Stretching and twisting of the DNA duplexes in coarse-grained dynamical models

2009 J. Phys.: Condens. Matter 21 474221

(http://iopscience.iop.org/0953-8984/21/47/474221)

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **21** (2009) 474221 (14pp) doi:10.1088/0953-8984/21/47/474221

Stretching and twisting of the DNA duplexes in coarse-grained dynamical models

Szymon Niewieczerzał and Marek Cieplak

Institute of Physics, Polish Academy of Science, Aleja Lotników 32/48, 02-668 Warsaw, Poland

Received 30 April 2009, in final form 25 June 2009
Published 5 November 2009
Online at stacks.iop.org/JPhysCM/21/474221

Abstract

Three coarse-grained molecular dynamics models of the double-stranded DNA are proposed and compared in the context of single molecule mechanical manipulation such as twisting and various schemes of stretching—unzipping, shearing, two-strand stretching and stretching of only one strand. The models differ in the number of effective beads (between two and five) representing each nucleotide. They all show similar behaviour, but the bigger the resolution, the more details in the force patterns. The models incorporate the effective Lennard-Jones potentials in the couplings between two strands and harmonic potentials to describe the structure of a single strand. The force patterns are shown to depend on the sequence studied. In particular, both shearing and unzipping for an all-AT sequence lead to lower forces than for an all-CG sequence. The unzipping patterns and the corresponding scenario diagrams for the contact rupture events are found to reflect the sequential information if the temperature is moderate and initial transients are discarded. The derived torque–force phase diagram is found to be qualitatively consistent with experiments and all-atom simulations.

This paper is dedicated to Dr Richard Palmer on his retirement from *Journal of Physics: Condensed Matter*.

1. Introduction

Manipulation of large biomolecules by means of atomic force microscopy, optical tweezers and other nanotechnological devices is playing an increasingly growing role in elucidating mechanisms of biologically relevant processes [1–3]. The dynamical data obtained through mechanical manipulation usually requires theoretical interpretation that can be reached through numerical simulations. This need is especially apparent when dealing with proteins (see, e.g., [4]) because of their strongly inhomogeneous network of interactions between the amino acids. The inhomogeneity results, for instance, in a protein-dependent pattern of peaks when the force of resistance to pulling at a constant speed is plotted against elongation (see, e.g. [5–8]).

All-atom simulations have contributed to an understanding of the large conformational changes in proteins induced by the mechanical manipulation (see, e.g., [9–11]). However, such simulations are inherently limited by the timescales and system sizes that can be studied. One way out is to accelerate the modelled processes by orders of magnitude relative to their experimental realizations, which may, undesirably, distort the physics involved. Another way is to use coarse-grained models. These models can also be applied to other processes involving large conformational changes such as folding or thermal unfolding of biomolecules. They may also find applications in studies of systems which are much larger in size.

In this paper, we focus on coarse-grained models describing the double-stranded DNA (dsDNA). The simplest coarse-graining scheme involves treating dsDNA as a polymer endowed with a local stiffness [12]. The resulting model seems to be appropriate in studies of stretching at high temperatures, when the entropic effects dominate, and in studies of phenomena related to fluid flow and the hydrodynamic interactions [13]. The corresponding characteristics dimension in this model can be measured by the Kuhn length or the hydrodynamic radius. The typical values of these parameters are of the order of 106 and 77 nm [13], respectively, which encompasses more than 200 distances (of 0.4 nm) between successive pairs of nucleotides in the dsDNA. This level of coarse-graining is thus too crude to be useful when describing

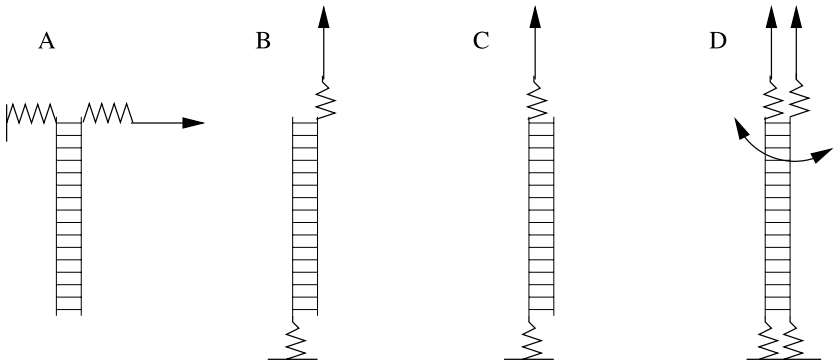

Figure 1. Four possibilities of manipulation of the DNA double helix.

mechanical manipulation of the system where nanometric features are detectable. Here, we discuss coarse-grained models with structures that are resolved at the level of a single nucleotide.

There are a variety of possible ways to manipulate the DNA duplex. The ones that we consider in this paper are schematically represented in figure 1. Unzipping, whether at constant speed or at constant force, corresponds to scheme A in this figure. The unzipping process involves breaking of one hydrogen bond at a time, resulting in a force, F, of the order of 13–15 pN, that rapidly undulates with the amplitude of the order of 1 pN as the pulling distance, d, is increased [14, 15]. The force pattern has also been found to depend weakly on the pulling speed, v_p [14], but would be expected to depend on temperature, T, more substantially. Finally, the dynamics should depend on the sequential details as it takes less force to unravel the two hydrogen bonds in the A–T pair compared to the three bonds in the G–C pair. Essevaz-Roulet *et al* [16] have estimated the average force for unzipping of the pure A–T and pure G–C double strands to be 10 pN and 15 pN, respectively. The corresponding values estimated by Rief *et al* [17] are 9 ± 3 pN and 20 ± 3 pN, respectively. Rief *et al* have also pointed out that such values must also depend on the precise definitions of the strings of the nucleotides.

Even though the variations in the F–d patterns when pulling at constant speed take place on the nucleotidic length scales, their interpretation in terms of the specifics of the sequence is difficult because of the noise due to thermal fluctuations. Recently, Baldazzi *et al* [18] has suggested, however, that if instead the mechanical unzipping is performed at constant force then Bayesian methods of the corresponding sequence prediction should be nearly error-free. Methods for extracting kinetic information from constant force experiments have been discussed recently in the context of DNA unzipping in a nanopore [19].

Schemes B and C involve stretching at the opposite ends of the duplex. The two schemes employ distinct mechanisms of resistance to stretching. Mechanism B involves shear which is responsible for generation of the strongest force clamps in biomolecules [4] and the maximum force obtained depends on the number of the bonds that are sheared simultaneously. Mechanism C, on the other hand, leads to localized unravelling and generates a size-independent force. When using micropipettes, as in the experiment by Cluzel *et al* [20], one probably combines schemes B and C. The resulting F–d pattern has three stages: one starts off with a long period of a nearly constant force, after which there is a steady increase (in the 120 pN range) which finally is followed by a sudden drop to zero. A similar pattern of behaviour arises in simulations involving anisotropic pressure [21] (these studies have been performed for a dsDNA with about 10 pairs of nucleotides).

Still another way of manipulating the dsDNA has been employed by Oroszi *et al* [22] and it involves applying a torque, G. Wereszczynski and Andricioaei [23] have generalized it still further, in their all-atom simulations, by considering a simultaneous application of a force and torque, as shown in scheme D in figure 1. They have predicted the existence of a rich phase diagram of possible structures on the F–G plane. There have been experimental studies involving torque produced in an optical trap [24, 22, 27]. Bryant *et al* [24] have found that the torque needed to transform the B-DNA conformation into the left-hand twisted L-DNA form is $G = -9.6$ pN nm (the negative sign indicates twisting against the native sense of turn in the dsDNA), while to transform it further into the Pauling-like (P-DNA) form the required torque is 34 pN nm. The experiments also reveal that the dsDNA molecule, when pulled (as in scheme D) with a force of about 65 pN [20, 25], undergoes transition into a stable overstretched form, which is about 60% longer then the B-DNA. At forces exceeding 150 pN, the dsDNA melts into two strands in an non-equilibrium way [26] with a significant force-loading-rate dependence and hysteresis. The exact structure of the stable overstretched form is not yet known. However, there are two hypotheses about it. The first is that both strands separate and stay away from one another as two ssDNA molecules. The second is that the B-DNA transits into a ladder-like dsDNA structure, where the hydrogen bonds remain intact, while the stacking interactions are largely disrupted. This form is denoted as the S-DNA structure. The recent experimental results [28] favour the first hypothesis since they suggest that the hydrogen bonds do break. Also an interesting analysis is presented in a theoretical study by Marenduzzo *et al* [29]. Their model involves self-avoiding coarse-grained polymers on a three-dimensional lattice. The lattice polymers are described in terms of the monomer size (the length of one base pair) and

the persistence length. When these parameters are selected as equal for both dsDNA and ssDNA then the stretching force is found to stabilize the double-stranded DNA structure (beneath the melting transition), independent of the applied force. On the other hand, the strands may separate beneath the melting temperature if the values of these parameters distinguish between dsDNA and ssDNA.

In this paper, we construct three variants of the nucleotide-based coarse-grained dynamical models of the DNA duplex. The coarse-graining method introduces several effective objects, referred to here as beads, that represent a single nucleotide. The models differ primarily by the number of beads involved. We compare the workings of the three models for a 22-base-pair system, or shorter, and use them to elucidate the mechanisms of rupture in processes corresponding to schemes A–D. One conclusion of our studies is that, even though various dynamical details differ between the models, all of them can be considered adequate and ready to be applied to larger systems. In particular, all of the coarse-grained models studied lead to a transition of the usual right-hand-twisted B-dsDNA form to the L-dsDNA form and to the P-DNA form on application of a appropriate torque. We cannot really decide about the validity of the two hypotheses regarding the nature of the stretched state. since we study short sequences and the effective potentials have longer ranges than true hydrogen bonds. Thus our model naturally leads to the overstretched DNA acquiring a form similar to that of the S-DNA.

