Article No. jmbi.1999.2798 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (1999) **289**, 1301–1326

JMB

DNA Stretching and Compression: Large-scale Simulations of Double Helical Structures

Konstantin M. Kosikov[1], Andrey A. Gorin[2], Victor B. Zhurkin[3]* and Wilma K. Olson[1]*

[1]*Department of Chemistry Rutgers, the State University of New Jersey, Wright-Rieman Laboratories, 610 Taylor Road Piscataway NJ 08854-8087, USA*

[2]*Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York NY 10021, USA*

[3]*Laboratory of Experimental and Computational Biology National Cancer Institute, National Institutes of Health Bethesda, MD 20892, USA*

Computer-simulated elongation and compression of *A*- and *B*-DNA structures beyond the range of thermal fluctuations provide new insights into high energy ''activated'' forms of DNA implicated in biochemical processes, such as recombination and transcription. All-atom potential energy studies of regular poly(dG)·poly(dC) and poly(dA)·poly(dT) double helices, stretched from compressed states of 2.0 Å per base-pair step to highly extended forms of 7.0 Å per residue, uncover four different hyperfamilies of right-handed structures that differ in mutual base-pair orientation and sugar-phosphate backbone conformation. The optimized structures embrace all currently known right-handed forms of double-helical DNA identified in single crystals as well as non-canonical forms, such as the original ''Watson-Crick'' duplex with *trans* conformations about the P—O5′ and C5′—C4′ backbone bonds. The lowest energy minima correspond to canonical *A* and *B*-form duplexes.

The calculations further reveal a number of unusual helical conformations that are energetically disfavored under equilibrium conditions but become favored when DNA is highly stretched or compressed. The variation of potential energy *versus* stretching provides a detailed picture of dramatic conformational changes that accompany the transitions between various families of double-helical forms. In particular, the interchanges between extended canonical and non-canonical states are reminiscent of the cooperative transitions identified by direct stretching experiments. The large-scale, concerted changes in base-pair inclination, brought about by changes in backbone and glycosyl torsion angles, could easily give rise to the observed sharp increase in force required to stretch single DNA molecules more than 1.6-1.65 times their canonical extension.

Our extended duplexes also help to tie together a number of previously known structural features of the RecA-DNA complex and offer a self-consistent stereochemical model for the single-stranded/duplex DNA recognition brought in register by recombination proteins. The compression of model duplexes, by contrast, yields non-canonical structures resembling the deformed steps in crystal complexes of DNA with the TATA-box binding protein (TBP). The crystalline TBP-bound DNA steps follow the calculated compression-elongation pattern of an unusual ''vertical'' duplex with base planes highly inclined with respect to the helical axis, exposed into the minor groove, and accordingly accessible for recognition.

Significantly, the double helix can be stretched by a factor of two and compressed roughly in half before its computed internal energy rises sharply. The energy profiles show that DNA extension-compression is related not only to the variation of base-pair Rise but also to concerted changes of Twist, Roll, and Slide. We suggest that the high energy ''activated'' forms calculated here are critical for DNA processing, e.g. nucleoprotein recognition, DNA/RNA synthesis, and strand exchange.

Abbreviations used: TBP, TATA box binding protein; MD, molecular dynamics; EM, electron microscopic; CRP, cyclic AMP receptor protein.

E-mail addresses of the corresponding authors: olson@rutchem.rutgers.edu; zhurkin@structure.nci.nih.gov.

0022-2836/99/251301–26 $30.00/0

Keywords: DNA stretching; conformational transitions; molecular simulation; RecA-DNA assembly; TBP-DNA complex

*Corresponding author

Introduction

A remarkable series of new physical manipulations of individual DNA molecules (Bensimon, 1996) has renewed interest in the large-scale conformational flexibility of the double helix. In addition to the long known solvent-induced transition of the slim, elongated *B*-type helix (Watson & Crick, 1953) into the wide, stubby *A*-form (Franklin & Gosling, 1953), *B*-DNA can now be converted to a novel over-stretched S conformation upon direct pulling (Bensimon, 1996; Cluzel *et al*., 1996; Smith *et al*., 1996; Baumann *et al*., 1997). This extended state, while anticipated nearly 50 years ago from the optical properties of fibers under tension (Wilkins *et al*., 1951), is only beginning to be understood. According to the best available molecular models (Cluzel *et al*., 1996; Konrad & Bolonick, 1996; Lebrun & Lavery, 1996), complementary base-pairing remains intact in the extended duplex (which is 1.6-1.65 times the normal 3.4 Å/bp contour length), the duplex is unwound, and base-pairs are inclined significantly with respect to the helical axis.

