

Reaction path ensemble of the B–Z-DNA transition: a comprehensive atomistic study

Mahmoud Moradi, Volodymyr Babin, Christopher Roland and Celeste Sagui*

Center for High Performance Simulations (CHiPS) and Department of Physics, North Carolina State University, Raleigh, NC 27695-8202, USA

Received August 10, 2012; Revised September 29, 2012; Accepted October 1, 2012

ABSTRACT

Since its discovery in 1979, left-handed Z-DNA has evolved from an *in vitro* curiosity to a challenging DNA structure with crucial roles in gene expression, regulation and recombination. A fundamental question that has puzzled researchers for decades is how the transition from B-DNA, the prevalent right-handed form of DNA, to Z-DNA is accomplished. Due to the complexity of the B–Z-DNA transition, experimental and computational studies have resulted in several different, apparently contradictory models. Here, we use molecular dynamics simulations coupled with state-of-the-art enhanced sampling techniques operating through non-conventional reaction coordinates, to investigate the B–Z-DNA transition at the atomic level. Our results show a complex free energy landscape, where several phenomena such as over-stretching, unpeeling, base pair extrusion and base pair flipping are observed resulting in interconversions between different DNA conformations such as B-DNA, Z-DNA and S-DNA. In particular, different minimum free energy paths allow for the coexistence of different mechanisms (such as zipper and stretch-collapse mechanisms) that previously had been proposed as independent, disconnected models. We find that the B–Z-DNA transition—in absence of other molecular partners—can encompass more than one mechanism of comparable free energy, and is therefore better described in terms of a reaction path ensemble.

INTRODUCTION

In 1979 the publication of the crystal structure of a d(CGC GCG)₂ duplex revealed a left-handed double-helix DNA with two antiparallel chains joined by Watson–Crick (WC) base pairs (1). This structure (Figure 1) was

termed Z-DNA because its sugar–phosphate backbone forms a characteristic zig-zag pattern. Left-handed Z-DNA is formed by dinucleotide repeats and occurs in sequences that alternate a purine–pyrimidine repeat, mainly CG (or GC). The *anti-syn* alternation of these base pairs underlies the zig-zag pattern and is due to the rotation of the guanine residue around its glycosidic bond, resulting in a *syn* conformation, whereas the cytosine residue retains its *anti* conformation. There is a high density of base sequences favoring Z-DNA near transcription start sites (2). Z-DNA is also induced by Z-DNA-binding proteins near the promoter region which boosts the transcription of the downstream genes (3). It is known that Z-DNA is highly immunogenic, and there are antibodies against it that have been used to map regions favorable to Z-DNA conformations (4–6). All in all, it is now thought that Z-DNA formation is closely related to gene expression, regulation and recombination (3,7–14).

Despite all the investigations about the biological role of left-handed Z-DNA, the microscopics behind the B–Z-DNA transition have remained controversial, with several different mechanisms proposed in the literature (15). Generally speaking, the models for the B–Z transition are classified into mechanisms that involve either base pair opening or base pair rotation without WC base pair breaking. Each of these mechanisms may or may not have an intermediate structure. Among the models without intermediate structure, the most popular base pair opening mechanism is the Wang model (1). It proposes that base pair opening occurs before base pair plane and phosphate backbone angle rotation within the core of the helix. The Harvey model (16), on the other hand, proposes that the B–Z transition happens through the successive flipping of base pair planes, without any disruption of the WC pairs. This process is supposed to be facilitated by breathing modes. Experimental evidence, however, seems to favor models with intermediate structures (17–21). The Saenger–Heinemann model postulates that the transition takes place through two A-DNA-like intermediates (the second one with a dinucleotide unit), without breaking of the base pairs (20). More recently,

*To whom correspondence should be addressed. Tel: +1 919 515 3111; Fax: +1 919 513 804; Email: sagui@ncsu.edu

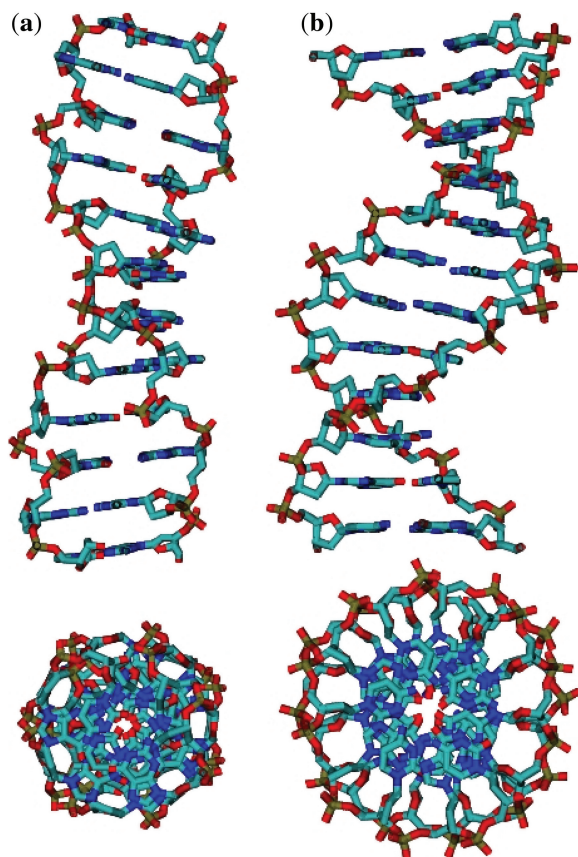


Figure 1. Structural differences of B- and Z-DNA conformations. Side and top views of (a) an ideal left-handed Z-DNA and (b) an ideal right-handed B-DNA with 12 base pairs. Hydrogen atoms are not shown.

the extrusion of bases was observed in the crystal structure of a B–Z junction (21). These extrusions and their propagation followed by the reformation of the base pairs in a new order also represents an alternative mechanism for the transition (21) (similar to Wang’s model, except that the bases rotate outside the double-helix core).

Computational studies have subsequently been used to examine these mechanisms at an atomic level. The methodology of these studies varied from the use of stochastic difference equations (19,22), to dynamic phase diagrams (23) and, most notably, large-scale atomistic molecular dynamics (MD) simulations (24,25). The initial investigations (19,24,26) proposed yet another model for the B–Z-DNA transition that involves the simultaneous formation of a stretched intermediate conformation with parallel phosphate backbones, the transient disruption of the WC hydrogen bond pairing, base flipping and the re-ordering into Z-DNA. In contrast, Lee *et al.* (25) carried out an explicit solvent simulation of the mechanism proposed by Ha *et al.* (21). This alternative mechanism is based on the stepwise propagation of a B–Z junction. The estimated transition barrier for this model was considerably less than the barrier obtained for the stretched intermediate model. The authors concluded that the B–Z transition involves a relatively fast growth of B-DNA or Z-DNA structures by a sequential propagation of B–Z junctions, once an initial junction is nucleated.

The crystal structure of a B–Z junction (21) has revealed the full extrusion from the helix of the two junctional bases. The formation of a B–Z junction requires little structural disruption, because it preserves the integrity of both the B- and Z-DNA helices, as well as the base stacking between the two helices. Although a B–Z junction is formed at the interface between B- and Z-DNA, a Z–Z junction is also commonly formed in sequences where the dinucleotide repeat is interrupted by single base pair insertions or deletions, which bring neighboring helices out of phase. The three-dimensional structure of a Z–Z junction has been recently described experimentally (27). The Z–Z junction studied by de Rosa *et al.* is stabilized by $Z\alpha$, the Z-DNA-binding domain of the RNA editing enzyme ADAR1, and consists of a single base pair AT in the middle of a CG sequence, with the resulting duplex sequence 5′-(CG)₃-A-(CG)₃-3′. The AT pair leads to a partial disruption of the helical stacking. In contrast to a B–Z junction, the bases of this structure are not fully extruded, and the stacking between the two left-handed helices is not continuous.

