and its ability to form kinks, which arise when certain motifs slide past others to form non-native contacts. A mechanism for nucleosome positioning is proposed in which sequence influences DNA-histone binding by altering the local base-pair-level structure when subject to the curvature necessary for binding.

DOI: 10.1103/PhysRevLett.106.238107 PACS numbers: 87.15.−v, 87.14.gk

Uncovering and understanding regulatory mechanisms for gene expression represents a grand challenge for biophysics and epigenetics. One of these mechanisms involves controlling the way in which genes are packed in the chromosome, rendering some regions more accessible for expression than others. In the chromosome, the basic packing unit is the nucleosome [Fig. 1(b)(i)]: a 147 base-pair (bp) long DNA segment wrapped around a protein octamer (histones). Any DNA segment can form nucleosomes *in vitro*, but the relative affinity of histones for different DNA sequences can differ by as much as 1000-fold (> 4 kcal/mol in free energy) [1]. Through analysis of nucleosome-bound DNA sequences, Segal *et al.* [2] discovered that AA-TT-TA bp steps are more likely to be found where the minor groove of DNA is at the DNA-histone interface. Crystal structures of nucleosomes [3] show no contacts between DNA bases and histones, leading to the proposal that sequence could influence nucleosome positioning indirectly by facilitating the tight bending required to wrap DNA around histones. Sequence could affect binding by either creating a kink (an abrupt and pronounced change in curvature) in a DNA segment (altering *bendedness*) or changing the amount of energy required to bend it (altering *bendability*). The main question addressed here is whether DNA's mechanical properties, at small length scales, depend on sequence, as nucleosome positioning data suggest. Experimental studies have provided some understanding of this issue [4–11], but attempts to distinguish bendedness from bendability have met with limited success [1]. While fully atomistic models are available for DNA [12], the computational demands associated with atomistic simulations of long strands preclude a systematic study of bendedness and bendability, and theoretical treatments have been scarce, largely because models capable of describing the mechanical properties of double-stranded DNA beyond the elastic limit and its ability to dehybridize locally on a sequence-dependent manner were not available until recently [13].

In this study we present a coarse-grained (CG) model of DNA [Fig. 1(a)] that reproduces experimental melting temperatures, their dependence on chain length, GC content, and ionic strength [14], describes hybridization [15], and captures local mechanical properties of the molecule (bp-step deformability). Langevin dynamics simulations are employed to examine the macroscopic mechanical properties of DNA as a function of sequence. Since the model allows for dehybridization, it enables examination of the effects of sequence on the overall stability of the DNA double helix when subject to bending constraints. Omitting the histones from the simulated systems allows

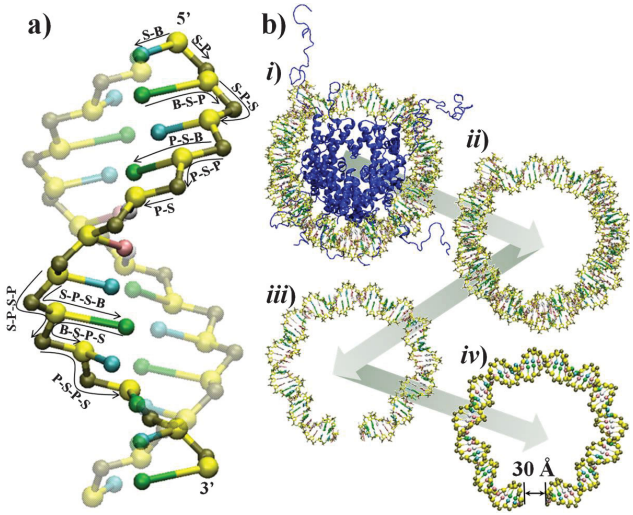

FIG. 1 (color). DNA model. (a) Each nucleotide is described by three interaction sites: S, sugar (yellow); P, phosphate (tan); and B, base (C, cyan; G, green; T, pink; A, white). Arrows define bonds, bends, and torsions in the model, defined in the order encountered when moving along the molecule from 5′ to 3′ (arrow direction). (b) DNA system studied. (i) Crystal structure of the nucleosome core particle [3], (ii) segment of DNA after removal of the histones (147 bp), (iii) central 73 bp, and (iv) CG version of the segment.

