

Breathing, bubbling, and bending: DNA flexibility from multimicrosecond simulations

Ari Zeida, Matías Rodrigo Machado, Pablo Daniel Dans, and Sergio Pantano*

Institut Pasteur de Montevideo, Calle Mataojo 2020, Montevideo, Codigo Postal 11400, Uruguay

(Received 4 July 2011; revised manuscript received 1 June 2012; published 3 August 2012)

Bending of the seemingly stiff DNA double helix is a fundamental physical process for any living organism. Specialized proteins recognize DNA inducing and stabilizing sharp curvatures of the double helix. However, experimental evidence suggests a high protein-independent flexibility of DNA. On the basis of coarse-grained simulations, we propose that DNA experiences thermally induced kinks associated with the spontaneous formation of internal bubbles. Comparison of the protein-induced DNA curvature calculated from the Protein Data Bank with that sampled by our simulations suggests that thermally induced distortions can account for $\sim 80\%$ of the DNA curvature present in experimentally solved structures.

DOI: [10.1103/PhysRevE.86.021903](https://doi.org/10.1103/PhysRevE.86.021903)

PACS number(s): 87.14.gk, 87.15.A–, 87.15.H–

I. INTRODUCTION

The double-stranded DNA (dsDNA) polymer presents a very stable structure with persistence lengths on the order of 50 nm or ~ 150 consecutive base pairs (bp). This stiffness results from the sum of relatively small energetic contributions (below $4k_B T$) [1], which come from the pairing and stacking of a large number of single nucleotides. Thermal excitation leads to the sporadic breaking of a single bp, giving place to the so-called “breathing” motion in which opening and closing can be detected by nuclear magnetic resonance (NMR) spectroscopy in the time range from picoseconds to nanoseconds [2]. The simultaneous opening of two or more base pairs is referred to as a “bubble”. This kind of energetically more expensive event is, consequently, less frequent. Under room conditions, bubbles from 2 to 10 bp with characteristic closing times of several microseconds have been deduced from fluorescence quenching experiments [3]. This highly dynamic behavior, spanning about 8 orders of magnitude in time, challenges theoretical methods as molecular dynamics (MD) or wormlike polymer models to properly describe the flexibility of DNA. In this regard, coarse-grained (CG) models offer a valid alternative to achieve an exhaustive exploration of the conformational space of macromolecules with a significantly reduced computational effort, yet capturing the essential physics at play. Recent examples of CG models applied to the study of nucleic acids include the work of Savelyev and Papoian, who developed an accurate scheme using one effective bead per nucleotide. This model, which explicitly included ionic strength, allowed for the characterization of structural transitions and other DNA properties [4]; Ortiz and de Pablo [5] measured the effects of sequence on the stability of DNA under bending conditions via umbrella sampling calculations using a three beads per bp model [6]. At a higher level of abstraction, Alexandrov *et al.* [7] used a unidimensional nonlinear Langevin dynamics model to examine the breathing dynamics around the transcription start site of an engineered DNA segment, which works as a highly efficient transcriptional promoter called super core promoter 1 (SCP1) [8]. They inferred a relationship between intrinsic DNA flexibility and transcriptional activity.

Aimed at providing new insight on the flexibility of double-stranded DNA, we conducted a series of MD studies on DNA filaments at the CG level using the model reported by Dans *et al.* [9] (Fig. 1). Our scheme is somewhat reminiscent of the early model of Zhang and Collins [10]. Nucleotides are represented by six effective beads, each of them placed on the Cartesian coordinates of one element in the all-atom representation. These beads are characterized by particular masses, partial charges, bonding, and van der Waals parameters, so as to condense the molecular information from its atomic neighborhood in one effective interaction site. This translates into a reduction of nearly 80% in the number of particles constituting a given molecular system. In addition to predict melting temperatures under different conditions and hydrogen bonding features, our CG scheme accurately reproduces the structure and dynamics of dsDNA. Among many other accurate CG models of DNA, a distinctive feature of our CG scheme is the possibility to furnish fully atomistic structures from the CG trajectories thanks to its complete backmapping capabilities [9].

In this paper, we explore the unbiased dynamic behavior of single-stranded DNA (ssDNA) and double-stranded DNA in a multimicrosecond time scale through a series of simulations. Breathing motions, bubbles, kinks formation, and bending of DNA are studied for experimentally characterized molecular systems of two different sequences and lengths. Finally, we also compare our results with DNA curvature calculated from protein-dsDNA complexes in the Protein Data Bank (PDB).