These artificially induced changes in conformation provide new clues to the response of DNA to proteins and other ligands. While many DNA-binding proteins take advantage of the molecular mechanisms involved in the $B \leftrightarrow A$ helical transformation: bending, unwinding, and displacing neighboring base-pairs (Calladine & Drew, 1984) through concerted changes in backbone and glycosyl torsion angles (Olson, 1976; Yathindra & Sundaralingam, 1976; Zhurkin *et al*., 1978), a growing number of proteins distort the double helix to conformational states outside the bounds of the natural $B \leftrightarrow A$ transition (see reviews by Guzikevich-Guerstein & Shakked, 1995; Suzuki & Yagi, 1995; Werner *et al*., 1996; Dickerson, 1998; Olson *et al*., 1998). The stretching of DNA by protein arises in many instances from the intercalation, or partial insertion, of amino acid side groups between adjacent base-pairs, although there is at least one example of non-intercalative DNA base separation (~5 Å) in a single-stranded DNA fragment bound to the RecA protein (Nishinaka *et al*., 1997, 1998). The latter structure, which is stabilized by novel sugar-base stacking, is thought to be relevant to the regularly elongated form of DNA in the RecA filament (Howard-Flanders *et al*., 1987) as well as to physically stretched forms of DNA. Molecular models of related structures suggest that sugar-sugar interactions may also be involved (Zhurkin *et al*., 1994a,b).

The precise mechanisms of DNA stretching are still unknown at the atomic level. Recent energy minimization and molecular dynamics (MD) simulations of extended duplex structures (Konrad & Bolonick, 1996; Lebrun & Lavery, 1996) have focused more on the global conformational features of over-stretched duplexes than the concerted movements of bases and sugar-phosphate backbone involved in the deformation processes. Moreover, there are discrepancies in the predicted conformational changes in the published studies. For example, the unwound ribbon-like structures found by pulling on the 3′-termini of complementary strands of oligonucleotide models are sometimes (Konrad & Bolonick, 1996), but not always (Lebrun & Lavery, 1996), accompanied by high base inclination. These differences presumably reflect a variety of built-in methodological distinctions, including (1) the nature of the assumed force fields, AMBER (Pearlman *et al*., 1995) *versus* JUMNA (Lavery *et al*., 1995), (2) the choice of independent structural variables (Cartesian *versus* internal coordinates), and (3) the route taken to change conformation (MD *versus* minimization). In addition, the stretched forms revealed in energy minimization depend on the method used to bring about chain extension, i.e. pulling the molecule on the 3′ and 5′ termini gives different results (Lebrun & Lavery, 1996). On the other hand, the corresponding simulations of *A* and *B*-DNA compression have not yet been performed.

Here we report a systematic all-atom potential energy analysis of the compression and elongation of regular DNA double helices. Rather than pushing or pulling the chain on the 5′ or 3′ ends with explicit energy terms, we solve a constrained minimization problem that identifies the conformations of a DNA duplex with fixed global helical extension. This method allows us to follow the longitudinal deformations of the double helix in terms of its ''macroscopic'' stretching coordinate, the helical rise. We optimize the configurations of both poly(dA)·poly(dT) and poly(dG)·poly(dC) homopolymers (of given helical extension) under high and low salt conditions. In order to sample conformation space as widely as possible, we adopt an implicit representation of the chemical environment with a distant-dependent dielectric constant and atomic charges modified to mimic counterion condensation (Young *et al*., 1997). Optimization of *A* and *B*-DNA helices reveals several structural ''hyperfamilies'' that exhibit significantly different energy profiles, backbone conformations, and base-pair step geometries. The combined data provide new insight into the deformations of duplex DNA and the nature of both highly stretched and protein-deformed double helices.

Importantly, the simulated duplex variability goes beyond the limits of the thermal fluctuations that occur under ''standard solution conditions''. Here we are interested in the large-scale deformations of DNA caused by bound proteins, such as TBP or RecA, or by direct physical manipulations. The biological role of these high energy ''activated'' states (Rich, 1959) is related to the improved accessibility of Watson-Crick base-pairs during the course of DNA ''processing'' (methylation, recombination, transcription, etc.). These forms, which are unfavorable in free DNA, are presumably stabilized by the binding of proteins. While the specific ways to facilitate DNA distortions are different for different proteins, one feature always remains the same: the conformational preferences of the duplex itself, which are the subject of this study.