one to discern whether the sharp bends observed in crystal structures of nucleosomes are a cause or an effect of the formation of the complex. All four segments considered here (Nuc, A-tract, Flex, and Stiff) consist of 73 bp, the length required to form one full turn around the histones [Fig. 1(b)(iii)]. In experiments, the Nuc and A-tract sequences were found to exhibit high [3] and low [16] affinity for the histones, respectively. Flex and Stiff sequences were engineered on the basis of our model for high and low flexibility, respectively. Details on the model and sequences are provided as supplemental material [17].

The potential of mean force, or free energy (ΔG), associated with loop formation (work required to bring the molecule's ends together) was measured in umbrella-sampling simulations. The resulting ΔG profiles (supplemental Fig. 1 [17]) are barely different from one sequence to another, indicating that at length scales of 73 bp the average mechanical properties of the molecules are similar. This is consistent with experimental findings of Widom and co-workers [2], who report differences in sequence probabilities in the nucleosome of up to 0.1, translating into ΔG differences of up to $0.5 k_B T$. Examination of smaller length scales, however, reveals profound distinctions. As done in the experimental analysis of DNA contours in Ref. [18], the probability distribution function $P(\theta, \Delta N_{\mathrm{bp}})$ of deflection angle θ between tangents to the double helix was calculated at points separated by ΔN_{bp}. Figure 2 shows the negative logarithm of $P(\theta, \Delta N_{\mathrm{bp}})$, a measure of the bending energy $E(\theta)/k_B T$ of a DNA segment at length scales defined by ΔN_{bp}. For $\Delta N_{\mathrm{bp}} = 7$ bp, the Flex and A-tract sequences exhibit a proclivity to form tight bends ($\theta > 1.2$ rad), whereas Nuc and Stiff do not. However, as ΔN_{bp} is increased—the resolution of the analysis becomes coarser—the differences between sequences disappear.

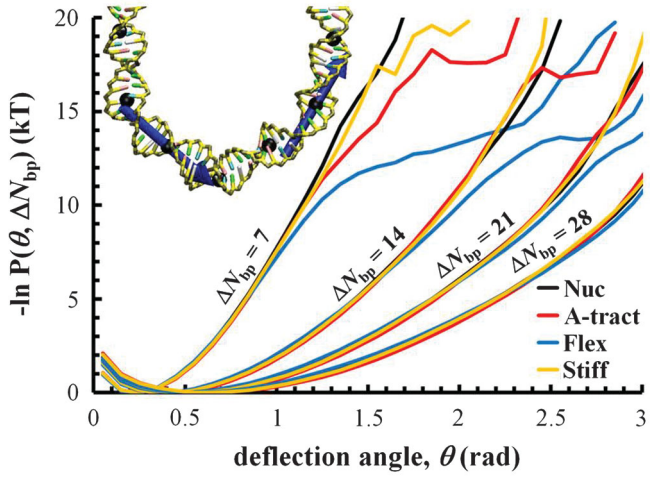

FIG. 2 (color). Microscopic flexibility. Negative logarithm of the observed probability distribution function $P(\theta, \Delta N_{\mathrm{bp}})$ of angles between tangents separated by ΔN_{bp} base pairs. The inset shows how the contour of the molecule is divided into 7 bp segments (marked by black spheres); two tangents are drawn, which are separated by $\Delta N_{\mathrm{bp}} = 21$.