In this article, we report on a large-scale, ‘atomistic’, MD simulation study of the B–Z transition and the structural characteristics of the B–Z and Z–Z junctions. The focus is on the mechanisms and free energy of the transition, which were calculated by means of the recently developed adaptively biased molecular dynamics (ABMD) (28,29) and steered molecular dynamics (SMD) (30) methods. Because so many combinatorial factors are involved in the relative equilibrium of these systems, we simplify the picture as much as possible: simple (CG) sequences of various lengths (at most interspersed with one AT base pair) and without chemical modifications, no other binding molecules, physiological temperature and pressure conditions, aqueous environment and the limit of infinite salt dilution. The latter, of course, considerably decreases the stability of Z-DNA with respect to B-DNA, but the use of enhanced sampling techniques allows for the crossing of very high energy barriers, and can thus throw light on the nature of the transition mechanisms, giving also a reference point for future studies of the effects of varying salt concentration. We observe a wide range of different phenomena in our simulations, including but not limited to, unpeeling, over-stretching, base pair extrusion and base pair flipping, never observed all together in previous experimental or computational studies. Our results are consistent with both a ‘stretch–collapse’ mechanism involving an S-DNA-like intermediate, and an extruded-basis zipper propagation mechanism. Moreover, the transition barriers for the two are quite comparable, and so in absence of other intervening molecules, these mechanisms are likely to compete with each other (at least for short DNA strands) during a B–Z-DNA transition.

MATERIALS AND METHODS

A description of the simulation methodology may be found in the Supplementary Methods, which explicitly discusses the free energy methods, provides detailed

definitions of the collective variables and gives all the relevant simulation details.

To calculate accurate free energy maps as a function of several sets of collective variables, we used the recently developed ABMD method (28,29). ABMD is a non-equilibrium MD method that belongs to the general category of umbrella sampling methods with a history-dependent biasing potential (31–35). This method provides for an elegant way of computing the free energy (36)—or potential of mean force (PMF)—in terms of a set of collective variables. The ABMD method is currently implemented in AMBER v.10–12 (37–39), and has already been applied to a variety of biomolecular systems (28,29,40–47). SMD (30) is another non-equilibrium MD method that makes use of a time-dependent biasing potential to ‘steer’ the system between two known states. This method is particularly useful for examining select pathways and mechanisms between two equilibrium states, as well as estimating the transition rates for these reactions (48). Here, we have used SMD to generate initial structures for our ABMD simulations and to qualitatively compare selected transition pathways and mechanisms.

Both the ABMD and SMD methods require that one or more collective variables be specified. The main collective variables used in this work include (i) the handedness (H) of the DNA strands, which quantifies the extent of left- and right-handed helical conformations; (ii) the radius of gyration (R_g), which is a measure of the length of the helical structures; (iii) a collective variable $X = \sin(\chi_{\text{purine}}) + \sin(\chi_{\text{pyrimidine}})$, which is defined in terms of the glycosidic torsion angles (χ) of a base pair; (iv) a collective variable N_{WC} , which indicates if a given base pair has the correct WC hydrogen bonding and (v) N_B (N_Z), which counts the number of base pairs that are in a B-DNA (Z-DNA) conformation. Right-handed helical turns have handedness $H > 0$ and left-handed helical turns have $H < 0$. Overall handedness of the helix is obtained by summing over all turns. Relevant values for X range from about -1.96 to -0.68 for right-handed A- or B-DNA-like structures; and from about 0.13 to 0.57 for left-handed Z-DNA-like structures. The X collective variable is therefore able to probe the specific structure of a given DNA base pair. For a base pair, $N_{\text{WC}} \simeq 1$ when the base pair is in a correct WC conformation, and $N_{\text{WC}} \simeq 0$ if the base pair is broken or in a non-WC base pair configuration.

Simulations were carried out for the DNA sequences $(\text{CG})_6$, which is a short-hand notation for $d(5'\text{-CGCGCGCGCGCG-}3')_2$, $(\text{CG})_4$ and $(\text{CG})_3\text{A}(\text{CG})_3$, using an implicit water model based on the generalized Born approximation with the surface area (GB/SA) term (49,50). The simulations used the ff99 version of the Cornell *et al.* (51) force field, with the leap-frog algorithm and a 1 fs timestep, along with Langevin dynamics at $T = 300$ K and a cut-off of 32 Å for the non-bonded interactions. Initial configurations were generated using the LEaP program of the AMBER v.9 simulation package (52) for B-DNA (for all three sequences) and 3DNA package (53,54) for Z-DNA [for the $(\text{CG})_6$ and $(\text{CG})_4$ sequences]. To generate the initial configurations for the B–Z and Z–Z junctions, we used a different procedure based on the SMD method as outlined in the Supplementary

Methods. All the MD simulations were run using the ‘SANDER’ program of AMBER v.11 package (38).

For the dodecamer $(\text{CG})_6$ duplex, ABMD simulations were used to construct the (H, R_g) free energy map. In addition, we performed two sets of SMD simulations steering the system between its B- and Z-DNA conformations: one SMD in the (H, R_g) space steering along the least free energy paths (LFEPs) (55) and the other SMD in the (N_B, N_Z) space. For the octamer $(\text{CG})_4$, we ran ABMD simulations for three different conformations: pure B-DNA, pure Z-DNA and the B–Z junction, using the collective variables (X, N_{WC}) defined on the base pair 4. Finally, we ran ABMD simulations for three $(\text{CG})_3\text{A}(\text{CG})_3$ systems: pure B-DNA, and Z–Z and the B–Z junctions. We used two distance-based collective variables: one defined between the O_4 atom of the thymine residue and N_6 atom of the adenine residue, $d(O_4^T - N_6^A)$ and the other defined between the N_3 atom of the thymine residue and N_1 atom of the adenine residue, $d(N_3^T - N_1^A)$.

RESULTS

B–Z-DNA transition

We now present our ABMD and SMD simulation results for a dodecamer DNA undergoing a B–Z transition. The two-dimensional (H, R_g) free energy plot is given in Figure 2. This free energy landscape displays a number of free energy minima, with the most prominent being associated with B-DNA and Z-DNA. The former represents the global minimum, whereas the latter is a local minimum with a free energy that is ~ 9.0 kcal/mol higher. We note that for this figure there is no particular distinction between structures such as B- and A-DNA, which are degenerate for this set of collective variables. The estimated error for this plot is about 2 kcal/mol, based on the umbrella correction runs.

On this free energy landscape we identified two different LFEPs associated with a stretch–collapse mechanism (solid line in the plot) and with a zipper mechanism (dashed line). In this plot, B- and Z-DNA configurations are color-coded. There are several other local minima, particularly in the lower part of the free energy landscape, that are neither B- nor Z-DNA. We have examined the

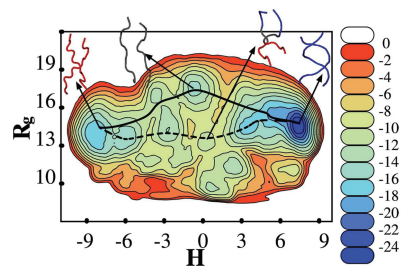


Figure 2. Free energy landscape of $(\text{CG})_6$ DNA. Two-dimensional (H, R_g) free energy landscape for $(\text{CG})_6$. Ribbon diagrams of the structure backbones associated with four minima are also shown, with blue (red) representing a B-DNA (Z-DNA) base pairs. Gray ribbons are in neither conformation. Two different LFEPs over this landscape are plotted with solid and dotted lines, associated with a stretch–collapse and a zipper-like mechanism, respectively.