2. An overview of the models used

Physical properties of the DNA double helix are quite distinct [30] compared to other biomolecules. Its strong stiffness comes from the braided nature of its structure combined with the presence of the base stacking interactions. Furthermore, the phosphate groups in the DNA backbone carry substantial electric charges. All of these features are employed by the cell's machinery in the processes of copying, transcribing and packaging of the DNA. For example, helicases, which unwind the double helix to provide single-stranded templates for polymerases, have evolved as motors that are capable of moving along the torsionally constrained DNA molecules. Topoisomerases break and reconnect the DNA to relieve a torsional strain that accumulates ahead of the replication fork [31, 32]. Finally, the DNA-binding proteins get docked to the DNA by means of guidance mechanisms which seem to be primarily electrostatic in nature.

Our models address the mechanical properties of the dsDNA and do not aim at determining the electrostatic potential outside of the duplex. The models are built in analogy to the Go-like models of proteins, especially in the specific implementation proposed in [33–36, 4], including in the context of topoisomerase I [37]. In the case of proteins, the model represents the system by its C^{α} atoms which are tethered together by harmonic interactions. The native contacts, such as the hydrogen bonds, are described by the Lennard-Jones potentials. The Langevin overdamped thermostat with random forces mimics fluctuational effects of the solvent. Sixty-one other variants of this basic model of a protein are discussed and compared in [38].

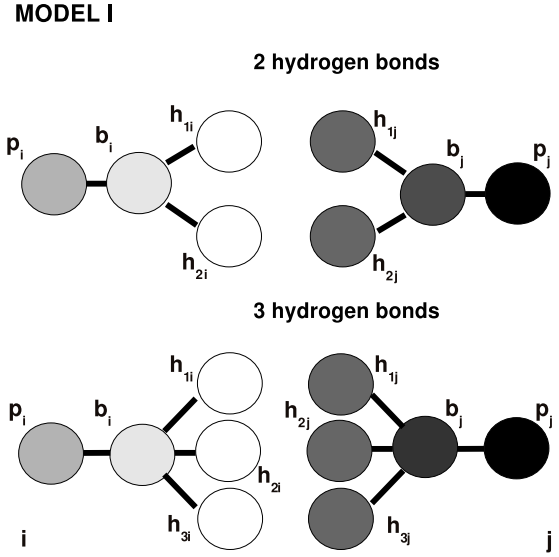

Figure 2. A schematic representation of model I of the dsDNA. It shows formation of two and three hydrogen bonds.

The dsDNA has a simpler elastic structure than proteins since a pair of nucleotides can bind only in two ways: either by forming two (A–T) or three (G–C) hydrogen bonds. When trying to build a coarse-grained model for DNA one is first inclined to assign a single bead to a nucleotide and to locate it at the phosphorus (P) atom. This may be acceptable for a single-strand DNA, provided the local chain stiffness terms are included. However, for the dsDNA this procedure would lead to a distance of 17 Å between the P atoms in a hydrogen-bonded pair of nucleotides. Such a relatively large distance would introduce too much mechanical instability into the model but appears to be adequate to study conformational changes in dsDNA nanocircles and submicron-sized plasmids with torsional stress [39]. A more detailed approach, denoted here as model I, involves representing the A and T nucleotides by four beads and the G and C nucleotides by five beads. One of the beads represents the phosphate group, another the sugar group, and the remaining beads participate in formation of either two or three hydrogen bonds, depending on the specificity. The hydrogen bond interactions are represented by the effective Lennard-Jones potentials and other bonds, being structural in nature, are described by the harmonic potentials with large elastic constants. The schematic construction of this model is shown in figure 2.

More simplified approaches involve a reduction in the number of beads representing each nucleotide. In a model denoted here as model II, we mimic the ribose–phosphate groups by one bead and the base by another bead as illustrated at the top of figure 3. In this model, the distinction between the A:T and G:C pairing interactions comes not through introduction of separate contacts for each hydrogen bond but through adjustment of the amplitude of the effective base–base Lennard-Jones potential by a factor of 2 or 3, respectively.

In between models I and II there is another model, denoted here as model III, that has been introduced by Knott *et al*

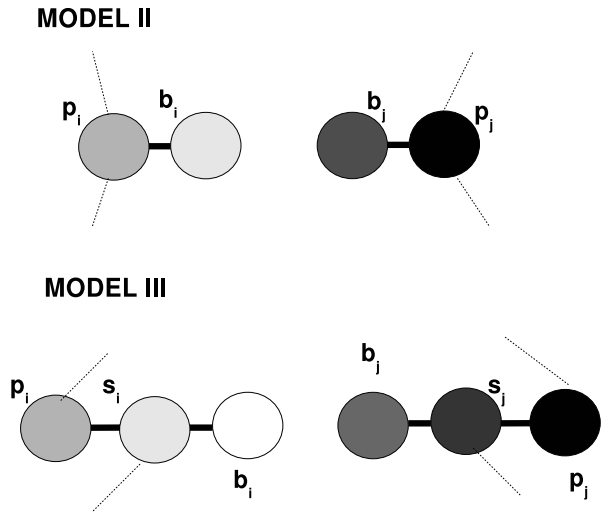

Figure 3. A schematic representation of models II and III. The thin lines indicate the way the backbone chains are constructed.

[40] in the context of the salt-induced melting. Model III involves three beads as illustrated in the bottom part of figure 3. The beads represent the phosphate, sugar and base groups correspondingly. In model III, the backbone chain of one strand is constructed by linking the sugar group to the P atom on the same nucleotide and to the P atom on a preceding nucleotide. Thus the backbone chain has the appearance of a zigzag line. In contrast, in models I and II, the backbone chain is formed by tethering the consecutive P atoms.

As a model system we consider the structure coded as 119D in the protein data bank (PDB), and shown in figure 4, which has been determined by Leonard and Hunter [41]. It corresponds to the palindromic sequence 5′-d(CGTAGATCTACGTAGATCTACG)-3′ which will be denoted here as S_1. (The letter d in the foregoing sequence signifies that in every nucleotide the sugar is deoxyribose.) Most of our results are obtained for S_1. However, when discussing the sequence dependence in schemes A and C, we also consider three other sequences. The first of these, denoted here as S_2, is the 21 base-pair long dsDNA molecule 5′-d(ACGTGATCGATGATAAGCTGT)-3′. The corresponding structure was determined by Masliah and Mauffret [42] and deposited in the PDB as 2JYK. The remaining two systems are 22 base-pair long and have been generated synthetically by an algorithm described later in this paper and which makes use of average geometrical parameters associated with a single base pair. Their sequences are poly-(dA-dT) and poly-(dG-dC). The convention used in this paper is that if a figure does not indicate a sequence then by default it relates to sequence S_1. Longer structures can be obtained, for instance, by repeating the basic units, or synthetically.

3. Model I: the 4- or 5-bead description

We start by introducing three different types of beads, p, h and b as illustrated in figure 2. The p-beads are meant to represent the backbone which is made of the phosphate and ribose groups. The p-bead is placed at the position of the C4* atom in the molecule of ribose. This placement achieves two goals. First, it represents the DNA phosphate chain by the p-beads. Second, it locates the p-bead close to the base beads. The C4* atom is the ribose ring atom that is closest to the phosphate group. The h-beads represent the 'head' atoms which may act either as donors or as acceptors in the hydrogen bonds. In the C nucleotides, the h-beads are located on the O2, N3 and N4 atoms of the bases while in the G nucleotides, on the O6, N1 and N2 atoms. Finally, in the A and T nucleotides they are on the N6, N1, and N3, O4 atoms, respectively. The h-beads are linked to their 'bases', i.e. to the supporting b-beads. In the native state, the b-beads are located halfway between the p-bead and the centre of mass of the h-beads at each nucleotide. The overall scheme results in about four- to fivefold reduction in the number of the degrees of freedom compared to the all-atom approach.

In table 1 we provide the Cartesian coordinates of the beads that define the model. Knowledge of these coordinates

Table 1. Cartesian coordinates of the beads in models I and II of the dsDNA. The coordinates depend on the identity of a nucleotide. The following description shows how to generate the DNA double helix for an arbitrary sequence. For the nucleotide, which is placed in the $(n+1)$th position in the first strand, coordinates of the related beads can be obtained through the following transformation: $x(n+1) = x\cos(n36°) - y\sin(n36°)$, $y(n+1) = x\sin(n36°) + y\cos(n36°)$, $z(n+1) = z + n3.4$, where x, y and z denote the starting coordinates as listed in the table. This transformation involves rotation around the z axis by 36°, and a shift by 3.4 Å along the z axis. The second strand in the helix is constructed in a similar way, but the initial coordinates have to be transformed from (x, y, z) into $(-x, y, -z)$ for every bead belonging to the first nucleotide of the second strand. Then one applies the prior transition to generate the positions of the sites belonging to the nucleotide paired with the $(i+1)$th nucleotide in the first strand. A similar construction for model III is presented in [40].