This result suggests that a boundary exists between length scales at which DNA sequence affects its mechanical properties—~3 turns—and length scales over which sequence effects are homogenized. This analysis was also performed for $\Delta N_{\mathrm{bp}} = 5$ (supplemental Fig. 2 [17]), and the same results were obtained. On one hand, at a length scale of 7 bp (~ 2.5 nm), A-tract and Flex appear more flexible in Fig. 2. On the other hand, equilibrium end-to-end distance values show that A-tract is the shortest ($R_{\mathrm{min}} = 243$ Å) and Flex is the longest ($R_{\mathrm{min}} = 251$ Å). These two observations can be explained by the argument that sequence can affect both bendedness, by forming kinks (A-tract case), and bendability, by making it more flexible (Flex case). The DNA model presented here can identify and distinguish these two effects from each other.

If A-tract forms kinks, where in the molecule do they arise? Are they distributed uniformly, or do they appear in specific regions of a segment? To answer this, the distribution $P(\theta, \Delta N_{\mathrm{bp}} = 7)$ of Fig. 2 was deconstructed into contributions from individual bps (see [17]), providing the probability of observing a kink at a specific position along a sequence, as shown in Fig. 3. The labels on the figure show that regions with a high probability of forming kinks consist of repeated base pairs, and the magnitude of the corresponding peak correlates with the length of the repeated base-pair segment. Longer repeated segments (such as AAAAAAAAA) have a higher probability of forming a kink, but even relatively short repetitive stretches (e.g., CC) can exhibit distinct kinks (albeit less frequently).

Insights into local effects of sequence are provided by quantifying the hybridization of the molecule. Figure 4 shows profiles of the fraction of time that a base spends in one of three states: (i) hybridized to its native base (f_{nat}), (ii) hybridized to a non-native neighboring complementary base ($f_{\mathrm{non\text{-}nat}}$), or (iii) dehybridized (f_{free}). These fractions add up to unity ($f_{\mathrm{nat}} + f_{\mathrm{non\text{-}nat}} + f_{\mathrm{free}} = 1$). Figure 4 shows that consecutively repeated regions, such as A-tracts,

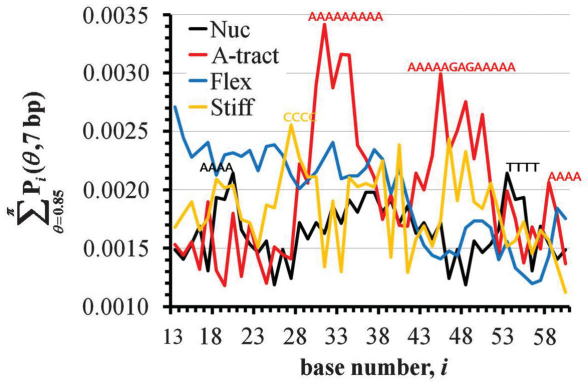

FIG. 3 (color). The integral of $P_i(\theta, 7\ \mathrm{bp})$ for $0.85 \leq \theta \leq \pi$, evaluated at each bp position along the sequence, measures the probability of observing high deflection angles (kinks) at that position. The labels show the specific motifs (e.g., consecutively repeated A bases) that are responsible for the largest peaks in the curves.

exhibit high $f_{\text{non-nat}}$ levels. This sustained non-native "hybridization" in repeated sequences is the likely cause for the formation of the kinks observed in Fig. 3. Because of its lack of repeated domains, the Flex molecule shows almost no $f_{\text{non-nat}}$. The Nuc segment does exhibit some non-native hybridization, but regions of high $f_{\text{non-nat}}$ are small and distributed uniformly throughout the molecule. Bendability also appears to correlate with $\langle f_{\text{non-nat}} \rangle$. The bending ΔG (from supplemental Fig. 1 [17]) are arranged from lower to higher magnitude in the same order as that observed for increasing $\langle f_{\text{non-nat}} \rangle$ in Fig. 4 (Flex → Nuc → Stiff → A-tract). This suggests that sustained non-native hybridization, particularly pronounced in the A-tract segment, is responsible for both the formation of kinks and changes in flexibility.