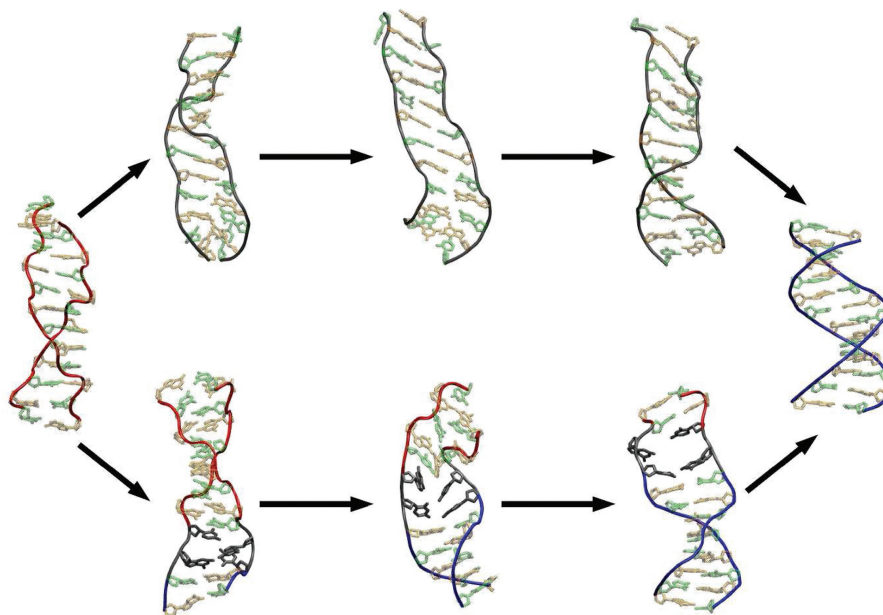


Figure 3. Alternative B–Z-DNA transition pathways. Snapshots of a Z- to B-DNA transition for the dodecamer (CG)₆ obtained from SMD simulations through the LFEP trajectories in Figure 4. The top series shows a stretch–collapse mechanism and the bottom series shows a base extrusion zipper mechanism. A ribbon-based representation of the backbone is shown, with a color scheme similar to that of Figure 2. In addition, the cytosine and guanine bases are colored green and orange, respectively. The bases of intermediate configurations are colored gray in the stepwise mechanism.

DNA structure at these points and, roughly speaking, these minima are associated with structures with disrupted base pairs, including anomalous non-WC base pairing, hydrogen bonding between the bases of the same DNA strands, etc.

Although Figure 2 shows one intermediate configuration for each LFEP trajectory, intermediate configurations obtained from SMD simulations are shown in Figure 3. The stretch–collapse mechanism involves relatively little disruption of the base pairs (Figures 2 and 3). Instead, the DNA stretches and unwinds, forming two almost parallel strands that preserve their WC base pairs, assuming an S-DNA-type conformation (56). This conformation then rewinds to form a double helix of the opposite handedness. This mechanism is associated with relatively large changes in R_g . Notice that SMD with R_g (or end-to-end distance) as a collective variable is an ideal tool to complement DNA stretching experiments, such as those carried by AFM and optical trap. In particular, for CG sequences, a B–S-DNA transition is proposed to occur after a threshold force (57). For larger forces, S-DNA gives way to ‘unpeeling’, where one strand falls off the other, and only the remaining single strand carries the tension. Our simulations can reproduce this scenario: if we steer the system from B-DNA up to a maximum at $(H, R_g) \sim (0, 20 \text{ \AA})$, S-DNA with WC pairs is observed; steering beyond 20 Å results in base pair breaking from one end to the other.

In contrast, the zipper-like mechanism involves almost no change in R_g . However, there are large changes of handedness within the structure, and the intermediate structures involve at least one B–Z junction, generally

with extruded base pairs. Steering the system through the LFEP associated with the zipper mechanism on the (H, R_g) landscape is quite challenging, because WC base pairs are broken, and the bases extruded. Thus, it can take quite a long time for the base pairs to re-assemble, a fact that is consistent with experiments (it takes from milliseconds to seconds for a single base flipping). Independent SMD simulations involving the (N_B, N_Z) set of collective variables confirm a zipper mechanism. Notice that, by definition, these variables determine how many base pairs will be in the B- or Z-conformation [and therefore (N_B, N_Z) preclude the stretch–collapse mechanism], but they are completely degenerate with respect to the order and the fashion in which the bases flip. Both short 10 ns and long 100 ns SMD simulations with the (N_B, N_Z) variables are consistent with a zipper mechanism for the B–Z-DNA transition.

These results clearly show that the SMD simulations along different trajectories on the (H, R_g) landscape are very sensitive to the trajectory chosen. In addition to the ‘physically probable LFEPs’ shown in Figure 2 and the corresponding transition mechanisms shown in Figure 3, one could choose any other trajectory joining the B- and Z-DNA minima. One of such ‘arbitrary’ paths, e.g. is obtained by going linearly from the minimum in B-DNA to $(H, R_g) = (0, 20 \text{ \AA})$, where the structure becomes almost a perfect S-DNA, and then linearly to the minimum in Z-DNA, where the structure collapses into a different left-handed structure. This left-handed structure has WC base pairing, and shares the same values of (H, R_g) than Z-DNA, but is still quite different from Z-DNA: it has a WC-type backbone direction

similar to Z[WC]-DNA (18), which is the same backbone direction as that of B-DNA. Supplementary Figure S3 compares this left-handed structure with conventional Z-DNA.

We have also monitored the glycosidic torsion angle (χ) of guanine and cytosine nucleotides along the B–Z transition path as obtained from SMD simulations associated with the stretch–collapse and the zipper mechanisms. In both mechanisms, the guanine bases change from *anti* to *syn*, whereas the cytosine bases start and end, as expected, in the *anti* conformation. In the stretch–collapse mechanism, there is no particular order for the change in χ and both guanines and cytosines flip from *anti* to *syn* although cytosines eventually come back to the *anti* conformation. The temporary *syn* conformation of the pyrimidines seems to correlate with the appearance of the S-DNA intermediate structure. In contrast, in the base extrusion zipper mechanism, the jump from *anti* to *syn* for the guanines takes place in an ordered, zipper-like fashion and cytosines stay in the *anti* conformation along the transition (Supplementary Figure S4 shows a time series of the χ torsion angles).

The stretch–collapse mechanism does not involve much disruption in the WC base pairing along the transition. Although we are using a small constraint on the base pairs, this constraint does not prevent the disruption of the base pairs, which is sometimes observed when the system is steered around the $H \approx 0$ region. In the base extrusion mechanism, on the other hand, the flipping of the glycosidic angle of a nucleotide happens only after the base pair associated with that nucleotide has been broken. We also note that even though the (N_B, N_Z) collective variables tend to disfavor extruded bases (because neither the B- nor Z-DNA reference structure has any broken base pair), all the base pairs break before their glycosidic angles flip. It is interesting that flipping always occurs at the peak of the extrusion process (Supplementary Figure S5 shows a time series of both the $N_3^G - N_1^C$ distance and the purine glycosidic angle for select base pairs as obtained from (N_B, N_Z) -based SMD simulations).

Finally, we have measured free energies along the normalized pathways for both transition mechanisms, as well as the associated total work versus time plots for the SMD simulations. These plots are shown in Figure 4, which reveals that both mechanisms have similar free energy profiles. In particular, we note that the work profile for the stretch–collapse mechanism is relatively smooth, whereas the zipper-like mechanism displays more step-like changes. Likewise, the free energy profiles of Figure 4a reflect this as well: there are more minima associated with the dashed line describing the zipper mechanism than with the continuous-line stretch–collapse mechanism.