II. METHODS

Our CG mapping scheme uses effective beads, which are placed on the positions of real atoms (Fig. 1). This offers the possibility to recover pseudoatomistic information from a CG trajectory by using internal coordinates defined for each atomistic nucleotide. Reconstruction of atomistic coordinates from CG configurations results in root mean square superposition on the order of 0.1 nm, which is well compared to the intrinsic variability observed during classical all-atoms simulations [9]. However, it is worth noting that our backmapping procedure tends to homogenize certain substates observed during the all-atom simulations. After backmapping, the ζ/ϵ torsions are always in the BI conformer, and α/γ torsions are reconstructed in the canonical $g - / g +$ distribution. Experimentally, nearly

*Corresponding author: spantano@pasteur.edu.uy

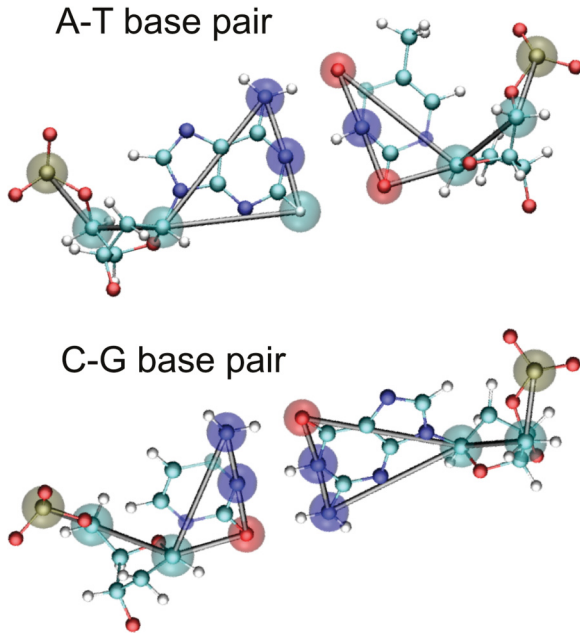


FIG. 1. (Color online) Coarse-grained model of DNA. Superposition of the CG and all-atom representations of the four nucleotides in a typical Watson-Crick interaction. Semitransparent spheres indicate the position of the atoms occupied by effective CG beads. Gray sticks indicate the connectivity between CG beads. See Refs. [9] and [11] for further details on the parameterization.

15% of the ζ/ϵ torsions are in the BII conformer, and some less frequent sequence-dependent shifts for the α/γ torsions are also observed. In our case, the sugar pucker is always reconstructed in the C2'-endo typical of the canonical B form. Nevertheless, as a result of the energy minimization performed as a final step of the backmapping procedure, we obtain 80% of C2'-endo conformations, whereas, the remaining conformers correspond mainly to C3'-exo.

To explore the dynamics of DNA, we choose different molecular systems. The first corresponds to the rationally designed SCP1, which sustains very high levels of transcription by RNA polymerase II [8]. The sequence of this dsDNA promoter is as follows:

5'-d(GCATGCCTGCAGGTACT**TATATAAGGGGGTGGGG**
GCGCGTTCTG_{T(C)}C_(G)CTCA₊₁GTCGCGA_(G)TCGAACA
_(G)CTCGAGCCGAGCAGACGTGCCTACGGACCGTCTAG
AGGATCC)-3'.

Bold letters indicate the AT-rich TATA box where the TATA binding protein binds. The underlined nucleotides correspond to the promoter region, whereas, the flanking sequences belong to the plasmid used in the experimental paper [8] and were included in the simulation to rule out possible end effects. The transcription starting site is indicated with a +1 subscript. Four mutations were introduced on the SCP1 sequence to study variations in the breathing profile as they reduce the transcription yield in *in vitro* experiments by 80% [7]. The subscripts between parentheses indicate the point mutations introduced in each of the precedent positions. The mutated SCP1 is named, hereafter, as mSCP1. MD simulations on systems SCP1 and mSCP1 were performed for 20 μ s. Loose

harmonic constraints of 3 kcal mol⁻¹ Å⁻² were used to mimic the continuation of the DNA within the plasmid.

Additionally, we also simulated a dsDNA segment analogous to the sequence called M18 in the paper by Altan-Bonnet *et al.* [3]. This corresponds to a dsDNA system with the following sequence:

5'-d(GGCGCCCAATATAAAATATTAAATGCGC)-3'.

Constraints as those used for SCP1 and mSCP1 were used only at the 3' end of the M18 dsDNA. This was intended to mimic the presence of a thymine tetraloop present in the experimental work, whose structure is unknown [3]. Five independent replicas of this system were simulated for 50 μ s each using different starting conformers. This is equivalent to a total sampling of 250 μ s.

Finally, two ssDNA, corresponding to the two separate filaments of M18, were simulated. Each of the two single-stranded filaments was simulated for 50 μ s, corresponding to a total of 100 μ s of sampling time for ssDNA.

All MD simulations were performed using AMBER 10 [12]. Temperature was controlled using a Langevin thermostat [13] with a collision frequency of 50 ps⁻¹ and a target temperature of 300 K. Calculations were carried out under the same conditions reported by us in Ref. [9] with the parameter's modification reported in Ref. [11], which allows for a time step of 20 ps. Electrostatic interactions were calculated using a cutoff of 1.8 nm within the framework of the generalized Born model for implicit solvation as implemented in AMBER. Within this approach, electrostatic screening of a monovalent salt at a concentration of 150 mM was included via the Debye-Huckel parameter [14].

III. RESULTS AND DISCUSSIONS

Aimed at studying the intrinsic flexibility of DNA, we first analyzed the dynamics of the SCP1. This is a rationally designed transcriptional promoter of RNA polymerase II. It is a very well characterized system where the first protein binding event corresponds to the interaction between the TATA binding protein and its cognate DNA target motif. This eventually generates a cascade of protein DNA binding events, which ultimately results in high levels of transcriptional activity [8]. Theoretical methods have pointed out how breathing profiles are directly correlated with transcriptional regulation in this system [7]. MD simulations may further contribute to this by furnishing structural and dynamic insights. Moreover, the wide time window needed for a proper description of DNA dynamics makes CG simulations an attractive alternative.