Our discussion so far has relied on results from a three-site-per-nucleotide model. To explore the atomistic origins of the above observations, we also performed extensive molecular dynamics simulations of an all-atom (AA) representation of the Nuc sequence in explicit water, with the ends constrained at a distance of 30 Å [distance required for nucleosome binding; Fig. 1(b)(iv)]. We note that the AA model was not parameterized to provide the melting behavior of DNA at experimentally observed temperatures [19], and one cannot ascertain whether the extent of non-native hybridization that it predicts is in fact accurate at the temperature considered in this work. Results included in supplemental Fig. 3 [17] suggest that the AA model is overly stable. The CG model was parameterized to provide accurate melting temperatures and heat capacities over a wide range of ionic strength, which at any given temperature quantify the fluctuations of the energy, including contributions arising from hybridization and electrostatic interactions. Given such issues, we simply seek to determine whether AA and CG simulations exhibit a similar, qualitative behavior with regards to non-native hybridization. Figure 5(a) provides a representative configuration from the simulation showing native and non-native hydrogen bonds. From a 10 ns trajectory (0.5% of a CG simulation), one can infer the propensity of a base to form a non-native hydrogen bond ($P_{\text{non-nat}}$); see [17]. In Fig. 5(b), C and G bases exhibit a higher $P_{\text{non-nat}}$ than A and T, and all bases exhibit a higher propensity for non-native hydrogen bonds in the 3′ direction than in the 5′ direction [schematic drawing under Fig. 5(b)]. $P_{\text{non-nat}}$ determined from the CG model [FIG. 5(d)] are in agreement with those from the AA model, but the ratio of 3′-to-5′ non-native hybridization is larger in the AA model. The CG simulations, which one is able to perform over much longer time intervals, also reveal some -2 and $+2$ non-native hybridization events, as shown in Fig. 5(c). Finally, when measuring the distance between bases when forming non-native contacts, one finds that the average contact distance in the 3′ direction is higher (~ 4.5 Å) than in the 5′ direction (~ 3.3 Å). This observation helps explain why the 3′ direction is preferred over the 5′ direction; it requires the helical structure to deform less.

A-tracts have generated controversy in past discussions of nucleosome positioning; tracts of five or more nucleotides were recently shown to be excluded from nucleosomes [16], but, prior to that work, they were thought to be stabilizing in nucleosome binding because of their high tendency to form bends in solution [11]. Figure 3 offers an explanation for these experimental observations:

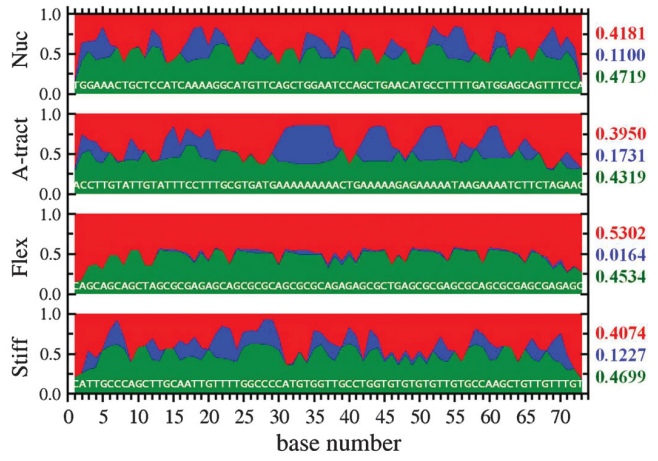

FIG. 4 (color). Hybridization profiles. Fraction of total simulation time during which a base was (i) hybridized to its native complementary base (f_{nat}, green), (ii) hybridized to a non-native complementary base ($f_{\text{non-nat}}$, blue), and (iii) dehybridized (f_{free}, red). Average values over all bases are shown on the right. Segment sequences are shown alongside the data.