B–Z and Z–Z junctions

In this section, we discuss the free energy maps and some structural characteristics of the B–Z and Z–Z junctions. First, we studied the free energy cost to flip the configuration of an inner base pair, while the rest of the base pairs remains in the same initial conformation. Figure 5 shows

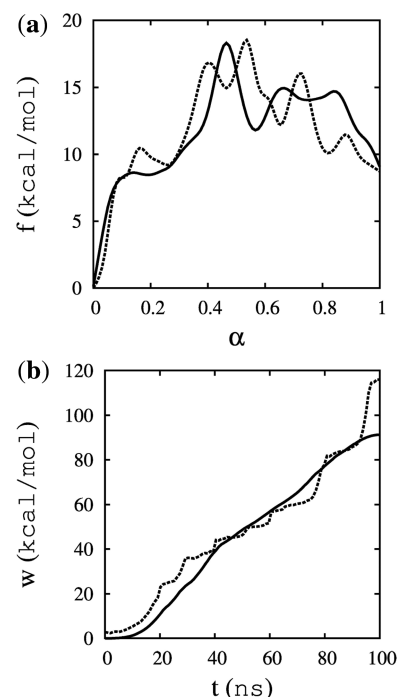


Figure 4. Free energy profile along different B–Z-DNA transition pathways. Here, (a) presents the free energy along two LFEPs over the (H, R_g) space as a function of the normalized path length α as the system moves from B- to Z-DNA. In (b), the total work as a function of time for SMD runs for the B–Z transition, moving along the stretch–collapse (solid line) and base extrusion zipper (dotted line) LFEPs, is plotted.

the (X, N_{WC}) ABMD free energy landscape for the flipping of the fourth base pair, given that the $(CG)_4$ duplex is in a (a) B-DNA; (b) Z-DNA or (c) B–Z junction conformation. In general terms, the minima associated with $N_{WC} \approx 1$ indicate a proper WC base pair, whereas a value of ≈ 0 is indicative of either broken base pairs, or non-WC hydrogen bonding. In terms of the collective variable X , a negative value < -0.7 is associated with B- or A-DNA base pairing whereas a positive value of ~ 0.1 – 0.6 is associated with Z-DNA base pairing (see ‘Methods’ section).

Figure 5 essentially shows the cooperativity among base pairs in DNA: in a pure B-DNA or pure Z-DNA environment, the fourth base pair also favors that conformation, with minima at $(X, N_{WC}) = (-0.8, 1)$ for B-DNA and $(0.5, 1)$ for Z-DNA. Alternatively, the forced disruption of the 4th base pair leads to a conformation where the base pair is not part of a WC bond, as shown by the minima at $N_{WC} \approx 0$. The presence of a B–Z junction, on the other hand, overwhelmingly favors non-WC base pairing, and switching of the fourth base pair between B- and Z-DNA conformations appears to be almost equally likely. Table 1 gives the free energies associated with the B- and Z-DNA minima relative to the minimum around $(0, 0)$ that represents the typical case of a broken base pair. Note that one can identify several other minima in these plots. In the case of $N_{WC} \approx 0$, they consist of a confusing set of broken bonds that correspond to a mixture of A-, B- and Z-DNA characteristics.

Particularly, there is a common deep minimum around $(-2, 0)$ that corresponds to an extreme case with both χ angles around -90° . However, our main focus here is on the minima associated with B- and Z-DNA conformations, and the relative stability of these conformations, when put in the middle of a B-DNA, Z-DNA or a B-Z junction. Sample configurations of a B-Z junction are shown in Figure 6a and b, whereas Supplementary Table S1 gives average values of six common structural characteristics, along with their standard deviation. This average is over the conformations of the ABMD trajectories weighted by their estimated probabilities based on the free energy maps. We used the 3DNA package (53,54) to measure the step parameters of individual conformations but only the steps with no disrupted WC base pairs were included in the averaging.

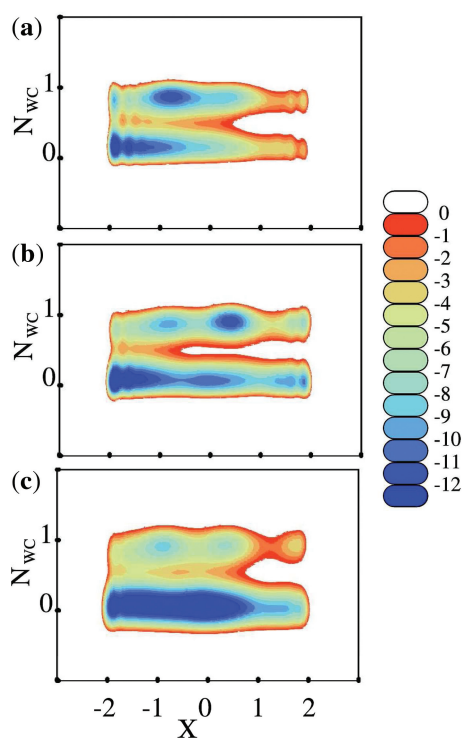


Figure 5. Free energy landscapes of a CG base pair in different conformations. Free energy maps of an 8 bp DNA defined in terms of N_{WC} and $X = \sin(\chi_C) + \sin(\chi_G)$ for the fourth base pair, with the rest of the strand in (a) B-DNA; (b) Z-DNA and (c) B-Z junction form.

Now we turn to the case of the purine-pyrimidine repeats of a duplex being interrupted by a single AT base pair, placed in the middle of a $(CG)_n$ duplex, as given by the $(CG)_3A(CG)_3$ duplex. A two-dimensional free energy plot for this situation is shown in Figure 7, with a set of collective variables defined in such a way as to track the hydrogen bonding of the AT base pair. Specifically, we define $d(O_4^T - N_6^A)$ as representing the distance between the O_4 atom of the thymine base and the N_6 atoms of the adenine base, whereas $d(N_3^T - N_1^A)$ is similarly defined for atoms N_3 and N_1 for thymine and adenine, respectively. The $(CG)_3$ parts are either in B- or Z-DNA conformations, so we have three cases: (i) B-DNA; (ii) a Z-Z junction or (c) a B-Z junction. The distances defined above vary with the type of base pairing, as illustrated in Figure 8. For instance, Figure 8a shows a typical WC base pairing in B-DNA, whereas Figure 8b, c and d show three typical Z-Z junctions, similar to those previously observed by de Rosa *et al.* (27). Here, Figure 8b demonstrates a reverse WC base pairing with two hydrogen bonds, whereas Figure 8c and d show structures with only one and no hydrogen bonds, respectively.

Figure 7 presents the free energy landscapes for these various junctions, whereas the free energies associated with the minima are given in Table 1 and structural characteristics obtained by averaging over the ABMD trajectories, weighted according to the corresponding free energy maps, are given in Supplementary Table S1. Figure 7a shows that the addition of the AT base pair does not alter the stability of the B-DNA duplex. The global minimum is centered about (3,3) (in Å) and corresponds to proper WC pairing (Figure 8a). Two other local minima are readily identified at (6,3) and (6,6), and correspond to configurations with a single and zero hydrogen bonds, respectively. The configuration at (6,3) resembles the configuration shown in Figure 8c. The free energy landscape of the Z-Z junction, on the other hand, is characterized by many more minima, some of which are quite broad. Note that the minimum at (3,3) has disappeared, indicating that the inserted base pair does not form a WC bond. The global minimum in this case has shifted to (3,6), with configurations characterized by a reverse WC bond as that shown in Figure 8b (where the second H-bond is between O_2 from thymine and N_6 from adenine). A second broad minimum of competing depth is centered around (8,6), which corresponds to configurations without

Table 1. Free energies of the select minima associated with DNA free energy maps

$(CG)_4$	Free energies		$(CG)_3A(CG)_3$	Global	Free energies			
	$(-0.8,1)$	$(0.5,1)$			Conformation	Minimum	$(3,3)$	$(6,3)$
B	-4.08	-0.26	B	(3,3)	0	3.68	10.47	5.87
Z	1.51	-1.89	Z-Z	(3,6)	4.08	0.12	0	0.33 ^a
B-Z	6.98	7.06	B-Z	(11,11)	3.27	3.19	6.27	2.02

Free energies (in kcal/mol) of some of the minima associated with the free energy maps, shown in Figures 5 and 7. The free energies for $(CG)_4$ structures (Figure 5) are measured relative to the minimum around (0,0) while the free energies for $(CG)_3A(CG)_3$ structures (Figure 7) are measured relative to the global minimum in each plot as listed in the table. All the position of the minima in the maps are given approximately as they vary in different maps.