Site	Coordinates (Å)		
	x	y	z
Model I			
p	−7.039	−2.284	−0.492
b (A)	−3.841	−0.516	0.110
$h1$ (A)	−0.498	0.194	0.286
$h2$ (A)	−0.829	2.308	1.138
b (T)	−4.439	−0.293	−0.489
$h1$ (T)	−2.026	0.614	−0.256
$h2$ (T)	−1.692	2.782	−0.717
b (G)	−3.737	−0.997	−0.186
$h1$ (G)	−0.793	2.423	0.731
$h2$ (G)	−0.449	0.300	0.131
$h3$ (G)	−0.005	−1.851	−0.500
b (C)	−4.603	−0.802	−0.163
$h1$ (C)	−1.768	2.928	−0.103
$h2$ (C)	−2.146	0.665	0.178
$h3$ (C)	−2.524	−1.552	0.428
Model II			
p	−7.039	−2.284	−0.492
b (A)	−2.443	0.261	0.832
b (T)	−3.197	1.046	−0.351
b (G)	−2.183	0.230	0.468
b (C)	−3.153	0.825	0.139

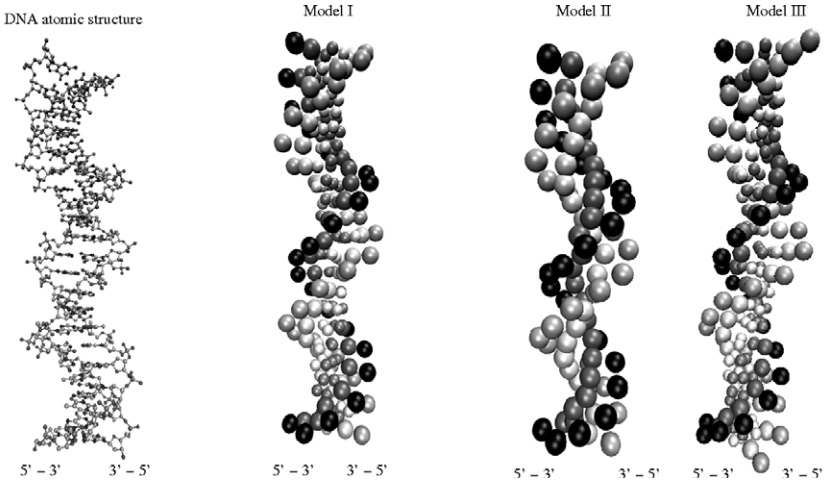

Figure 4. The atomic representation of the 119D dsDNA structure is shown on the left. The remaining panels show the corresponding coarse-grained representations considered in this paper.

should allow for a generation of a synthetic dsDNA based on the sequence. The coordinates have been obtained by making averages of the geometric parameters in the PDB structure corresponding to the basic sequence studied.

The p-beads are tethered into two separate chains and thus form two backbones. The tethering is accomplished through the elastic potential

$$V^{pp}_{i,i+1} = K_b \cdot (\vec{r}_{p_i} - \vec{r}_{p_{i+1}} - d_{p_i p_{i+1}})^2, \tag{1}$$

where the index i enumerates consecutive nucleotides and $d_{p_i p_{i+1}}$ are the distances between the consecutive p-beads in the native state. These distances vary from bead to bead. Their mean is equal to 5.8 Å and the standard deviation is close to 0.3 Å. The mean geometrical parameters cited in this description can be used in a general sequence-dependent construction of a synthetic dsDNA structure. The elastic constant is taken to be equal to $K_b = 50\varepsilon\text{Å}^{-2}$, where ε is the energy scale corresponding to the internucleotidic hydrogen bonds, as defined below.

The steric constraints of the DNA sugar–phosphate backbone are represented by the following two potentials for the bond and dihedral angles of p-beads' backbone:

$$V^{\mathrm{B}} = \sum_{i=\mathcal{N}_1}^{\mathcal{N}-2} K_\theta(\theta_i - \theta_{0i})^2, \tag{2}$$

$$V^{D} = \sum_{i=1}^{\mathcal{N}-3}[K^1_\phi(1+\cos(\phi_i - \phi_{0i})) + K^3_\phi(1+\cos 3(\phi_i - \phi_{0i}))]. \tag{3}$$

The bond angle θ_i is measured between the p_i–p_{i+1} and p_{i+1}–p_{i+2} bonds, and the dihedral angle ϕ_i is an angle between two planes: one of them is determined by the p_i–p_{i+1} and p_{i+1}–p_{i+2} bonds, and the second one by the p_{i+1}–p_{i+2} and p_{i+2}–p_{i+3} bonds, where the subscript 0 indicates the native values and $\mathcal{N}$ denotes the number of nucleotides in one chain. We take $K_\theta = 20\varepsilon/(\text{rad})^2$, $K^1_\phi = 1.0\varepsilon$, $K^3_\phi = 0.5\varepsilon$ in analogy to [43].

All of the interbead interactions within one nucleotide are taken to be harmonic so that the corresponding potentials are

$$V^{\nu\mu}_{i,i} = \sum_{\nu,\mu} K_b \cdot (\vec{r}_{\mu_i} - \vec{r}_{\nu_i} - d_{\mu_i \nu_i})^2, \tag{4}$$

where the indices ν and μ label beads belonging to the ith nucleotide. The equilibrium distances $d_{\mu\nu}$ take values as in the native structure and they range from 2.3 ± 0.1 Å for the neighbouring h-beads (according to the notation in figure 2, pairs: h_{1j}–h_{2j} and h_{2j}–h_{3j}) to 3.6 ± 0.7 Å between the b-bead and h-beads.

The h-beads on one chain are capable of making hydrogen bond contacts with the h-beads on the opposite chain. In the simplest version, we follow the prescription used previously for proteins and describe these contacts by the Lennard-Jones potential:

$$V^{hh}_{i,j} = 4\varepsilon\left[\left(\frac{\sigma_{h_i h_j}}{r_{h_i h_j}}\right)^{12} - \left(\frac{\sigma_{h_i h_j}}{r_{h_i h_j}}\right)^{6}\right], \tag{5}$$

where i and j are the paired residues and $r_{h_i h_j} = |\mathbf{r}_{h_i} - \mathbf{r}_{h_j}|$. The parameters $\sigma_{h_i h_j}$ are chosen so that each contact in the native conformation is stabilized in the minimum of the potential. Essentially, $\sigma_{h_i h_j} = 2^{-1/6} d_{h_i h_j}$. The value of $d_{h_i h_j}$, the distance between the h-beads-making bond, is equal to 2.66 ± 0.14 Å. For proteins, the choice of the form of the contact potential has turned out to be of much less importance than the correct determination of the contact map [38].

The stacking interactions between consecutive b-beads in each chain are also accounted for by the Lennard-Jones interactions:

$$V^{bb}_{i,i+1} = 4\varepsilon\left[\left(\frac{\sigma_{b_i b_{i+1}}}{r_{b_i b_{i+1}}}\right)^{12} - \left(\frac{\sigma_{b_i b_{i+1}}}{r_{b_i b_{i+1}}}\right)^{6}\right]. \tag{6}$$

The distance between the stacking pairs of the b-beads is 4.43 ± 0.42 Å in the native structure.

All of the interactions discussed above arise in the native state. However, distorted conformations may lead to

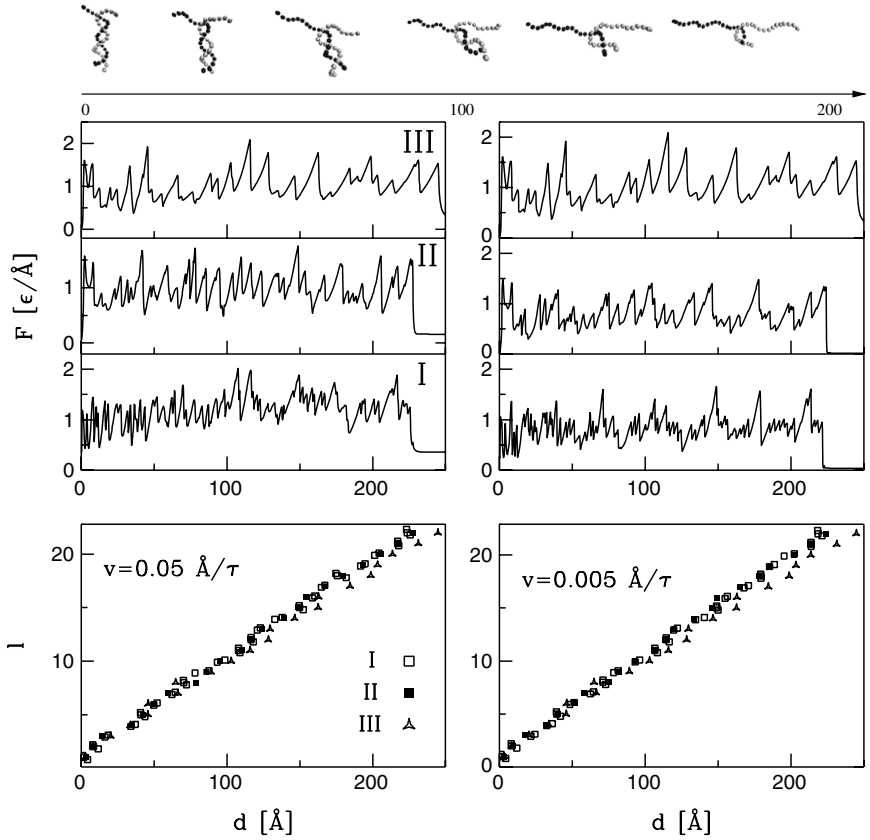

Figure 5. The A-type stretching at $\widetilde{T} = 0$. The snapshots at the top show subsequent conformations of the pulled dsDNA in model II for $v_p = 0.05$ Å$/\tau$. The panels below on the left correspond to $v_p = 0.05$ Å$/\tau$ and those on the right to $v_p = 0.005$ Å$/\tau$. The middle panels show the F–d curves for the three models as indicated. At the end of the process, the two chains get fully separated and the force drops to 0. The bottom panels show the corresponding scenarios of unfolding.

new interactions. In the spirit of the Go-like models, we describe these by purely repulsive potentials (the Lennard-Jones potentials which are cut at the minimum and shifted). The hard-sphere diameters of the h- b- and p-beads are taken to be equal to 2.0 Å, 3.4 Å and 6.0 Å, respectively. The large effective size of the p-bead prevents the chains from crossing and self-crossing.