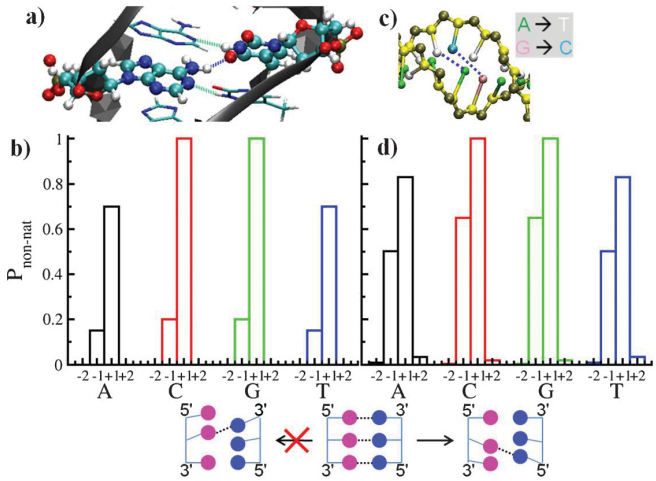

FIG. 5 (color). Analysis of non-native contacts in Nuc segment at AA (a),(b) and CG (c),(d) levels. (a) Representative configuration from AA simulation showing native (green dashed line) and non-native (blue dashed line) hydrogen bonding. (b), (d) Propensity of a base to form a non-native contact ($P_{\text{non-nat}}$) as a function of base identity and bond direction. Positive ($+1$, $+2$) and negative (-1, -2) values correspond to non-native contacts in the 3′ and 5′ directions, respectively. (c) Configuration from CG simulation showing non-native contacts. Dashed lines identify the native pairs: A (green) is "hybridized" to a non-native T (white) in the $+2$ direction, the contact between the G (pink) and C (cyan) is broken, and they remain dehybridized.

A-tracts longer than 5 bp enable the formation of kinks, but these are highly localized and do not occur at regular intervals throughout the chain. The molecule therefore behaves as a collection of rods and hinges which would not have the uniformity required to accommodate a continuous, regular wrapping around the histone. The experimental observation [11] that A-tracts exhibit a tighter bend at the $3'$ end than at the $5'$ end is consistent with our findings in Fig. 5; a preference exists for the direction in which bases form non-native pairs. The sequence GGGCCC has also been shown experimentally to exhibit a net bend [20,21]. The fact that this is a repeated sequence supports our result that the formation of kinks is tightly correlated to the ability of the segment to form non-native contacts. In our simulations, a bending constraint is imposed and the DNA double helix is free to adopt any twist that allows it to reach the lowest energy possible (under the constraint). However, AA studies in which a twist is imposed by closing the DNA loop (nanocircle) have also reported kinking and denaturation [22,23]. The AA description makes it computationally expensive to explore sequence space in these systems, and past studies could not narrow down the relationship between sequence, kinking, and denaturation that has been identified here. A study that extended the classical elastic rod DNA model to include the possibility of denaturation [13] also observed denaturation in small nanocircles. That study looked at loop-length effects instead of sequence, and given a simpler description of denaturation (each bp denatured in an independent manner without taking into account couplings between neighbors) observations of sequence effects with that model would be necessarily different from those reported here, particularly for repeated tracks.