^aIn this case, the minimum is around (7.7,6.2).

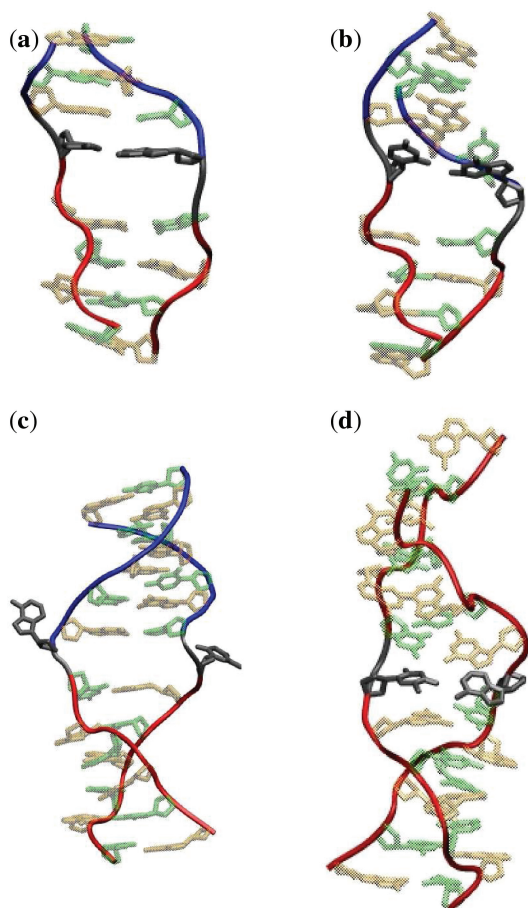


Figure 6. Different conformations of B–Z and Z–Z junctions. The top structures show a B–Z junction for a $(CG)_4$ duplex (a) with WC base pairing; and (b) without WC base pairing. The bottom figures show structures for a $(CG)_3 A(CG)_3$ duplex with (c) a B–Z junction and (d) a Z–Z junction. Color scheme is similar to that of Figure 2, with adenine and thymine bases colored gray.

any hydrogen bond at all, i.e. similar to Figure 8d. We also observe a local minimum centered at (6,3), corresponding to a structure with a single hydrogen bond between the O_4 and N_6^A of the thymine and adenine bases (Figure 8c). This configuration resembles the experimentally observed structure of the HEPES-free Z–Z junction (27). Finally, the AT base pair inserted in a B–Z junction shows no preference for any kind of hydrogen bonding. In this case, the global minima are clustered in a valley centered around (11,11), which is just too far for any significant hydrogen bonding. Figure 6c and d shows sample configurations of such B–Z and Z–Z junctions. To obtain a rough estimate of the stacking among the two different segments of the junctions, we measured the angle between the third principal axis components of the two segments of DNA. This strategy defines the ‘kink’ angle that has been reported experimentally (27). We applied this definition to the ‘non-equilibrium’ structures, just to get an estimate of the ‘kink’ due to the junctions. In the Z–Z junctions, 90% of the data fall between 11° and 37° with an average around 29° . In the B–Z junctions, 90% of the data fall between 3° and 16° with an average around 8° .

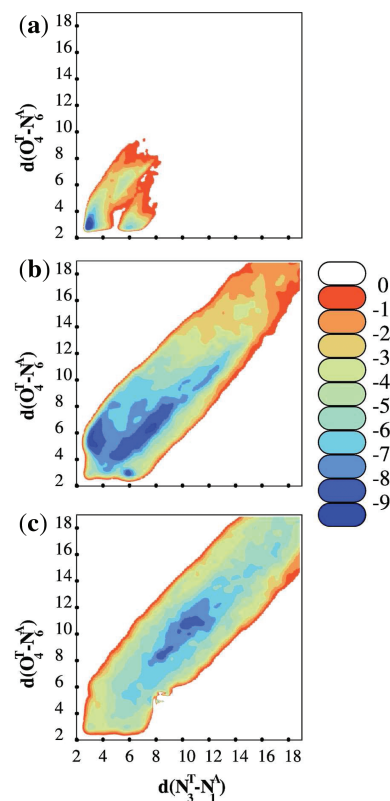


Figure 7. Free energy landscapes of an AT base pair in different conformations. Free energy maps of a $(CG)_3 A(CG)_3$ double helical DNA defined in terms of the collective variables $d(O_4^T - N_6^A)$ (i.e. the distance between the O_4^T and N_6^A of the thymine and adenine bases) and $d(N_3^T - N_1^A)$ (i.e. the N_3^T and N_1^A atoms of the thymine and adenine bases) with the rest of the structure having a conformation corresponding to (a) B-DNA; (b) Z–Z junction and (c) B–Z junction. Distances are given in Å.

These estimates indicate that the extrusion of the A and T bases in the B–Z junction causes only a small kink angle, thus tending to preserve the base pair stacking; whereas the lack of base pair extrusion in the Z–Z junction tends to disrupt the base pair stacking (experimentally, the kink angle in the Z–Z junction is 27°).

DISCUSSION

The atomistic study of the B–Z-DNA transition is quite challenging. Experimentally, crystal structures are used to obtain atomic details of the static structures, and techniques involving thermodynamics, circular dichroism and nuclear magnetic resonance (NMR) determine general features of the transition, but the atomistically detailed dynamics have remained elusive. Fully atomistic MD becomes invaluable in complementing the experimental data and addressing the inherent structural complexity of the B–Z transition, but these studies have been slow to emerge. One of the main difficulties is given by the time scales involved: a single base flipping event takes from milliseconds to seconds, and a full B–Z-DNA transition of a short oligomer occurs on a timescale from seconds to minutes. Thus, in addition to accurate force fields,

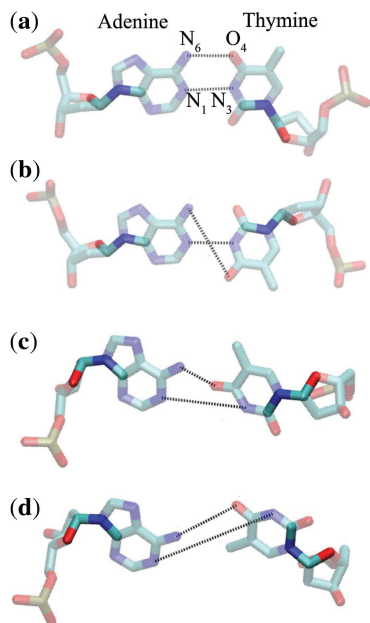


Figure 8. Different forms of AT base pairing. Schematic representation of the different conformations of the AT base pair in a $(CG)_3A(CG)_3$ DNA duplex. Here, (a) shows WC base pairing ($N_1^A-N_3^T$ and $N_6^A-O_4^T$); (b) reverse WC base pairing ($N_1^A-N_3^T$ and $N_6^A-O_2^T$); (c) base pairing with only a single hydrogen bond ($N_6^A-O_4^T$) and (d) no base pairing, respectively. The distance between the O_4 of the thymine base and N_6 of the adenine base, and the distance between N_3 of the thymine base and N_1 of the adenine base are explicitly marked. The glycosidic angles are also highlighted.

atomistic MD needs to be supplemented with special enhanced sampling techniques in order to cover the required timescales.