All beads are endowed with the same mass, m, and the equation of motion of each is described by the Langevin equation

$$m\ddot{\vec{r}} = -\gamma(\dot{\vec{r}}) + \vec{F}^c + \vec{\Gamma}, \tag{7}$$

which provides thermostating and mimics dynamical effects of the solvent. Here $\vec{r}$ is the position of the bead, $\vec{F}^c$ is the net force on it due to potentials, γ is the friction coefficient and $\vec{\Gamma}$ is a white noise term with the dispersion of $\sqrt{2\gamma k_B T}$, where k_B is the Boltzmann constant and T is the temperature. The dimensionless temperature, $k_B T/\varepsilon$, will be denoted by $\widetilde{T}$. The friction coefficient γ is equal to $2m/\tau$, where τ is a characteristic timescale. The dynamics are meant to be overdamped so the characteristic timescale corresponds to a diffusional passage of a molecular distance (∼3 Å) and is thus of the order of 1 ns. For small damping, τ would correspond to a timescale of (ballistic) oscillations in the Lennard-Jones well which is significantly shorter. The equations of motion are solved by the fifth-order predictor–corrector scheme [44].

As the average value of energy for hydrogen bond interaction in dsDNA, we chose 0.6 kcal mol^{-1}, while in [45] it was chosen as around 0.5-0.7 kcal mol^{-1} and in [40]: 0.66 kcal mol^{-1}. On average, 2.5 hydrogen bonds are created between the bases in dsDNA. Hence the total average energy of interaction between paired bases in the DNA helix is about 1.5 kcal mol^{-1} in our models. This choice is consistent with $\widetilde{T} = 0.4$ corresponding to $T = 300$ K. The corresponding unit of force, ε Å^{-1} should then be of the order of 100 pN. In the entropic limit, the hydrogen bond potentials matter much less than the thermal fluctuations. In our model, this starts to happen at $\widetilde{T}$ of about 0.5–0.6.

4. Model II: the two-bead model description

In the two-bead model, we consider beads denoted by p and b at each nucleotide as shown in figure 3. The p-beads are placed in positions of the C4* atom and mimic the phosphate–ribose chain of the DNA molecule. The harmonic tethering potential, as well as the bond and the dihedral angle potentials, are introduced in analogy to model I. The b-bead in each nucleotide is placed in the geometrical centre of the base. The absence of the h-beads of model I is compensated by introducing hydrogen-bond-like interactions between the b-beads. The average Cartesian coordinates of the model beads

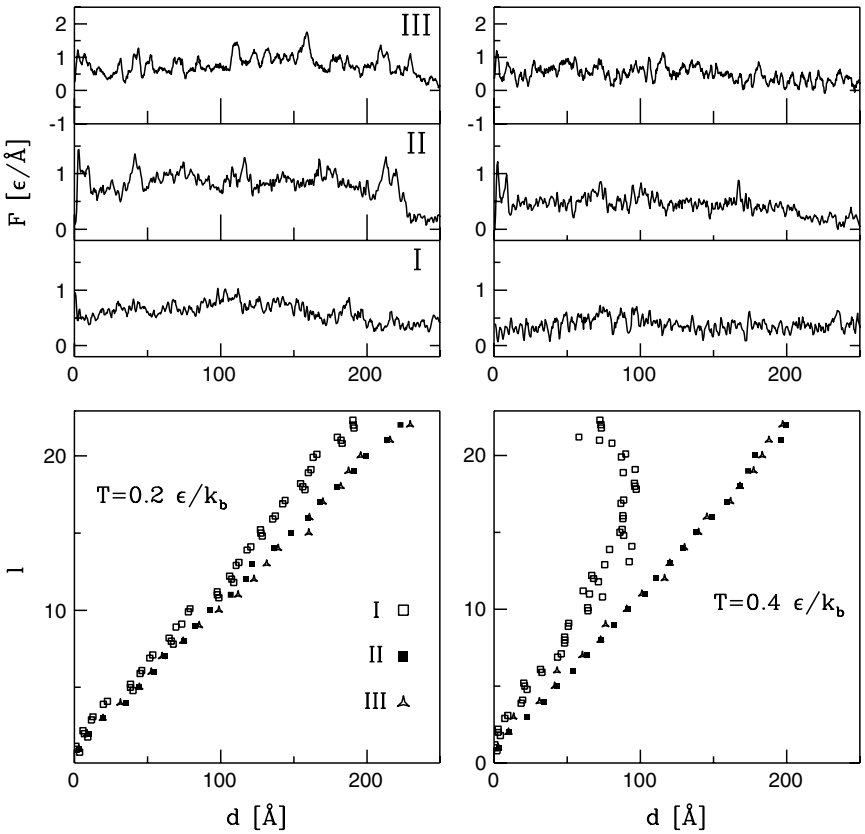

Figure 6. Stretching in the A-type mode at $\widetilde{T} = 0.2$ (panels on the left) and at $\widetilde{T} = 0.4$ (panels on the right). The pulling velocity is 0.05 Å/τ.

are provided in table 1. In order to distinguish between the A:T and G:C pairs in the DNA sequence, we strengthen the amplitude of the corresponding Lennard-Jones potential by a factor of 2 or 3, respectively. The stacking potential, between the neighbouring b-beads along each DNA is described as in model I. The hard-sphere diameters of the beads remain defined as in model I. The average value of the bead mass is $m = 162$ g mol^{-1} and the distance between the beads that make effective hydrogen bonds is 5.5 ± 0.8. We have determined that the persistence length in model II at $\widetilde{T} = 0.4$ is about 50 nm.

5. Model III: the three-bead description

Model III, introduced in [40] and shown schematically in figure 3, provides a three-bead description. It differs from model II primarily as a result of a different treatment of the backbone chain. In model II, the sugar and phosphate groups of the same nucleotide are represented by one bead, whereas in model III the two groups are represented by separate beads so that the backbone chain is formed by connecting sugar bead(s) of one nucleotide to the phosphate bead of the nucleotide that follows in the sequence.

The distinction between the phosphate and sugar groups is important in this model because it facilitates the introduction of electrostatic charges on the phosphate beads. The charges are introduced to describe interactions of the DNA with ions in the solvent but they also affect the p–p distances through the resulting Coulombic repulsion. The corresponding potential is given by $V_{\mathrm{elec},ij} = \sum \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon r_{ij}} \mathrm{e}^{-r_{ij}/\kappa_{\mathrm{D}}}$, where κ_{D} is the Debye constant, and its value depends on the ionic strength of the solution. For standard ionic strengths, κ_{D} ranges from 11 to 15 Å (e.g. when $[\mathrm{Na}^{2+}] = 50$ mM one obtains $\kappa_{\mathrm{D}} = 13.6$ Å). Here, we do not take this term into account, as its effect on the p–p distances is minor and because our focus is on mechanical manipulations and not on the effects resulting from variations of the ionic strength.

Other potentials used in this model are analogous to those used in models I and II. The exception is the base pairing potential. We describe it by the effective Lennard-Jones potential whereas Knotts *et al* [40] used the 10–12 potential:

$$V^{\mathrm{bp}} = \sum_{\text{base pairs}} 4\varepsilon_{\mathrm{bp}}\left[5\left(\frac{\sigma_{\mathrm{bp}}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{\mathrm{bp}}}{r_{ij}}\right)^{10}\right], \qquad (8)$$

where $\varepsilon_{\mathrm{bp}}$ depends on the type of base pair (AT or GC, while $\varepsilon_{\mathrm{GC}} = 3/2\varepsilon_{\mathrm{GC}}$) and σ_{bp} is around 2.9 Å for all paired bases.

Hyeon and Thirumalai [45] have recently considered a model of the RNA hairpin in which every nucleotide is represented by three beads which correspond to the phosphate, sugar and base groups which is analogous to model III of the double helix and to the model of Knotts *et al* [40]. Similar to [40], the Debye–Hückel potential between the phosphate beads is introduced to account for screening by condensed counter ions and for the hydration effects. However, there

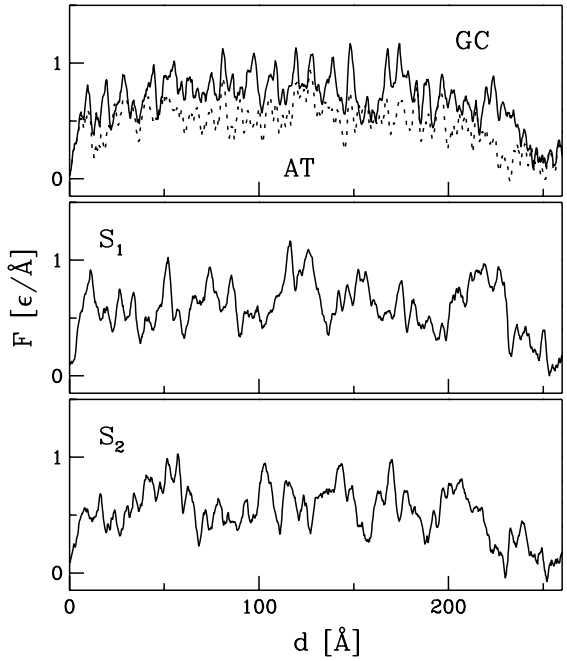

Figure 7. The A-type stretching in model II for four sequences at $\widetilde{T} = 0.2$ and $v_p = 0.05$ Å$/\tau$. The top panel corresponds to unzipping of the poly-(dG-dC) and poly-(dA-dT) strands. The remaining panels correspond to sequences S_1 and S_2 as indicated.