The breathing profile of the SCP1 system, presented in Fig. 2(a), suggests a good conservation of the Watson-Crick hydrogen bond pattern. Despite the high prevalence of the canonical dsDNA conformation, two different kinds of events are present with different time scales: (i) Fleeting and widespread breathing profiles in the range of picoseconds to nanoseconds, (ii) long-lasting disruption of one or two consecutive bp with closing times on the order of 1 μ s. The first and second kinds of events are represented by a punctuated orange (or gray in the printed version) and white patterns in Fig. 2(a), respectively.

Analysis of the trajectory indicates a large flexibility in the dsDNA filament [Fig. 2(b)], which is apparently uncorrelated

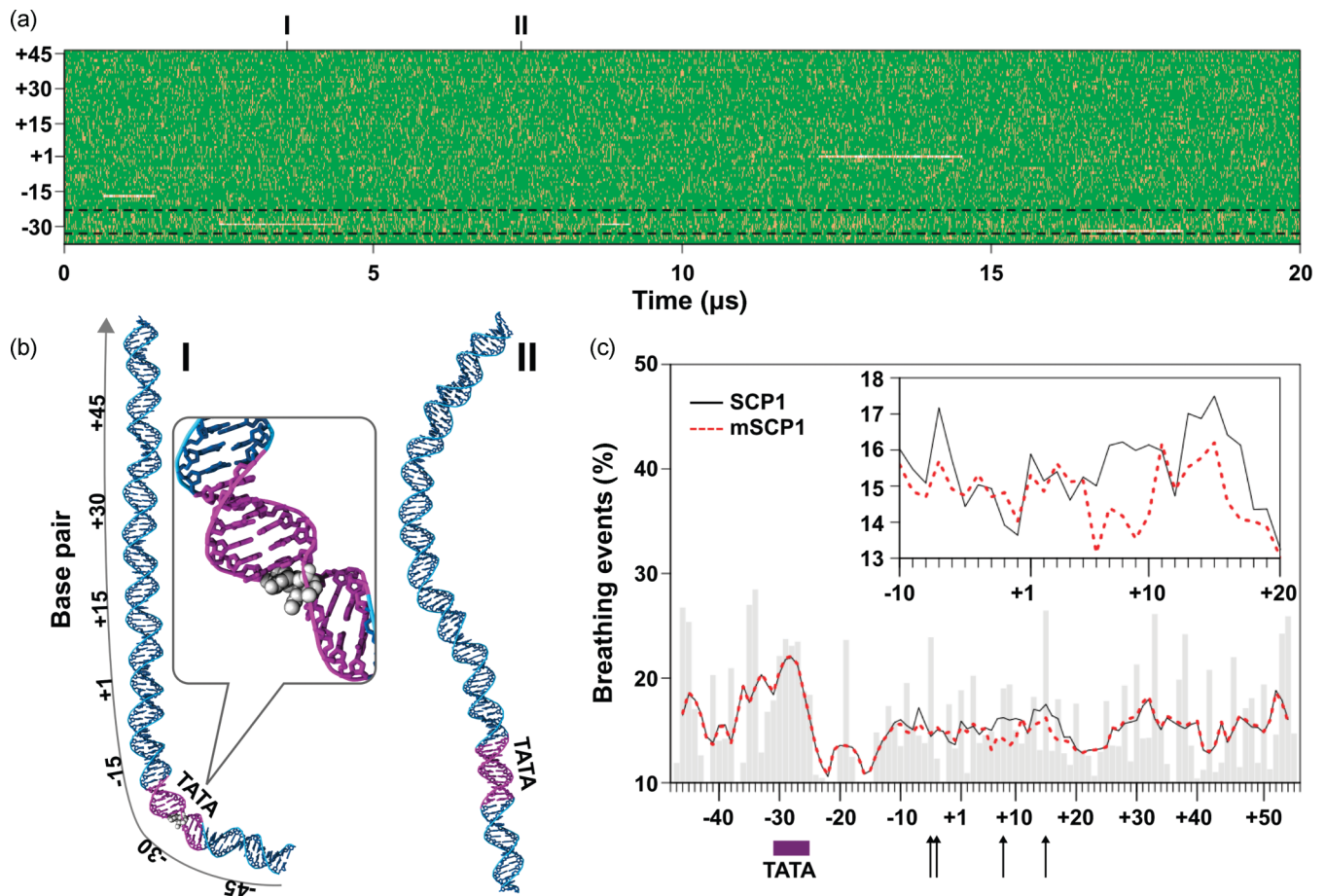


FIG. 2. (Color online) Simulations of SCP1 and mSCP1 dsDNA. (a) Breathing profile along the MD simulation of the SCP1 segment. Only the 80 bp corresponding to the promoter (underlined sequence in the Methods section) are shown for clarity. The distance between complementary bases in each bp is presented with different colors. Green (or dark gray in the printed version) color represents closed states (inter bp distances lower than 0.4 nm); open states are divided in two ranges: orange (light gray in the printed version): from 0.4 to 0.6 nm and white: more than 0.6 nm. The dashed lines delimit the AT-rich TATA box motif. (b) Representative conformers obtained along the trajectory to illustrate the flexibility of the filament. The roman numbers on top of panel (a) indicate the point in the trajectory where they come from. Inset: The single bp involved in the long-lasting disruption within the TATA box (conformer I) is shown with a space-filling representation. (c) Cumulative counting of the breathing events expressed as the percentage of time in which a bp is open between 0.4 and 0.6 nm. Gray bars are calculated for each bp in segments of the trajectory of the SCP1 filament. The continuous black line corresponds to an adjacent average every five data points. The dashed lines are calculated along the trajectory of mSCP1. The TATA box and the mutation sites are indicated by a rectangle and arrows, respectively. The inset shows a closeup on the region of the sequence where differences between SCP1 and mSCP1 are relevant.

with the fleeting breathing motion. However, we noticed that the disruption of even a single bp resulted in a marked curvature [Fig. 2(b)]. Therefore, we decided to characterize both events separately. With this aim, we calculated the cumulative sum of short breathing events of each individual bp excluding the segments of the trajectory where long-lasting disruptions are present [i.e., where the white color is present in Fig. 2(a)]. Although the fleeting breathing events may look apparently uncorrelated, the cumulative counting reveals that the AT-rich TATA box region experiences a more frequent breathing pattern. This seems to be a characteristic signature of the AT-rich tract since regions with higher nucleotide heterogeneity as the central segment, from positions -9 to $+15$, display a sensibly reduced breathing motion [Fig. 2(c)].

In contrast, the G-rich track immediately downstream of the TATA box, from positions -26 to -10 , presents the lowest

breathing occurrence. It is also worth noticing that the only peak within this region is centered on the thymine at position -19 . This underlines the capacity of the model to pinpoint the effects of single nucleotides within a given sequence context.