Three relatively recent computational studies address the B–Z-DNA transition. Lim and Feng (19,26) used stochastic difference equations (which find the minimum potential path based on the principle of minimum action) to look at the heptamer duplex $d(GCGCGCG)_2$ [same as $d(CGCGCGC)_2$] in implicit solvent (19); Kastenzholz *et al.* (24) used targeted molecular dynamics (TMD) to look at the hexamer $d(GCGCGC)_2$ in explicit water; and Lee *et al.* (25) used TMD and umbrella sampling to study the propagation of a single step in the nanomer $d(GCGCGCGCG)_2$ [same as $d(CGCGCGCGC)_2$] also in explicit solvent. In a similar fashion, we considered various oligomers in aqueous solution formed by traditional bases (not chemically modified) and ignored interactions with other molecules. Our study, like that of Lim and Feng, was carried out in implicit solvent, and therefore ignored salt effects. Interestingly, the study by Kastenzholz *et al.* found the same stretch intermediate mechanism first proposed by Lim and Feng, in spite of the different force field, the different computational methodology, and the different approach to solvation. In addition, the transition pathway also appeared to be qualitatively insensitive to the ionic-strength conditions. These results are very encouraging for our own work, where the computational time saved by using the implicit solvent was employed in exploring different regions of phase space via various relevant order parameters, which would have been prohibitively expensive with explicit waters.

The use of multiple order parameters in this work not only allows for a more extensive exploration of the phase space but also unifies the apparently contradictory results between different models for the B–Z-DNA transition. The driving force behind the use of the TMD method in previous studies is based on the argument that the reaction coordinate for the transition is unknown. However, there are a few restrictions involved in the use of TMD. The TMD algorithm allows for the characterization of a gradual conformational change between an initial and a reference structure. A restraining potential for a conformation \mathbf{x} at time t proportional to $[R(\mathbf{x},t) - R_T(t)]^2$ is applied to the system. Here, $R(\mathbf{x},t)$ is the root mean square deviation (RMSD) of the conformation \mathbf{x} from the reference structure, and $R_T(t)$ is the target RMSD value that decreases linearly with time to drive the system from the initial to the reference state. The RMSD imposes constraints on all the atomic coordinates and it is forced to vary linearly; and therefore can potentially be quite constraining for certain transition pathways. Thus, undesirable numerical artifacts can affect the activation energy and the entropy, especially in the transition state. In addition, the free energy associated with the RMSD reaction coordinate tends to diverge to infinity at the end point of the transition, where there is an artificial entropy reduction upon decreasing the RMSD (58,59). Our approach involves the definition of relevant collective variables, which are crucial for the enhanced sampling techniques used. In particular, we find that our definition of the two-dimensional order parameter (H, R_g) is both adequate and convenient. First of all, it is important to describe a symmetry-breaking transition with a collective variable that accurately reflects the symmetry-breaking properties. Handedness unequivocally characterizes the two symmetries ($H < 0$ for left-handed symmetry, $H > 0$ for right-handed symmetry, and $H = 0$ for structures with no net handedness such as transition states and S-DNA). As extensively discussed in the Supplementary Methods, handedness is considerably more convenient than the base pair step parameter of twist (which, among other things, is not well defined for a broken base pair), and quantitatively quite close to the overall helical twist. The addition of the radius of gyration then provides for a phase-space (H, R_g) with enough complexity to contain standard and less standard DNA conformations. Although the minima corresponding to B- and Z-DNA are readily recognized, we also observe structures corresponding to A-DNA, S-DNA and even a transient Z[WC]-type DNA (along with other ‘no-name’ structures). The following is a description of our main results.

Mechanisms for the B–Z-DNA transition

In the (H, R_g) free energy map for the dodecamer $(CG)_6$ (Figure 2), the minimum corresponding to Z-DNA is (9 ± 2) kcal/mol higher than the global minimum corresponding to B-DNA. This is ~ 1.5 kcal/mol per dinucleotide, which is consistent with experimental data that give 0.66 kcal/mol per CG repeat (17), and in general, < 2 kcal/mol per dinucleotide repeat (60). Lee *et al.* (25) obtained 0.93 kcal/mol per dinucleotide with the base

extrusion zipper mechanism at 0.1 M salt concentration in explicit water. As our simulations exclude salt, the relative free energy of Z-DNA with respect to B-DNA is expected to be higher, and our results reflect that. We note that Kastenholz *et al.* obtained an energy difference between Z-DNA and B-DNA that varies between ~ 10 kcal/mol for the 4 M solution and 14 kcal/mol for the 0 M solution per dinucleotide, which considerably exceeds the experimental numbers. The LFEP algorithm identifies two pathways in the free energy map that are associated with a stretch-collapse and with a zipper mechanism. Free energies associated with these paths (Figure 4) are very similar for both mechanisms and thus, at least for this polymer length, there is no reason to favor one over the other. The largest free energy barrier in both cases is about (19 ± 2) kcal/mol, which is roughly in agreement with recent results (25) (the height and shape of the barriers are not entirely comparable, as the results from Lee *et al.* involve two structures where an existing junction is displaced by a single dinucleotide, whereas our studies involve the entire dodecamer including the end cases without junctions). We have used SMD to study the transition mechanisms along these two paths and below we discuss the results in more detail.

Stretch-collapse mechanism.

The mechanism along the solid line LFEP path in the (H, R_g) free energy landscape essentially corresponds to that proposed by Lim and Feng (19,26) and Kastenholz *et al.* (24). The main features of the model are the formation of a stretched intermediate conformation with parallel phosphate backbones, and the refolding of this intermediate structure into Z-DNA. The stretching of DNA [notice the larger radius of gyration in the (H, R_g) landscape] reduces steric constraints: the increase in rise prevents clashes between the bases as they perform the large amplitude motions necessary for rotations; and the stretching of the backbone decreases the backbone energy thus compensating for the increase in energy due to conformational changes in the nucleotides. One difference between our study and the previous studies is that we observe relatively little disruption of the base pairs that maintain their WC hydrogen bonds. This means that the stretched intermediate with its parallel strands corresponds to the 'S-DNA conformation' (56). Note that the energy barriers obtained in the previous studies are considerably higher than the values we measure. This could be partly due to the disruption of the WC base pairs. In addition, (19) (that uses the same force field and implicit solvent) represents a solution to a zero-temperature pathway, whereas (24) (which measures barriers of about 30 kcal/mol for the 4 M solution and 36 kcal/mol for the 0 M solution, for an hexadecamer) could suffer from inaccuracies in the force field or the TMD algorithm (because the numbers do not agree with experimental numbers). It is interesting that in a seemingly different context, our observations from the (H, R_g) -based simulations are in line with the conclusions (61) from recent experimental studies that there are two competing modes of over-stretching for a DNA with free ends: (i) formation of a S-DNA, associated with a reversible fast dynamics

and (ii) unpeeling of one strand from the other starting from one end, associated with an irreversible slow dynamics. Even though the selection of these modes depends on salt concentration, DNA sequence and temperature, the point at which S-DNA and unpeeling are in coexistence is quite close to physiological conditions (57). Although we observe both phenomena in our simulations, the B-Z transition can take place only when the S-DNA formation mode is dominant, if stretching occurs. It is shown experimentally (61) that high salt concentration and CG sequence (i.e. conditions that stabilize Z-DNA) both favor the S-DNA formation mode over unpeeling. Therefore, the stretch-collapse mechanism of B-Z transition with a S-DNA-like intermediate seems to be plausible for a short CG-based DNA with free ends.

Base extrusion zipper mechanism.