Sequence is thought to play an important role in nucleosome stability and dynamics by influencing the conformational and mechanical properties of a DNA segment in contact with the histones. The results presented in this Letter have uncovered a number of plausible sources of that sequence dependence. Free-energy calculations of unprecedented accuracy on a model of DNA that accounts for both sequence-dependent bp-step deformability and base pairing have revealed that the ΔG associated with the formation of a 73 bp loop shows a small but statistically significant dependence on the sequence of the molecule. Sequence affects the flexibility of DNA at a length scale smaller than $\sim$3 DNA turns. These effects arise from the formation of kinks (modification of bendedness) or from an increase in the segment's ability to bend (modification of bendability), as seen in the A-tract and Flex sequences, respectively. Local melting plays a major role on the molecule's ability to bend; kinks arise in a repeated sequence when strands slide past each other and the bases make non-native contacts with neighboring complementary bases. This phenomenon was corroborated through detailed AA simulations and was shown to correlate with the length of the repeated segment and with the identity of the bases involved. Our results serve to establish that DNA exhibits a sequence-dependent preference for bending, independent of the presence of a protein. When considering these and experimental results, the picture that emerges for a description of the DNA-histone binding process is one in which all sequences have similar probabilities of forming the required bends, but differences that arise at the local bp level enable molecules to adopt conformations that are highly dependent on the sequence. Such local mechanical effects, which partially rely on local dehybridization, are encoded in the sequence and, as proposed by Widom and co-workers, in all likelihood affect the ability of the molecule to interact with and wrap around the histones.

Useful discussions with E. J. Sambriski, M. Hoefert, and G. S. Freeman are acknowledged. V. O. was supported by an NLM grant to the University of Wisconsin CIBM Training Program (NLM T15LM007359). J. J. d. P. is grateful to the National Science Foundation for financial support.

*depablo@engr.wisc.edu

[1] J. Widom, Q. Rev. Biophys. **34**, 269 (2001).
[2] E. Segal *et al.*, Nature (London) **442**, 772 (2006).
[3] C. A. Davey *et al.*, J. Mol. Biol. **319**, 1097 (2002).
[4] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., *Nucleic Acids, Structures, Properties, and Functions* (University Science, Sausalito, CA, 2000).
[5] W. H. Taylor and P. J. Hagerman, J. Mol. Biol. **212**, 363 (1990).
[6] J. D. Kahn, E. Yun, and D. M. Crothers, Nature (London) **368**, 163 (1994).
[7] A. Bacolla *et al.*, J. Biol. Chem. **272**, 16 783 (1997).
[8] E. Vacano and P. J. Hagerman, Biophys. J. **73**, 306 (1997).
[9] J. M. Schurr, B. S. Fujimoto, P. Wu, and L. Song, in *Topics in Fluorescence Spectroscopy*, edited by J. R. Lakowicz (Plenum, New York, 1992), Vol. 3, p. 390.
[10] T. M. Okonogi *et al.*, Biophys. J. **78**, 2560 (2000).
[11] D. MacDonald *et al.*, J. Mol. Biol. **306**, 1081 (2001).
[12] R. Lavery *et al.*, Nucleic Acids Res. **38**, 299 (2010).
[13] T. B. Liverpool, S. A. Harris, and C. A. Laughton, Phys. Rev. Lett. **100**, 238103 (2008).
[14] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, Biophys. J. **96**, 1675 (2009).
[15] E. J. Sambriski, V. Ortiz, and J. J. de Pablo, J. Phys. Condens. Matter **21**, 034105 (2009).
[16] G.-C. Yuan *et al.*, Science **309**, 626 (2005).
[17] See supplemental material at http://link.aps.org/supplemental/10.1103/PhysRevLett.106.238107 for a description of the DNA model, sequences used, methods for data analysis, and supporting information.
[18] P. A. Wiggins *et al.*, Nature Nanotech. **1**, 137 (2006).
[19] W. D. Cornell *et al.*, J. Am. Chem. Soc. **117**, 5179 (1995).
[20] I. Brukner *et al.*, Nucleic Acids Res. **19**, 3549 (1991).
[21] I. Brukner *et al.*, Nucleic Acids Res. **21**, 1025 (1993).
[22] F. Lankaš, R. Lavery, and J. H. Maddocks, Structure **14**, 1527 (2006).
[23] S. A. Harris, C. A. Laughton, and T. B. Liverpool, Nucleic Acids Res. **36**, 21 (2008).