It has been demonstrated that point mutations at positions -4 , -5 , $+8$, and $+15$, which are distant from the TATA box, are impaired severely in the transcription levels. Moreover, a correlation between transcriptional activity and a change in the opening probability around position $+1$ has also been inferred from a different simulation approach [7]. Therefore, we sought to perform a MD simulation of mSCP1. The quantification of the breathing pattern along the dynamics of mSCP1 showed differences with that of SCP1 only within regions separated, at most, 10 bp from the mutation site, i.e., in the neighborhood of the transcription starting site [Fig. 2(c)]. The breathing profile presented in Fig. 2(c), which can be related to the opening

probability of each bp, is in very good qualitative agreement with the description presented by Alexandrov *et al.* [7]. Our simulations are in line with their conclusion that the four point mutations modulate the dynamics around the transcription starting site, leaving unaffected different protein-DNA binding sites present in the promoter.

In addition to the breathing profiles, MD simulations also grant the possibility of acquiring structural insight. Along the dynamics, the SCP1 filament may experience a significant flexion [Fig. 2(b)]. Besides relatively short breathing movements on the order of the nanosecond, the simulation of the SCP1 also presented five long-lasting events scattered along the trajectory and the sequence. Three of these occur within the TATA box [Fig. 2(a)]. These events involve the opening of one or two consecutive bp with closing times from hundreds of nanoseconds up to nearly 3 μ s. These more pronounced disruptions of the Watson-Crick pattern are related to a marked bending and, eventually, kinking [Fig. 2(b)]. Unfortunately, these long-lasting and apparently more relevant events happened in a time scale which is difficult to reach even for our CG scheme. Therefore, we sought to further explore this second kind of phenomenon in a smaller and computationally more affordable system, which allowed for longer simulation times. With this aim, we set up a series of simulations using the sequence M18 reported on in the Methods section. This sequence seems particularly well suited to our paper since it is relatively short and contains two GC “clamps” flanking an AT-rich track, which has been reported to favor the spontaneous formation of bubbles by fluorescence quenching experiments [3]. Moreover, significant breathing movements were previously reported by us for this system [9]. Hence, we speculated that a considerable increase in the sampling time could reveal a more complex behavior in terms of bubble formation and, eventually, more significant conformational changes. Therefore, we performed five independent simulations, each lasting 50 μ s, i.e., a total sampling of 250 μ s.

In analogy with the previous cases, the simulation of this shorter system revealed a complex behavior, which includes breathing, bubbling, and kinking. Moreover, we also observed the reversible separation of the 5' end of the double helix (fraying), and partial melting-rehybridization [Fig. 3(a)]. In addition, the significantly longer sampling time helps to get a clearer discrimination of breathing and bubbling patterns, which are present at significantly different time scales [compare both panels of Fig. 3(a)]. Spontaneous and transient disruption of Watson-Crick interactions, involving 2–10 bp, occurred with no apparent correlation along the MD trajectories. The shortest temporal events are in the picosecond to nanosecond range in agreement with NMR measurements [2]. The overall integrity of the double helix is well preserved during these events [conformers I and II in Fig. 3(b)]. Additionally, a reduced number of bubbles spontaneously appear in the microsecond range [Fig. 3(a)].