The dotted line LFEP on the (H, R_g) free energy landscape corresponds to the zipper mechanism first proposed by Ha *et al.* (21). The zipper mechanism involves the sequential propagation of a B-Z junction involving two base pairs with extrusion out of the helix core. Our results are in qualitative agreement with the results of Lee *et al.* (25), who carried out an explicit solvent simulation of an 'existing' B-Z junction moving through a single dinucleotide step. Their estimated transition barrier for a heptadecamer was 13 kcal/mol, which at the time was considerably less than the reported barrier for the stretched intermediate model; leading the authors to conclude that the zipper mechanism is a highly probable pathway for the B-Z transition. Our simulations (which for the results in Figures 2 and 3 also involve the nucleation of the initial B-Z junction) indicate that this is just one of the possible pathways associated with the transition. The zipper mechanism was further confirmed by our simulations employing the (N_B, N_Z) order parameter [specifying the number of base pairs in the B- or Z-conformations] that unlike the order parameter used by Lee *et al.* (25)] is completely degenerate with respect to the order in which the bases flip.

In spite of being quite different, both mechanisms are compatible with a B-Z-DNA transition. The simulations are carried out in short oligomers with free ends, and do not address the problem of how the transition begins in a very long or a circular segment of B-DNA *in vivo*. Biomolecules such as helicases or negative supercoiling can unwind the helix and thus favor a stretch intermediate mechanism, whereas other molecules such as ADAR1 may favor the base extrusion mechanism. It has been argued that the stretch model (that requires the concerted deformation of the entire backbone and rotation of base pairs) may not be an adequate model to describe the B-Z transition, as the barrier of the transition can be expected to increase linearly with the length of the DNA, leading to an exponential slow growth of the transition. However, this is not necessarily the case if one generalizes the concept of a B-Z junction. The junction can be defined as a DNA stretch where at least one base pair has neither the B nor the Z conformation. The junction can be as small as 2-3 bp, such as in the zipper mechanism described above, or as large as 10-12 bp [examples of various lengths for the B-Z junctions as measured experimentally are given in

Ref. (15)]. This means that the stretch intermediate structure could be considered as a B–Z junction, at least for the nucleation stage. So, unless more is known about how the transition is initially triggered, none of the alternative mechanisms can be ruled out on the basis of energetics alone.

B–Z and Z–Z-DNA junctions

Convenient definitions of new order parameters allow for the characterization of the B–Z and Z–Z junctions. From Figure 5, it is clear that neither Z-DNA nor B-DNA favor the nucleation of a pair of opposite handedness in the middle of the duplex. However, in both cases there is the possibility of base pair disruption ($N_{WC} = 0$) that might ultimately result in a nucleation event. Free energy maps associated with $(CG)_3A(CG)_3$ structures indicate that the B–Z junction precludes hydrogen bonding for the AT pair, but the Z–Z junction allows for several possibilities, although the WC base pairing is not likely. In the B–Z junction, an estimate of the average kink angle for the B–Z junction gives a small value of $\sim 8^\circ$, which indicates that the base stacking between the two helical segments is mainly preserved. In the Z–Z junction, the kink angle is estimated as $\sim 29^\circ$, which indicates a partial disruption of the base stacking between the two helical segments. This is in agreement with the conclusions of recent experimental studies (21,27). The experimental results show that a Z–Z junction stabilized by $Z\alpha$ and in presence of a HEPES molecule is in reverse WC base pairing (as in our Figure 8b), whereas a HEPES-free structure is more likely to have one (as in Figure 8c) or zero (as in Figure 8d) hydrogen bonds. As we do not have an intercalated HEPES molecule, the reverse WC pairs cannot be directly compared. In the experiment, the AT base pairs are oriented perpendicular to the plane of the other bases. In the simulation, both A and T bases in the reverse WC pairs tend to be in the *anti* conformation. When the hydrogen bonds are broken, adenine tends to be in the *syn* conformation, whereas the thymine bases switch between *anti* (especially in the case with one H-bond) and *syn* conformations. These results are also in agreement with the crystal structure results.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Methods, Supplementary Table 1, Supplementary Figures 1–5 and Supplementary References [62–69].

ACKNOWLEDGEMENTS

We thank the NC State HPC Center for extensive computational support.

FUNDING

The National Science Foundation [NSF-1021883, NSF-1148144 to C.S. and C.R.]. Funding for open access charge: NSF grants.

Conflict of interest statement. None declared.

REFERENCES

- Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, **282**, 680–686.
- Schroth, G., Chou, P. and Ho, P.J. (1992) Mapping Z-DNA in the human genome – computer-aided mapping reveals a nonrandom distribution of potential Z-DNA forming sequences in human genes. *J. Biol. Chem.*, **267**, 11846.
- Oh, D., Kim, Y. and Rich, A. (2002) Z-DNA binding proteins can act as potent effectors of gene expression in vivo. *Proc. Natl Acad. Sci. USA*, **99**, 16666.
- Lipps, H.J. (1983) Antibodies against Z-DNA react with the macronucleus but not the micronucleus of the hypotrichous ciliate *Stylonychia mytilus*. *Cell*, **32**, 435–441.
- Herbert, A. and Rich, A. (1999) Left-handed Z-DNA: structure and function. *Genetica*, **106**, 37.
- Lancillotti, F., Lopez, M., Alonso, C. and Stollar, B. (1985) Locations of Z-DNA in polytene chromosomes. *J. Cell Biol.*, **100**, 1759.
- Rich, A., Nordheim, A. and Wang, A.H. (1984) The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.*, **53**, 791–846.
- Kmieć, E., Angelides, K. and Holloman, W. (1985) Left-handed DNA and the synaptic pairing reaction promoted by *ustilago* re1 protein. *Cell*, **40**, 139.
- Blaho, J. and Wells, R. (1987) Left-handed Z-DNA binding by the *reca* protein of *Escherichia coli*. *J. Biol. Chem.*, **262**, 6082.
- Jaworski, A., Hsieh, W., Blaho, J., Larson, J. and Wells, R. (1987) Left-handed DNA in vivo. *Science*, **238**, 773.
- Schwartz, T., Rould, M., Lowenhaupt, K., Herbert, A. and Rich, A. (1999) Crystal structure of the Z alpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science*, **284**, 1841.
- Vinogradov, A. (2003) DNA helix: the importance of being GC-rich. *Nucleic Acids Res.*, **31**, 1838.
- Champ, P., Maurice, S., Vargason, J., Champ, T. and Ho, P. (2004) Distributions of Z-DNA and nuclear factor I in human chromosome 22; a model for coupled transcriptional regulation. *Nucleic Acids Res.*, **32**, 6501.
- Wang, G., Christensen, L. and Vasquez, K. (2006) Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl Acad. Sci. USA*, **103**, 2677.
- Fuertes, M.A., Cepeda, V., Alonso, C. and Pérez, J.M. (2006) Molecular mechanisms for the B-Z transition in the example of poly[d(G-C) · d(G-C)] polymers a critical review. *Chem. Rev.*, **106**, 2045–2064.
- Harvey, S. (1983) DNA structural dynamics - longitudinal breathing as a possible mechanism for the B-reversible-Z transition. *Nucleic Acids Res.*, **11**, 4867.
- Goto, S. (1984) Characterization of intermediate conformational states in the B ↔ Z transition of poly(dG - dC) · poly(dG - dC). *Biopolymers*, **23**, 2211.
- Ansevin, A. and Wang, A. (1990) Evidence for a new Z-type left-handed DNA helix - properties of Z(WC)-DNA. *Nucleic Acids Res.*, **18**, 6119.
- Lim, W. and Feng, Y. (2005) The stretched intermediate model of B-Z DNA transition. *Biophys. J.*, **88**, 1593.
- Saenger, W. and Hienemann, U. (1989) Raison d'être and structural model for the B-Z transition of poly d(G-C)*poly d(G-C). *FEBS Lett.*, **257**, 223.
- Ha, S.C., Lowenhaupt, K., Rich, A., Kim, Y.G. and Kim, K.K. (2005) Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature*, **437**, 1183.
- Elber, R., Ghosh, A. and Cardenas, A. (2002) Long time dynamics of complex systems. *Acc. Chem. Res.*, **35**, 396.
- Maji, J. and Bhattacharjee, S.M. (2010) Dynamic phase transition in the conversion of B-DNA to Z-DNA. *Europhys. Lett.*, **92**, 58004.
- Kastenholz, M.A., Schwartz, T.U. and Hünenberger, P.H. (2006) The transition between the B and Z conformations of DNA