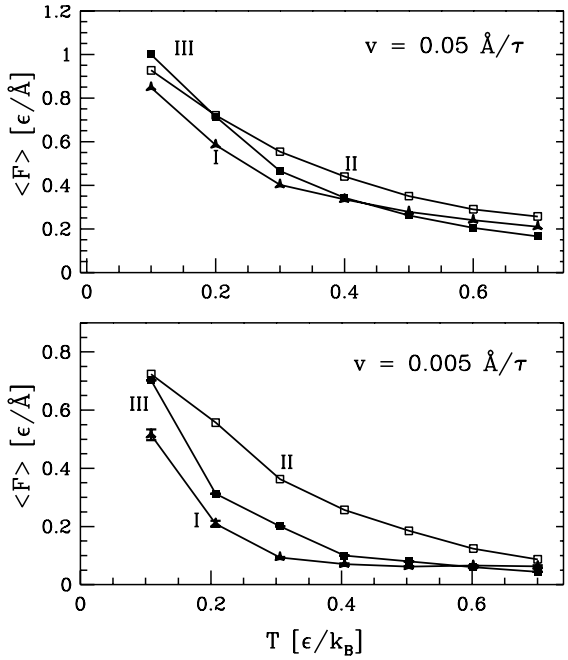

Figure 8. The average force arising during the A-type stretching at $v_p = 0.05$ Å$/\tau$ and at $v_p = 0.005$ Å$/\tau$ as a function of $\widetilde{T}$ for the three models.

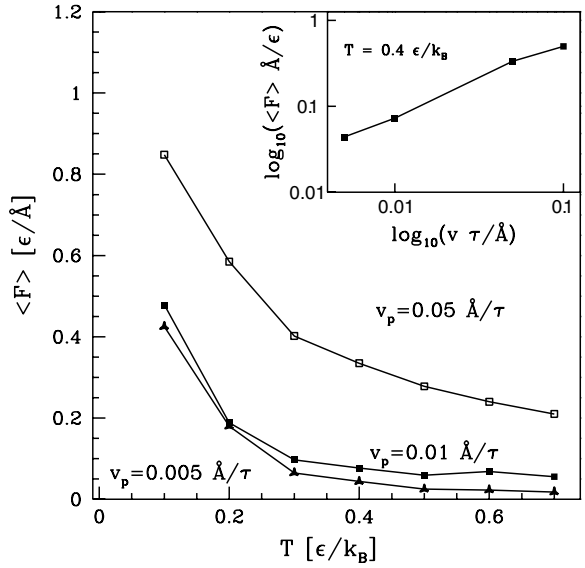

Figure 9. The average force arising during the A-type stretching for three different unravelling velocities in model I. The inset shows the log–log plot of $\langle F \rangle$ versus v_p.

are differences pertaining to the nonbonded potentials. In addition to the base pairing potential, Hyeon and Thirumalai introduce a possibility of stacking interactions between the base beads. Such interactions do not arise between the successive nucleotides, but may arise in, for example, the head of an RNA hairpin. These stacking interactions are responsible for the existence of the more complicated conformations, like the hairpin, that the RNA may adopt. As to the base pairing potential, Hyeon and Thirumalai take the Lennard-Jones potential without making a distinction between the number of hydrogen bonds involved. It affects the base beads which are within a distance of 7 Å and the corresponding depth of the potential is 1.8 kcal mol^{-1}. Non-native base bead interactions are repulsive and correspond to the energy parameter of 1.0 kcal mol^{-1}.

Another simple model of DNA has been recently proposed by Ouldridge *et al* [46] in the context of the self-assembly of DNA nanostructures in which the twisting character of the individual strands is disregarded. In this model, every nucleotide is represented by a softly repulsive sphere of diameter $l = 6.3$ Å and by another smaller sphere attached (nearly) rigidly to the centre of the repulsive sphere at a distance of $0.3l$ away from it and perpendicularly to the backbone. The smaller sphere provides a centre of attraction to another small sphere and thus plays the role of a base in the DNA strand. Four types of base site are considered and the Lennard-Jones attraction links the complementary bases. Additionally, the model incorporates a monomer-to-monomer bending energy to provide stiffness. In the ground state, two strands run parallel to each other like in a β-sheet in proteins.

Throughout this paper, we use the open square, solid square and triangle symbols to denote results corresponding to models I, II and III, respectively.

A convenient way to characterize the unravelling process is by providing the distance at which a given contact breaks down for the last time. A contact is said to be broken if the corresponding distance exceeds 1.5 times the length parameter σ in the Lennard-Jones potential associated with the contact. For models II and III, where there is only one connection between paired bases (through the b-beads) the contacts are

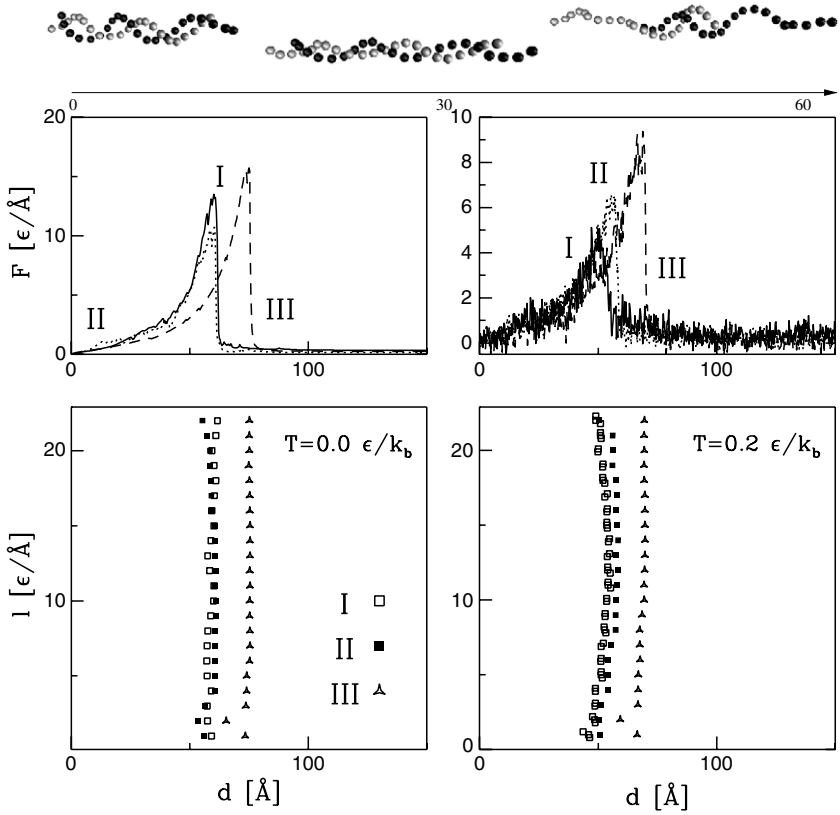

Figure 10. The B-type stretching for $v_p = 0.05$ Å$/\tau$ at two different temperatures for the three models. The panels on the left represent the results corresponding to $\widetilde{T} = 0$, and the results shown in the right panels correspond to $\widetilde{T} = 0.2\,\varepsilon/k_B$. The snapshots presented at the top show conformations during the stretching process of model II at $\widetilde{T} = 0$.

labelled by the nucleotide number, l, as counted from the side that is being pulled. In the case of model I, where each nucleotide can participate in either two or three contacts, the graphical representation uses the ordinates of $l-0.1$ and $l+0.1$ for residues A and T whereas it involves $l-0.2$, l and $l+0.2$ for residues G and C.

6. Unzipping at constant speed in scheme A

The stretching scheme A shown in figure 1 leads to the unzipping process in which the hydrogen bonds break by starting from the end that is being pulled. These bonds are enumerated by the index l which is being counted from the pulling end. Figure 5 shows the force versus displacement patterns at $\widetilde{T} = 0$ for two different values of the pulling velocity, v_p, of 0.05 and 0.005 Å$/\tau$ and for the three models discussed. At temperatures that are lower than $0.12\ \varepsilon/k_B$, the F–d patterns are qualitatively similar to the ones shown in figure 5 in the sense that the individual force peaks can usually be related to unravelling of specific base pairs in the DNA sequence.

In the initial stages of unravelling, all bonds that exist in the system get adjusted to some extent and it is only later on that the unravelling process becomes more site-specific. This is well seen in model I for $d < 50$ Å, especially for the higher pulling speed, where all force peaks are quite similar. In the other two models, this transient distance is much shorter because the number of adjustable bonds is smaller. However, after this transient stage is passed, one can read off the pair sequence of the double helix from the F–d patterns easily because the higher peaks arise due to breaking of the G:C bonds. The smaller the pulling speed the crisper the recognition of the sequence.