Bubbles lasting picoseconds to nanoseconds do not translate into large conformational changes. However, during microsecond long bubbles, the separation between opposite phosphates was increased, on average, by ~ 0.7 nm. The maximum increase in the interphosphate distance reached 1.7 nm but only for short periods (on the order of a few

nanoseconds). This is consistent with the experimental determinations indicating that bubbles between 2 and 10 bp arise spontaneously at room temperature and under physiological salt conditions with lifetimes in the range of 20–100 μ s [3]. It is uncertain, however, if these separations can fully account for the variations in the fluorescence quenching measured by Altan-Bonnet *et al.* [3].

The maximum temporal extension of the bubbles observed in our simulations is ~ 3 μ s in contrast with the nearly 1 order of magnitude higher bubbles deduced for this system from fluorescence quenching experiments [3]. This may suggest that a sampling time on the order of multimilliseconds may be needed to observe larger distortions in the backbone. In this context, it is important to recall the agreement between the lifetime of short-lived bubbles observed in our simulations and the NMR spectroscopy [2]. Although it is tempting to extrapolate this agreement to the range of the multimicroseconds, it is important to consider that the time scales sampled by CG simulations need to be, in general, interpreted with care.

A quantitative estimation of the hydrogen bond pattern (and, hence, on the stability of the double helix) can be acquired from the inter bp distances depicted in Fig. 4(a). Considering that a bp is formed at a distance below 0.4 nm [6,9], we observe that the first three bp are most frequently found in an open configuration. Furthermore, there is a rise in the inter bp distance near the middle of the double helix where long bubbles arise [compare also with Fig. 3(a)]. Besides the average and standard deviations, it is also worth paying attention to the extreme values. Although the minimum distances are limited by van der Waals contacts, relatively high maximum values are found, especially for bp numbers 18–21. Figure 4(a) provides a geometric view on the fact that, despite the global stability of the double helix, large distortions may spontaneously arise in the dsDNA at room conditions. These conclusions are in line with the results obtained by others using Langevin dynamics simulations [15].

To characterize the ordering in the open conformations, we calculated an order parameter, defined as $(\cos \langle \alpha_i \rangle)^2$, where α is the angle between the planes of individual bases between two consecutive nucleotides for all the i steps along the filament. This parameter provides an indication of the stacking between consecutive bases independent of the bp formation. The order parameter was calculated over the entire MD trajectories of M18 (i.e., 250 μ s). In order to establish a comparison level, we also simulated two independent 50 μ s long trajectories for ssDNA (see the Methods section). Moreover, to focus on the more distorted conformers within the double helix, we also selected 5 μ s of the trajectory corresponding to the two bubbles centered on microseconds 13 and 22 on the trajectory presented in Fig. 3(a). The analysis was carried out independently for that selected piece of the trajectory since averaging over the entire trajectory flattens out the results.

The ssDNA filaments showed, as expected, a large variation corresponding to a highly unstructured molecule [Fig. 4(b)]. In stark contrast, the calculation on the dsDNA trajectories shows a low ordering only in the first two bp, which is indicative of the frequent but limited fraying at the extremity of the molecule. The order rapidly increases, converging to nearly unitary values already after the third base pair. The