- investigated by targeted molecular dynamics simulations with explicit solvation. *Biophys. J.*, **91**, 2976–2990.
25. Lee, J., Kim, Y., Kim, K. and Seok, C. (2010) Transition between B-DNA and Z-DNA: free energy landscape for the B-Z junction propagation. *J. Phys. Chem. B*, **114**, 9872.
 26. Lim, W. and Feng, Y. (2005) Applying the stochastic difference equation to DNA conformational transitions: a study of B-Z and B-A DNA transitions. *Biopolymers*, **78**, 107.
 27. de Rosa, M., de Sanctis, D., Rosario, A.L., Archer, M., Rich, A., Athanasiadis, A. and Carrondo, M.A. (2010) Crystal structure of a junction between two Z-DNA helices. *Proc. Natl Acad. Sci. USA*, **107**, 9088–9092.
 28. Babin, V., Roland, C. and Sagui, C. (2008) Adaptively Biased Molecular Dynamics for free energy calculations. *J. Chem. Phys.*, **128**, 134101.
 29. Babin, V., Karpusenko, V., Moradi, M., Roland, C. and Sagui, C. (2009) Adaptively Biased Molecular Dynamics: An umbrella sampling method with a time-dependent potential. *Int. J. Quantum Chem.*, **109**, 3666–3678.
 30. Izrailev, S., Stepaniants, B., Isralewitz, S., Kosztin, D., Lu, H., Molnar, F., Wiggers, W. and Schulten, K. (1998) Steered molecular dynamics. In: *Computational Molecular Dynamics: Challenges, Methods, Ideas*. Springer-Verlag, Berlin, Germany, pp. 39–65.
 31. Huber, T., Torda, A.E. and van Gunsteren, W.F. (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J. Comput. Aided Mol. Des.*, **8**, 695–708.
 32. Laio, A. and Parrinello, M. (2002) Escaping free-energy minima. *Proc. Natl Acad. Sci. USA*, **99**, 12562–12566.
 33. Iannuzzi, M., Laio, A. and Parrinello, M. (2003) Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys. Rev. Lett.*, **90**, 238302.
 34. Bussi, G., Laio, A. and Parrinello, M. (2006) Equilibrium free energies from nonequilibrium Metadynamics. *Phys. Rev. Lett.*, **96**, 090601.
 35. Lelievre, T., Rousset, M. and Stoltz, G. (2007) Computation of free energy profiles with parallel adaptive dynamics. *J. Chem. Phys.*, **126**, 134111.
 36. Frenkel, D. and Smit, B. (2002) *Understanding Molecular Simulation. Computational Science Series*. Academic Press, Waltham, Massachusetts, USA.
 37. Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Crowley, M., Walker, R.C., Zhang, W. et al. (2008) *AMBER 10*. University of California, San Francisco.
 38. Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M. et al. (2010) *AMBER 11*. University of California, San Francisco.
 39. Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M. et al. (2012) *AMBER 12*. University of California, San Francisco.
 40. Babin, V., Roland, C., Darden, T.A. and Sagui, C. (2006) The free energy landscape of small peptides as obtained from Metadynamics with umbrella sampling corrections. *J. Chem. Phys.*, **125**, 2049096.
 41. Babin, V. and Sagui, C. (2010) Conformational free energies of methyl- α -l-iduronic and methyl- β -d-glucuronic acids in water. *J. Chem. Phys.*, **132**, 104108.
 42. Moradi, M., Babin, V., Roland, C., Darden, T. and Sagui, C. (2009) Conformations and free energy landscapes of polyproline peptides. *Proc. Natl Acad. Sci. USA*, **106**, 20746.
 43. Moradi, M., Babin, V., Roland, C. and Sagui, C. (2010) A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers. *J. Chem. Phys.*, **133**, 125104.
 44. Moradi, M., Lee, J.G., Babin, V., Roland, C. and Sagui, C. (2010) Free energy and structure of polyproline peptides: an *ab initio* and classical molecular dynamics investigation. *Int. J. Quantum Chem.*, **110**, 2865–2879.
 45. Moradi, M., Babin, V., Sagui, C. and Roland, C. (2011) A statistical analysis of the PPII propensity of amino acid guests in proline-rich peptides. *Biophysical J.*, **100**, 1083–1093.
 46. Moradi, M., Babin, V., Sagui, C. and Roland, C. (2011) PPII propensity of multiple-guest amino acids in a proline-rich environment. *J. Phys. Chem. B*, **115**, 8645–8656.
 47. Moradi, M., Babin, V., Roland, C. and Sagui, C. (2012) Are long-range structural correlations behind the aggregation phenomena of polyglutamine diseases? *PLoS Comput. Biol.*, **8**, e1002501.
 48. Moradi, M., Sagui, C. and Roland, C. (2011) Calculating relative transition rates with driven nonequilibrium simulations. *Chem. Phys. Lett.*, **518**, 109–113.
 49. Onufriev, A., Bashford, D. and Case, D.A. (2000) Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **104**, 3712–3720.
 50. Onufriev, A., Bashford, D. and Case, D.A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins*, **55**, 383–394.
 51. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
 52. Case, D.A., Cheatham, T.E. III, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The AMBER biomolecular simulation programs. *J. Computat. Chem.*, **26**, 1668–1688.
 53. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
 54. Lu, X. and Olson, W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.
 55. Ensing, B., Laio, A., Parrinello, M. and Klein, M. (2005) A recipe in the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B*, **109**, 6676–6687.
 56. Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J.L., Chatenay, D. and Caron, F. (1996) DNA: an extensible molecule. *Science*, **271**, 792.
 57. Cocco, S., Yan, J., Leger, J., Chatenay, D. and Marko, J. (2004) Overstretching and force-driven strand separation of double-helix DNA. *Phys. Rev.*, **70**, 011910.
 58. Schlitter, J. (1993) Estimation of absolute and relative entropies of macromolecules using the covariance-matrix. *Chem. Phys. Lett.*, **215**, 617.
 59. Schlitter, J., Swegat, W. and Mulders, T. (2001) Distance-type reaction coordinates for modelling activated processes. *J. Mol. Mod.*, **7**, 171.
 60. Ho, P. (1994) The non-B-DNA structure of d(CA/TG)(N) does not differ from that of Z-DNA. *Proc. Natl Acad. Sci. USA*, **91**, 9549.
 61. Fu, H., Chen, H., Marko, J. and Yan, J. (2010) Two distinct overstretched DNA states. *Nucleic Acids Res.*, **38**, 5594.
 62. Silverman, B.W. (1986) Density Estimation for Statistics and Data Analysis. In: *Monographs on Statistics and Applied Probability*. Chapman and Hall.
 63. Raiteri, P., Laio, A., Gervasio, F.L., Micheletti, C. and Parrinello, M. (2006) Efficient reconstruction of complex free energy landscapes by multiple walkers Metadynamics. *J. Phys. Chem.*, **110**, 3533–3539.
 64. Jarzynski, C. (1997) Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, **78**, 2690–2693.
 65. Crooks, G.E. (2000) Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, **61**, 2361–2366.
 66. Shirts, M.R., Bair, E., Hooker, G. and Pande, V.S. (2003) Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.*, **91**, 140601.
 67. Kosztin, L., Barz, B. and Janosi, L. (2006) Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality. *J. Chem. Phys.*, **124**, 064106.
 68. Sinden, R. (1994) *DNA Structure and Function*. Academic Press, Waltham, Massachusetts, USA.
 69. Tepper, H.L. and Voth, G.A. (2005) A coarse-grained model for double-helix molecules in solution: Spontaneous helix formation and equilibrium properties. *J. Chem. Phys.*, **122**, 124906.