At $\widetilde{T} = 0.4$, we observe the peaks with the maximal value of $0.39\ \varepsilon$ Å^{-1} (which corresponds to about 40 pN)—figure 6. These values are about 20 pN higher than the experimental results [14, 15], where for $v_p = 200$ nm s^{-1} one obtains peaks of 18–20 pN

The bottom panels of figures 5 and 6 show the scenario diagrams arising in scheme A. At sufficiently low temperatures the zipping process is seen to be proceeding linearly in time with minor differences in slopes between the models. At higher temperatures, like for $\widetilde{T} = 0.4$ in model I, the scenario diagram data points acquire a curved appearance. This indicates that the thermal fluctuations rupture bonds at the idle (and not anchored) end of the dsDNA before the mechanical unzipping process gets to them.

We now discuss the issue of the sequence dependence. Figure 7 compares the F–d curves among the four sequences studied in model II and at $\widetilde{T} = 0.2$. The patterns corresponding to poly-(dG-dC) dsDNA are seen to lie higher, by about $0.2\ \varepsilon$ Å^{-1}, than for poly-(dA-dT) dsDNA which is qualitatively consistent with the experiments [16, 17]. It should also be noted that the pattern corresponding to S_2 is clearly distinct

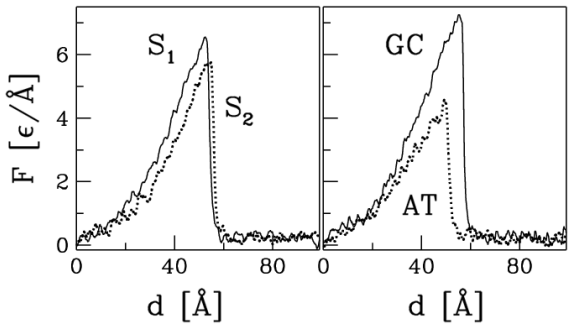

Figure 11. The B-type stretching for four sequences studied within model II. The lines are for $\widetilde{T} = 0.2$ and $v_p = 0.05$ Å$/\tau$. The left-side panel corresponds to sequences S_1 and S_2, while the right-side one to the poly-(dG-dC) and poly-(dA-dT) strands.

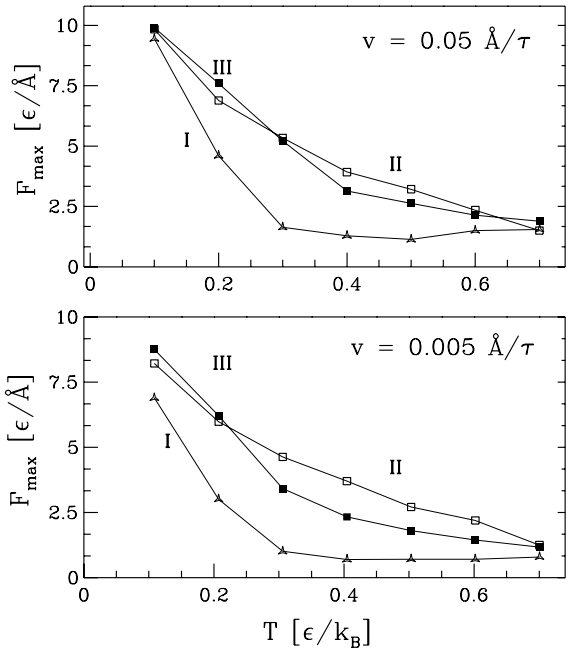

Figure 12. The maximal force for the B-type stretching for $v_p = 0.05$ Å$/\tau$ and for $v_p = 0.005$ Å$/\tau$ as a function of temperature for the three models.

from the one associated with S_2, underscoring the sensitivity to a sequence. For $\widetilde{T} = 0$ the force peaks have values of 1–2 ε Å^{-1} in all three models. On increasing the temperature, the peaks and $\langle F \rangle$, i.e. the force averaged over the duration of the full unravelling process (till F drops to 0), get lowered in a monotonic fashion, as shown in figure 8, which presents the results for unravelling with velocities $v_p = 0.05$ Å$/\tau$ and $v_p = 0.005$ Å$/\tau$. At the higher temperatures, model II yields the biggest mean forces, independent of the pulling speeds. Models I and III result in comparable mean forces at these temperatures and it depends on the velocity which one is the stronger of the two.

On lowering the unravelling velocity, the average forces decrease. Figure 9 presents the average forces for model I for three unravelling velocities. There is a big decrease of the $\langle F \rangle$ between the $v_p = 0.05$ Å$/\tau$ and $v_p = 0.01$ Å$/\tau$, while the

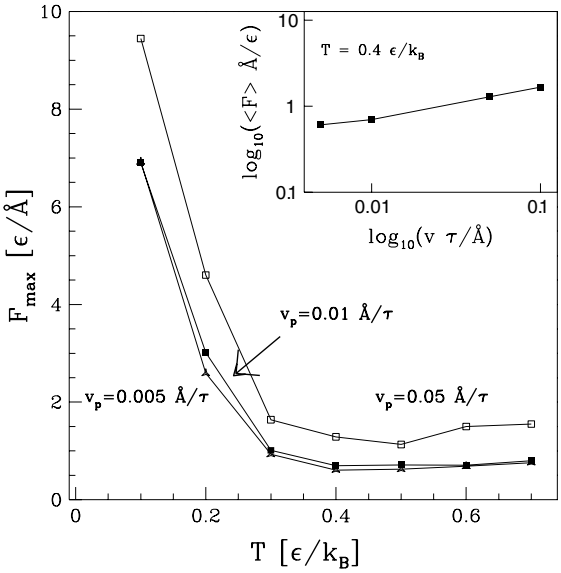

Figure 13. The B-type stretching in model I. The main figure shows F_{max} as a function of $\widetilde{T}$ for three values of v_p. The inset shows the log–log plot of F_{max} versus v_p.

results obtained for $v_p = 0.01$ Å$/\tau$ and $v_p = 0.005$ Å$/\tau$ are close to each other. Generally, the smaller the velocity, the smaller the $\langle F \rangle$ values. The similar dependences are observed for models II and III. Around $\widetilde{T} = 0.3$ the rapid decrease in $\langle F \rangle$ switches to a nearly constant behaviour at higher temperatures.

7. Stretching at constant speed in scheme B

In the B-type stretching, one chain is made to slide along its companion until the two chains separate as shown in figure 10. The F–d curves display a major peak due to shear which is an order of magnitude larger than the force peaks observed in scheme A in all of the three models studied. The emergence of this major peak is due to an increasingly cooperative resistance to manipulation of many contacts that are sheared simultaneously. Once the rupture takes place, the force drops down to the level corresponding merely to the thermal noise. The cooperation level appears to be the greatest in model III, followed by model II. In each of the models, the rupture of all contacts is nearly simultaneous.

The shear-generated force maximum depends also on the sequence as illustrated in figure 11 for model II at $\widetilde{T} = 0.2$. The maximal force obtained for S_2 is about 0.5 ε Å^{-1} lower than for S_1. The difference arises due to two circumstances: S_2 is shorter than S_1 by one base pair and, in addition, it has nine G:C pairs and not ten as in S_1. Furthermore, the maximum force corresponding to the poly-(dG-dC) sequence is substantially larger than for poly-(dA-dT). One needs the force of 4.3 ε Å^{-1} to separate poly-(dA-dT) by shearing and 6.7 ε Å^{-1} to separate the poly-(dG-dC) dsDNA molecule.

The maximal force peak dependence on temperature shown in figure 12 for the two velocities indicated is similar to what was observed for $\langle F \rangle$ in scheme A, especially at low temperatures. Models II and III are found to yield comparable

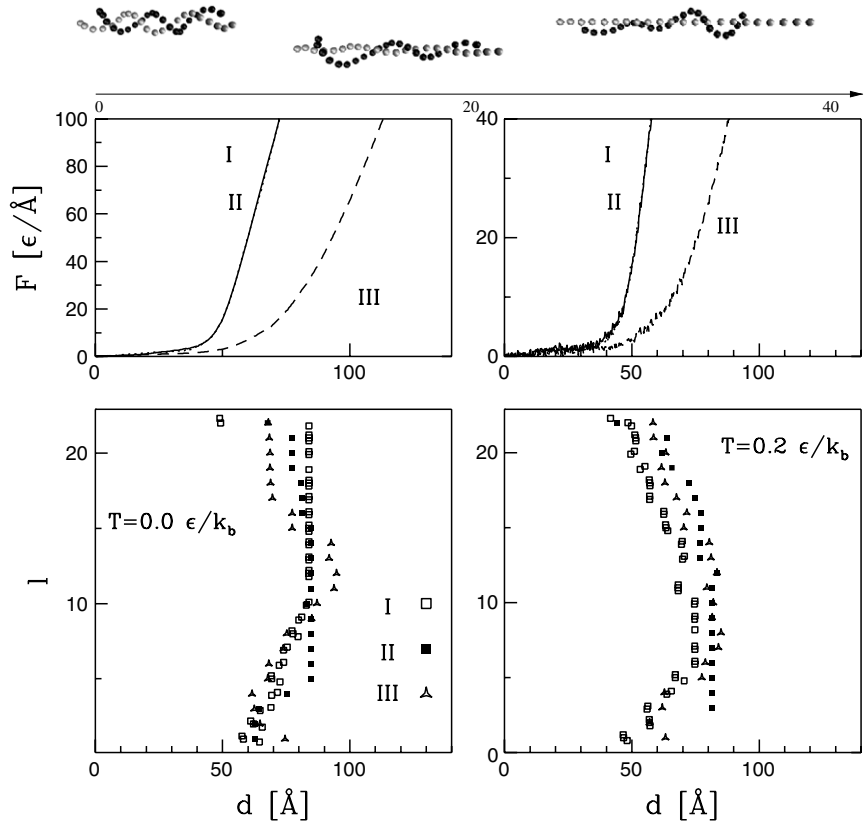

Figure 14. The C-type stretching for $v_p = 0.05$ Å$/\tau$ at two different temperatures for the three models. The panels on the left represent the results corresponding to $\widetilde{T} = 0$ and the results shown on the right panels correspond to $\widetilde{T} = 0.2\varepsilon/k_B$. The snapshots presented at the top show conformations during the stretching process of model II at $\widetilde{T} = 0.0$.

forces which are also noticeably larger than in model I except at the low temperature end. For the B-type stretching, the dependence on the stretching velocity is weak as demonstrated in figure 13. The inset demonstrates that the dependence is nearly logarithmic.