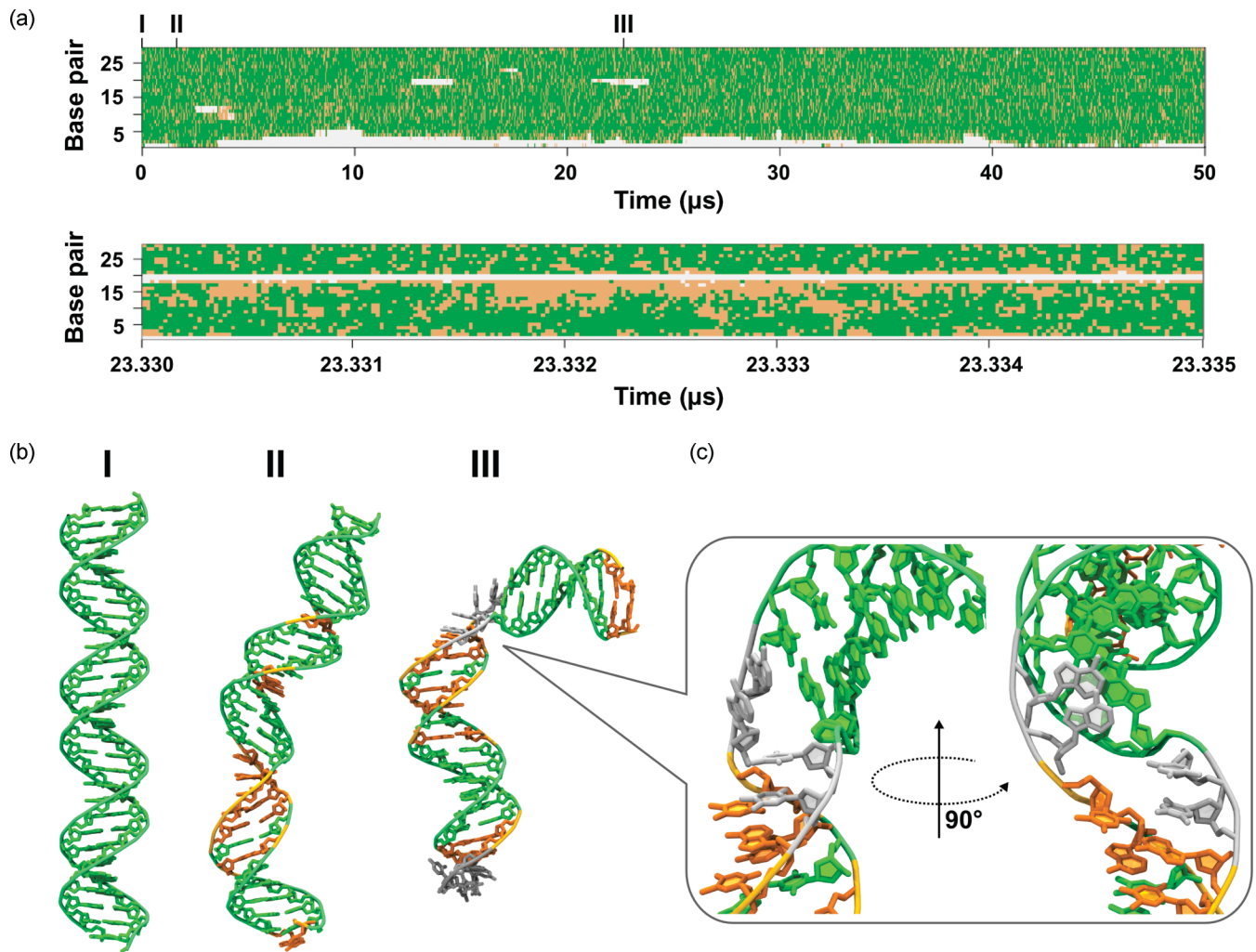


FIG. 3. (Color online) Simulations of M18 dsDNA. (a) Breathing profile along the MD simulation of the M18 segment. Color code corresponds to that of Fig. 2. Only one out of five independent 50 μ s trajectories is shown for shortness. The bottom panel shows an inset of 5 ns within the 23rd microsecond where the fine structure of short-lived bubbles can be clearly appreciated. (b) All-atoms reconstituted (backmapped) molecular representations of three different representative conformers found during the simulations. The roman numbers indicate the point on the trajectory where they are taken from. Nucleotides are colored according to panel (a). (c) Two different orientations of a closeup on the kinked conformer shown in (b) III.

selected 5 μ s trajectory containing bubbles showed a similar behavior with the exception of the highly bubbling region between bp 18 and 21. High levels of stacking between consecutive bases are kept even for bubbles with average base pair separations between 0.4 and 0.6 nm [Fig. 4(b)]. A significant decrease in the ordering is only observed when the base pair separation is, on average, higher than 0.6 nm. This supports the idea that locally denaturated regions retain some degree of order related to the base-base stacking [Fig. 3(c)], which would contribute to reduce the energy needed for bubble formation [3]. To exclude the possibility that the residual stacking in locally denaturated regions may arise from an artifact of our CG force field, we calculated the persistence length on the ssDNA filaments along the 100 μ s of the trajectory. Since the structure of ssDNA is mainly stabilized by the stacking between contiguous bases and electrostatic repulsion between phosphate groups, parametrization defects would result in deviations in the persistence length from

experimental results. The persistence length can be explicitly calculated from our simulations as $\Sigma^n \langle \cos \gamma_k \rangle$, where γ is the angle between the vectors perpendicular to the first and k th bases; the average is calculated over all the frames collected from the simulation, and the sum runs on all the n steps in the polymer. The persistence length of ssDNA resulted in 1.41 residues. This value is in very good agreement with experimental data [16], which suggest that the linearity of the ssDNA chain is completely lost already after only two residues.

Within the relatively large time window explored here, DNA breathing (but not bubbling) translates in a continuous flexion without compromising the integrity of the double helix. Clustering analysis of the curvature of DNA along the 250 μ s of simulations showed that the most visited conformation is not a completely straight one. A histogram of the total bend of the dsDNA calculated using the program CURVES+ [17] is shown in Fig. 4(c). We found that DNA bending follows

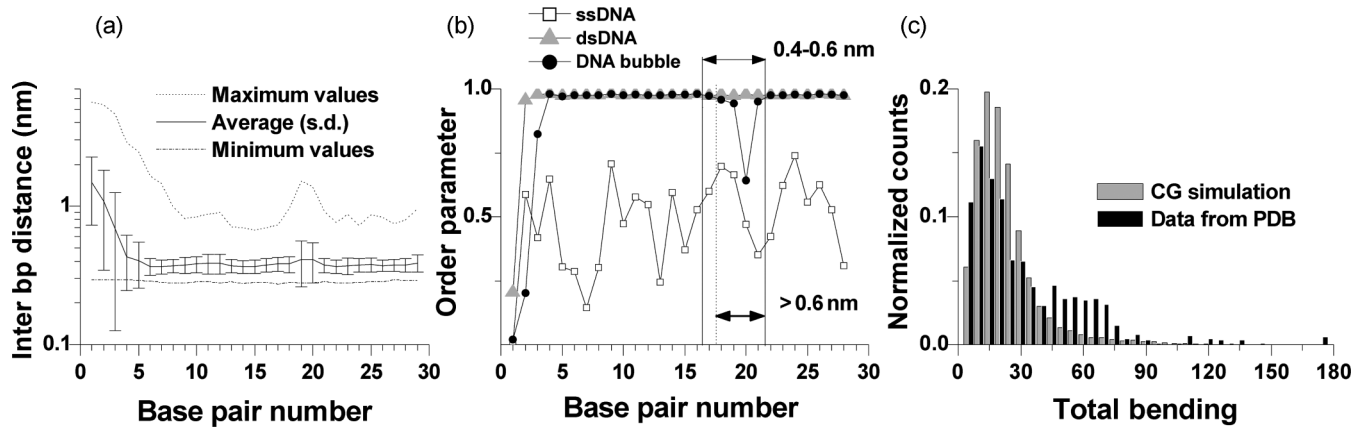


FIG. 4. Statistical analysis of DNA conformations on the M18 segment. (a) Continuous line: average inter bp distances as a function of the bp number. Standard deviations (s.d.) are reported as error bars. The maximum and minimum values measured during the dynamics are shown as dotted and dashed lines, respectively. For the sake of clearness, a logarithmic scale is used in the vertical axis. (b) Open squares: order parameter versus sequence calculated for ssDNA, gray triangles: dsDNA, and black circles: bubbled dsDNA trajectories. The arrows indicate the opening distance between bp in the region of the molecule where the bubbles occur. (c) Gray bars: histogram of the total bending of the dsDNA filament calculated over 250 μ s of simulation. The black bars show the distribution of the total bend calculated for the protein-DNA complexes from the PDB (see main text).

an asymmetric distribution with a peak between 10° and 15° . Continuously curved DNA conformers can be found up to a maximum bend of $\sim 50^\circ$. Higher bending is only observed in the presence of bubbles, which may generate kinked conformations. Taking into account these kinked conformers, the bending distribution extends up to 130° [Fig. 4(c)]. This suggests that thermal fluctuations induce a number of conformations which can be roughly divided into three categories [Fig. 3(b)]: (i) nearly canonical and straight B-DNA, (ii) continuously bent DNA without significant bubbles in the double helix, and (iii) conformations with widely opened bubbles, which may generate kinked double helical filaments. The first two categories can be very well described by standard semiflexible polymer models [18], which predict a maximum thermally induced curvature of nearly 1 rad in double-stranded segments with extensions close to their persistence length. However, the kinked conformers arising in our simulations correspond to rare events with typical occurrence times on the microsecond scale.

The DNA deformability suggested by our results could be essential for defining its biophysical properties. In fact, sharp kinks related to spontaneous bubble formation have been proposed to explain DNA cyclization probabilities $> 10^4$ times larger than those predicted by standard semiflexible polymer models for ~ 100 bp long DNA segments [19] (i.e., shorter than their persistence length). Similar protein-independent DNA flexibility has also been reported using different experimental techniques [20–22]. According to a model proposed by Yan and Marko [23], thermal fluctuations are enough to generate hinges involving at least 3 bp long bubbles increasing the flexibility and the probability of cyclization. Contrasting interpretations for the high flexibility of DNA in terms of the disruption of Watson-Crick interactions or the continuous stacking between nucleobases has also been provided by Du *et al.* [24] and Geggier *et al.* [25]. The possibility of accessing nearly atomistic information from our CG simulations provides a structural picture for the idea that single bp openings

may occur spontaneously in dsDNA. Bubbles may reach an extension up to ten consecutive bp with a disruption of the Watson-Crick hydrogen bond pattern, but residual ordering related with the base-base stacking is retained even in the bubbled region. When these distortions occur, they translate into a sharp kink in the double helix [Figs. 3(b) and 3(c)].

Simple geometrical calculations indicate that, at least, a single kink with an angle $\geq 120^\circ$ plus the continuous bending of the rest of the chain is enough to join the two extremities of 100 bp dsDNA filaments in a teardrop shape. The probability of the occurrence of kinks higher than 120° calculated from the distribution of the bending angles [Fig. 4(c)] is on the order of 10^{-4} , supporting the spontaneous occurrence of rare conformational defects in DNA as responsible for cyclization of relatively short DNA segments.

The biological relevance of this phenomenon is highlighted if we compare our results with the bending measured from protein-dsDNA complexes reported in the Protein Data Bank. With this aim, we considered all the protein-dsDNA complexes solved at a resolution higher than 0.25 nm as taken from the human curated protein DNA interface database (<http://melolab.org/pdibd>) [26]. There is a fairly good correspondence between the angular distributions of free DNA (from our simulations) and the set of nearly all the experimentally determined protein-dsDNA complexes [Fig. 4(c)].

It also has to be taken into account that sequence-dependent effects are expected to modulate DNA deformability [27]. Recently, using another CG model combined with umbrella sampling techniques, Ortiz and de Pablo showed how sequence can change the ability of DNA to form kinks in nucleosome positioning segments [5]. Although sequence-dependent effects have not been explored in this paper (neither in the simulations nor in the analysis of the experimental DNA-protein complexes), our results suggest that thermally driven bending and metastable kinking of DNA is an intrinsic property of the double helical architecture.