8. Stretching at constant speed in scheme C

The C-type stretching has an entirely different nature than the manipulations in schemes A and B. In this scheme, only one chain undergoes active stretching and this effect in turn influences the companion chain through the hydrogen bond contacts. Figure 14 illustrates the mechanics of this kind of manipulation. The snapshots (as obtained within model II) show that a full extension of one chain results in a substantial distortion of the other chain. Furthermore, F depends on d in a monotonic fashion. There are no force peaks even at $\widetilde{T} = 0$. The F–d curves for models I and II coincide and a bigger force arises faster than in model III because of a more direct transmission of tension between the p-beads.

The scenario diagrams also look distinct compared to scheme A and in a way which is more sensitive to the temperature. At $\widetilde{T} = 0$ many contacts are broken nearly simultaneously. At finite temperatures, the contacts at the extremities get ruptured before unravelling of the contacts in the middle in each of the models studied. The higher the $\widetilde{T}$, the earlier the particular contacts break down. We have observed insignificant dependence of the rupture distances on the pulling velocity. The whole process results in unravelling both of the hydrogen bond and of the stacking contacts.

9. Stretching at constant force and constant torque in scheme D

We now consider the tensile and torsional manipulations of the dsDNA and focus on the determination of the corresponding force–torque phase diagram.

We introduce the torsional stress of the dsDNA molecule in the following way. At one end of the molecule we choose two vectors defining the plane. The first vector is defined by the positions of the extreme p-beads at the chosen end of the dsDNA. The second vector defining the plane is a cross-product of the first vector and the dsDNA axis (which is defined by the midpoints of the extreme beads on both ends of the DNA molecule). In the plane constructed in the way described above we add two more beads as shown in figure 15 so that a square frame of four beads is formed. All beads in this frame are connected by the springs as in other structural bonds. The extreme beads on the other end of the molecule are anchored at their starting positions.

The torsion is applied to the DNA molecule by application of a force to each of the four beads in the square frame. The

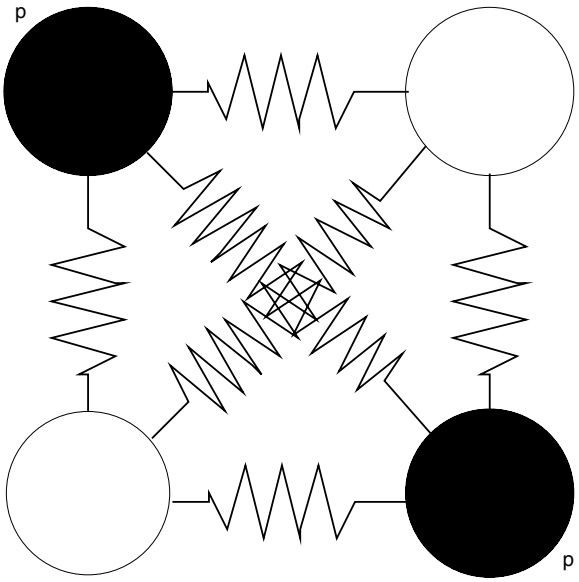

Figure 15. A schematic view of the beads mounted at the DNA end in order to apply torque through them. All beads are connected with the springs.

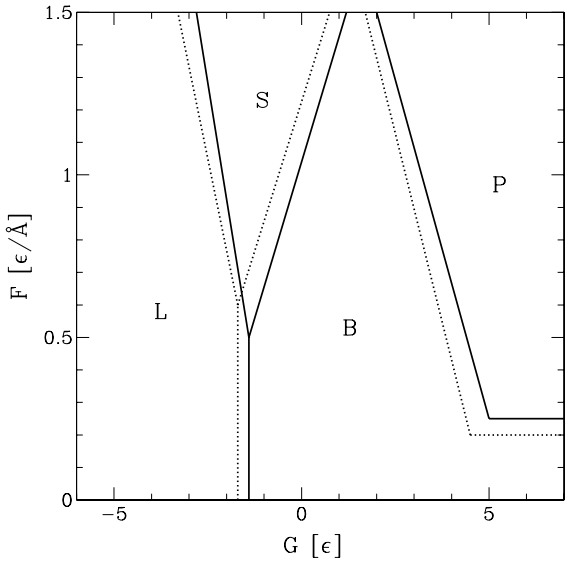

Figure 16. The phase diagram representing the final type of dsDNA structure obtained after applying the stretching force, F, and the torque, G, in model II. L denotes the L-DNA form and S signifies stretched chains, in which the backbones form straight lines. B denotes the original B-type structure. P corresponds to the Pauling form of the DNA, where the backbones get closer to each other and bases stay outside of the helix. The solid lines represent results obtained for $\widetilde{T} = 0.4$, while the dashed ones for $\widetilde{T} = 0.2$.

torque is perpendicular to the frame. The torsion is considered positive if it agrees with the sense of the twist in the double helix in the B-form and it is considered negative otherwise. The stretching force F is a resultant force applied to all of the four beads in the frame and along the dsDNA axis.

We first consider model II at two temperatures: 0.2 and 0.4 $\varepsilon/k_{\mathrm{B}}$. The results are presented in the phase diagram in figure 16. The boundaries of the phases are approximate and are quite similar in both models. These phase diagrams are also similar in appearance to those established experimentally [24, 22, 27] and theoretically [47, 24].

We start from the dsDNA B-form structure and observe the transitions into other phases of the DNA structure. At $\widetilde{T} = 0.4$, the DNA transforms into the L-form under the torque G of around $-1.5\ \varepsilon$. At the lower temperature, this transformation occurs for G of around 1.8–2.0 ε. The simulations which lead to the L-form with the value of the torque being close to this limiting value may last for up to $15\,000\tau$. This time becomes significantly shorter for larger values of G.

The experimentally characterized overstretched DNA molecule is extended by about 60% compared to the B-DNA form. We denote these structures by S-DNA as our effective hydrogen bonds are found not to be broken. In models I and II the elongation of the system leads to tightening of the bonds between p-beads along the chains, which finally leads to a significant increase in the applied force. Thus the S-form region in the dsDNA phase diagram corresponds to structures in which the backbone forms a straight line, without imposing the condition of 60% overstretching. In $\widetilde{T} = 0.4$ such structures occur while the force of 0.5 ε Å^{-1} is applied. For lower temperature there is needed a bit larger force for about 0.05–0.1 ε Å^{-1}. The above values were obtained for applied torque G of 1.5 ε and 1.7 ε, respectively, for temperatures of 0.4 ε and 0.2 ε.

The Pauling form is obtained when both the stretching force and the positive twisting torque are large. The smallest value of force needed to transform the system into the DNA P-form is 0.25 ε Å^{-1}, while the torque applied must be of a value around 5.0 ε. For larger stretching forces, G decreases to 2.1 ε at $F = 1.5\ \varepsilon$ Å^{-1}. In the P-form form, the p-beads come closer together while the remaining beads (b and h) become exposed and face out of the helix.

10. Stretching at constant angular speed in scheme D

In order to study stretching at a constant angular speed, we anchor the bottom beads and attach two frames to the top. Each of these frames is as described in the previous section and they coincide initially. The beads in one frame are connected to their twins by elastic springs. As the outer frame rotates at a constant angular speed, these springs get stretched and impose a twist on the inner frame which is glued to the DNA. This construct facilitates determination of the resistive torque as it is accomplished by monitoring stretching of the interframe springs.

Figure 17 shows the torque of resistance to twisting as a function of the angle of rotation of the outer frame. Two magnitudes of the angular speed, ω, are used, $0.000\,14\frac{1}{\tau}$ and $0.000\,69\frac{1}{\tau}$, which differ by the factor of 5. We also probe two senses of the twisting: agreeing with the helical twist ($\omega > 0$) or opposing it ($\omega < 0$). The former leads to overtwisting and an indefinite growth in the resistive torque due to an increasing infringement of the steric constraints. The latter results in unwinding and in a transition from the B-form to the L-form.