IV. CONCLUSIONS

To summarize, the set of simulations presented here provide nearly atomistic details of the dynamics of DNA in the time scale of the multimicroseconds, yet unexplored. The simulation of the SCP1 system presents breathing profiles, which are consistent with transcription experiments and other theoretical methods. In addition, we obtain the indication that even one single bp disruption may originate marked kinks in the double helical filament. Comparison with mSCP1 suggests that the effect of point mutations can propagate up to 10 pb in the double helix.

The relatively long simulations of the M18 sequence highlight the relevance of long bubbles spontaneously appearing in dsDNA. These bubbles of up to 10 bp may originate marked kinks in the double helix. These kinks present noncanonical but still ordered structural motifs, which are stable in the multimicrosecond time scale. The relatively long lifetime of these conformers could have a deep biological relevance. In fact, comparison of the bending produced by spontaneously

arising kinks with most of the experimentally determined protein-dsDNA complexes shows good correspondence. This suggests that thermally induced deformation of the double helix could be sufficient to overcome the free energy barrier needed to obtain about 80% of the DNA-protein complexes currently known.

ACKNOWLEDGMENTS

We thank J. J. Cifuentes and F. Melo for help with the analysis of the information from Protein-DNA interactions Database. This work was supported by ANII, Agencia Nacional de Investigación e Innovación, Programa de Apoyo Sectorial a la Estrategia Nacional de Innovación INNOVA URUGUAY (Agreement No. 8 DCI-ALA/2007/19.040 between Uruguay and the European Commission). M.R.M. is supported by a fellowship from CSIC-UdelaR. P.D.D. and S.P. appreciate support from the National Scientific Program of ANII (SNI) and from the Basic Science Development Program of Uruguay (PEDECIBA).

-
- [1] J. SantaLucia, Jr., *Proc. Natl. Acad. Sci. USA* **95**, 1460 (1998).
 - [2] M. Gueron and J. L. Leroy, *Methods Enzymol.* **261**, 383 (1995).
 - [3] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, *Phys. Rev. Lett.* **90**, 138101 (2003).
 - [4] A. Savelyev and G. A. Papoian, *Proc. Natl. Acad. Sci. USA* **107**, 20340 (2010).
 - [5] V. Ortiz and J. J. de Pablo, *Phys. Rev. Lett.* **106**, 238107 (2011).
 - [6] T. A. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo, *J. Chem. Phys.* **126**, 084901 (2007).
 - [7] B. S. Alexandrov, V. Gelev, S. W. Yoo, L. B. Alexandrov, Y. Fukuyo, A. R. Bishop, K. O. Rasmussen, and A. Usheva, *Nucleic Acids Res.* **38**, 1790 (2010).
 - [8] T. Juven-Gershon, S. Cheng, and J. T. Kadonaga, *Nat. Methods* **3**, 917 (2006).
 - [9] P. D. Dans, A. Zeida, M. R. Machado, and S. Pantano, *J. Chem. Theory Comput.* **6**, 1711 (2010).
 - [10] F. Zhang and M. A. Collins, *Phys. Rev. E* **52**, 4217 (1995).
 - [11] L. Darré, M. R. Machado, P. D. Dans, F. E. Herrera, and S. Pantano, *J. Chem. Theory Comput.* **6**, 3793 (2010).
 - [12] D. A. Case *et al.*, AMBER 10, University of California, San Francisco, 2008.
 - [13] R. W. Pastor, B. R. Brooks, and A. Szabo, *Mol. Phys.* **65**, 1409 (1988).
 - [14] J. Srinivasan, M. W. Trevathan, P. Beroza, and D. A. Case, *Theor. Chem. Acc.* **101**, 426 (1999).
 - [15] O. C. Lee, J. H. Jeon, and W. Sung, *Phys. Rev. E* **81**, 021906 (2010).
 - [16] S. B. Smith, Y. Cui, and C. Bustamante, *Science* **271**, 795 (1996).
 - [17] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, *Nucleic Acids Res.* **37**, 5917 (2009).
 - [18] P. J. Hagerman, *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265 (1988).
 - [19] T. E. Cloutier and J. Widom, *Mol. Cell* **14**, 355 (2004).
 - [20] N. A. Becker, J. D. Kahn, and L. J. Maher III, *J. Mol. Biol.* **349**, 716 (2005).
 - [21] C. Yuan, H. Chen, X. W. Lou, and L. A. Archer, *Phys. Rev. Lett.* **100**, 018102 (2008).
 - [22] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, *Nat. Nanotechnol.* **1**, 137 (2006).
 - [23] J. Yan and J. F. Marko, *Phys. Rev. Lett.* **93**, 108108 (2004).
 - [24] Q. Du, A. Kotlyar, and A. Vologodskii, *Nucleic Acids Res.* **36**, 1120 (2008).
 - [25] S. Geggier, A. Kotlyar, and A. Vologodskii, *Nucleic Acids Res.* **39**, 1419 (2011).
 - [26] T. Norambuena and F. Melo, *BMC Bioinf.* **11**, 262 (2010).
 - [27] A. Perez, F. J. Luque, and M. Orozco, *Acc. Chem. Res.* **45**, 196 (2012).