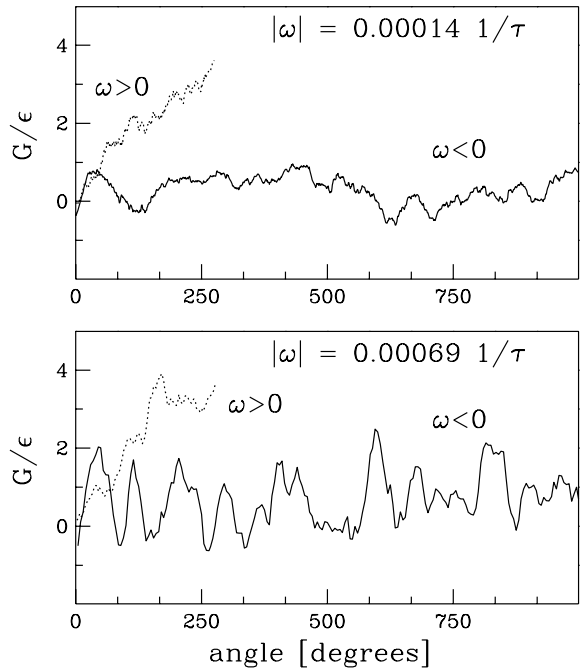

Figure 17. Torque of resistance to the twist as a function of the angle of rotation for two magnitudes of the angular velocities in $\widetilde{T} = 0.4$ as indicated. The system consists of 20 base pairs and the simulations have been performed within model II. The dashed lines correspond to a sense of the twist that agrees with helical rotation of the dsDNA. The solid lines correspond to the opposite sense of the twist.

The results clearly depend on the twisting speed. In particular, the average torques are 0.319 ε and 0.890 ε for the smaller and faster negative angular speeds, respectively. The peaks in the torque result from the distortions, but no contacts get ruptured when one infers about them from the distance-based criterion.

11. Conclusions

The coarse-grained models of the dsDNA discussed here allow for studies of features at the level of a single nucleotide. These models are found to be fairly equivalent and indicate that sequence-specific events can be observed in mechanical manipulations of various kinds when performed at low and moderate temperatures. However, this capability may become borderline around room temperature.

The models proposed here should be useful when studying DNA–protein complexes and when assisting nanotechnological DNA assembly processes theoretically as they involve substantially larger size scales than a single DNA. Examples of such processes are described in [48–50]. These models could also serve as systems that lead to construction of models of RNA and then to the viral RNAs confined by proteinic capsids. They should also find usage in molecular-level dynamical studies of topologically non-trivial conformations in DNA, such as knots [51–53] in analogy to [54, 55].

Acknowledgments

Many fruitful discussions with P Szymczak are appreciated. This work has been supported by grant N N202 0852 33 from the Ministry of Science and Higher Education in Poland and by the European Union within the European Regional Development Fund, through grant Innovative Economy (POIG.01.01.02-00-008/08).

References

[1] Bustamante C, Bryant Z and Smith S B 2003 *Nature* **421** 423–6
[2] Koster D A, Croquette V, Dekker C, Shuman S and Dekker N H 2005 *Nature* **434** 671–4
[3] Marko J F 1997 *Proc. Natl Acad. Sci. USA* **94** 11770–2
[4] Sułkowska J I and Cieplak M 2007 *J. Phys.: Condens. Matter* **19** 283201
[5] Rief M, Gautel M, Oesterhelt F, Fernandez J M and Gaub H E 1997 *Science* **276** 1109–12
[6] Fowler S B, Best R B, Toca herrera J L, Rutherford T J, Steward A, Paci E, Karplus M and Clarke J 2002 *J. Mol. Biol.* **322** 841–9
[7] Yang G, Cecconi C, Baase W A, Vetter I R, Breyer W A, Haack J A, Matthews B W, Dahlquist F W and Bustamante C 2000 *Proc. Natl Acad. Sci. USA* **97** 139–44
[8] Carrion-Vazquez M, Oberhauser A F, Fisher T E, Marszalek P E, Li H and Fernandez J M 2000 *Prog. Biophys. Mol. Biol.* **74** 63–91
[9] Lu H and Schulten K 1999 *Chem. Phys.* **247** 141–53
[10] Paci E and Karplus M 2000 *Proc. Natl Acad. Sci. USA* **97** 6521–6
[11] Pabon G and Amzel L M 2006 *Biophys. J.* **91** 467–72
[12] Marko F and Siggia E D 1994 *Macromolecules* **27** 981
Marko F and Siggia E D 1995 *Macromolecules* **28** 8759
[13] Jendrejack R M, de Pablo J J and Graham M D 2002 *J. Chem. Phys.* **116** 7752–9
[14] Bockelmann U, Essevaz-Roulet B and Heslot F 1997 *Phys. Rev. Lett.* **79** 4489–92
[15] Bockelmann U, Thomen Ph, Essevaz-Roulet B, Viasnoff V and Heslot F 2002 *Biophys. J.* **82** 1537–53
[16] Essevaz-Roulet B, Bockelmann U and Heslot F 1997 *Proc. Natl Acad. Sci. USA* **94** 11935–40
[17] Rief M, Clausen-Schaumann H and Gaub H E 1999 *Nat. Struct. Biol.* **6** 346–9
[18] Baldazzi V, Cocco S, Marinari E and Monasson R 2006 *Phys. Rev. Lett.* **96** 128102
[19] Dudko O K, Hummer G and Szabo A 2008 *Proc. Natl Acad. Sci. USA* **105** 15755–60
[20] Cluzel P, Lebrun A, Heller C, Lavery R, Viovy J-L, Chatenay D and Caron F 1996 *Science* **271** 792–4
[21] Luan B and Aksimentiev A 2008 *Phys. Rev. Lett.* **101** 118101
[22] Oroszi L, Galajda P, Kirei H, Bottka S and Ormos P 2006 *Phys. Rev. Lett.* **97** 058301
[23] Wereszczynski J and Andricioaei I 2006 *Proc. Natl Acad. Sci. USA* **103** 16200–5
[24] Bryant Z, Stone M D, Gore J, Smith S B, Cozzarelli N R and Bustamante C 2003 *Nature* **424** 338–41
[25] Smith S B, Cui Y and Bustamante C 1996 *Nature* **271** 795–9
[26] Clausen-Schaumann H, Rief M, Tolksdorf C and Gaub H 2000 *Biophys. J.* **78** 1997–2007
[27] Allemand J F, Bensimon D, Lavery R and Croquette V 1998 *Proc. Natl Acad. Sci. USA* **95** 14152–7
[28] Shokri L, McCauley M J, Rouzina I and Williams M C 2008 *Biophys. J.* **95** 1248–55
[29] Marenduzzo D, Maritan A, Orlandini E, Seno F and Trovato A 2009 *J. Stat. Mech.* L0400
[30] Albert B, Bray D, Lewis J, Raff M, Roberts K and Watson J D 1994 *Molecular Biology of the Cell* 3rd edn (New York: Garland Publishing)
[31] Leppard J B and Champoux J J 2005 *Chromosoma* **114** 75–85
[32] Corbett K D and Berger J M 2004 *Annu. Rev. Biophys. Biomol. Struct.* **33** 95–118

[33] Cieplak M and Hoang T X 2003 *Biophys. J.* **84** 475–88
[34] Cieplak M, Hoang T X and Robbins M O 2002 *Proteins* **49** 114–24
[35] Cieplak M, Hoang T X and Robbins M O 2004 *Proteins* **56** 285–97
[36] Kwiecińska J I and Cieplak M 2005 *J. Phys.: Condens. Matter* **17** S1565–80
[37] Szklarczyk O, Staroń K and Cieplak M 2009 *Proteins* at press
[38] Sułkowska J I and Cieplak M 2008 *Biophys. J.* **95** 3174–91
[39] Trovato F and Tozzini V 2008 *J. Phys. Chem.* B **112** 13197–200
[40] Knotts T A IV, Rathore N, Schwartz D C and de Pablo J J 2007 *J. Chem. Phys.* **126** 84901
[41] Leonard G A and Hunter W N 1993 *J. Mol. Biol.* **234** 198–208
[42] Masliah G, Rene B, Zargarian L, Fermandjian S and Mauffret O 2008 *J. Mol. Biol.* **381** 692–706
[43] Clementi C, Nymeyer H and Onuchic J N 2000 *J. Mol. Biol.* **298** 937
[44] Allen M P and Tildesley D J 1987 *Computer Simulation of Liquids* (New York: Oxford University Press)
[45] Hyeon C and Thirumalai D 2005 *Proc. Natl Acad. Sci. USA* **102** 6789–94
[46] Ouldridge T E, Johnston I G, Louis A A and Doye J P K 2009 *J. Chem. Phys.* **130** 065101
[47] Sarkar A, Léger J F, Chatenay D and Marko J F 2001 *Phys. Rev.* E **63** 051903
[48] Seeman N C 2003 *Nature* **421** 427–30
[49] Mitchell J C, Harris J R, Malo J, Bath J and Turberfield A J 2004 *J. Am. Chem. Soc.* **126** 16342–3
[50] Goodman R P, Schaap I A T, Tardin C F, Erben C M, Berry R M, Schmidt C F and Turberfield A J 2005 *Science* **310** 1661–5
[51] Krasnow M A, Stasiak A, Spengler S J, Dean F, Kollert T and Cozzarelli N R 1983 *Nature* **304** 559–60
[52] Weber C, Stasiak A, De Los Rios P and Dietler G 2006 *Biophys. J.* **90** 3100–5
[53] Weber C, De Los Rios P, Dietler G and Stasiak A 2006 *J. Phys.: Condens. Matter* **18** S161–71
[54] Sułkowska J I, Sułkowski P, Szymczak P and Cieplak M 2008 *Phys. Rev. Lett.* **100** 058106
[55] Sułkowska J I, Sułkowski P, Szymczak P and Cieplak M 2008 *Proc. Natl Acad. Sci. USA* **105** 19714–9