Results and Discussion

Overview of hyperfamilies

The minimization of conformational energy yields four hyperfamilies of right-handed helical structures with characteristic sugar-phosphate backbones (Table 1). Each hyperfamily contains both *A* and B-like duplexes with sugars puckered respectively in C3′-*endo* and C2′-*endo* domains and phosphodiester and glycosyl torsions concomitantly perturbed in well established directions, i.e. the O3′—P phosphodiester ζ angle is 30-40° higher and the glycosyl torsion χ is ∼30-90° lower in *A versus B*-duplexes. These conformational distinctions are consistent with all available crystallographic data (Schneider *et al*., 1997) and persist over the complete range of simulated stretching and compression (2.0-7.0 Å/bp helical rise).

Despite the differences in sugar puckering, the *A* and *B*-like structures in each hyperfamily exhibit the same qualitative behavior during the complete course of stretching and compression (Figures 1-5). The similar responses of backbone and bases to longitudinal deformations support early ideas that *A* and *B*-like forms occupy parallel ravines in (torsion angle/helical parameter) conformation space and that these forms are simply shifted by differences in P···P and C1′···C1′ distances (Zhurkin *et al*., 1978). Here we find additional sets of related duplexes sharing other common structural relationships as well as similar longitudinal deformabilities.

The energetic and structural changes in the ''high salt'' regime, which is thought to mimic the natural cellular environment of the double helix, persist as well under simulated ''low salt'' conditions (see Figure 1 and discussion below). Here we highlight the important conformational distinctions among the hyperfamilies at ''high salt'' in terms of the preferred ranges of dihedral angles: *trans* (*t*), *gauche*$^{\pm}$ ($g^{\pm}$) for backbone torsions of 180(±60)° and ±60(±60)°, and *anti*, *high anti*, and *syn* for O4′—C1′—N9(N1)—C8(C6) glycosyl torsions of 30(±45)°, 120(±45)°, and 210(±45)°, respectively. Terms such as stretch, stretch per step, global stretching, and extension are used synonymously with ''helical rise'' in the text. At the same time, dimer ''Rise'' and other local parameters (designated by capital letters) are used to characterize the base-pair step geometry. The local parameters (Figure 2) describe the positioning of coordinate frames embedded in adjacent base-pairs, whereas the helical parameters (Figure 3) are defined with respect to a global duplex axis (see Methods). While we focus attention on the computed behavior of poly(dG)·poly(dC) polymers, the variation of

Table 1. Torsion angles, in degrees, for *A* and *B* families of forms

	P	α O3′—P—O5′—C5′	β P—O5′—C5′—C4′	γ O5′—C5′—C4′—C3′	ε C4′—C3′—O3′—P	ζ C3′—O3′—P—O5′	χ O4′—C1′—N9(N1)—C8(C6)
A_1	20 to 30	270 to 290	170 to 190	40 to 80	180 to 210	260 to 300	10 to 35
B_1	120 to 150	280 to 300	180 to 190	40 to 60	180 to 190	220 to 270	35 to 70
A_t	−10 to 20	**150** to **170**	170 to 190	**160** to **180**	200 to 220	280 to 300	−5 to 10
B_t	130 to 180	**160** to **180**	150 to 200	**160** to **190**	180 to 200	220 to 270	20 to 45
A_2	20 to 30	290 to 310	**110** to **130**	60 to 70	**250** to **270**	**200** to **220**	40 to 55
A_k	15 to 25	280 to 300	**110** to **130**	55 to 65	**265** to **280**	**205** to **235**	30 to 55
B_2	160 to 180	240 to 280	**150** to **160**	40 to 70	**240** to **270**	**160** to **190**	**130** to **145**
A_3	−25 to 15	250 to 280	190 to 210	50 to 70	180 to 200	250 to 280	**100** to **105**
B_3	140 to 170	270 to 280	190 to 200	40 to 60	190 to 210	220 to 260	**130** to **135**
		260 to 320		30 to 80			
NDB, *A*-DNA[a]	−20 to 60		140 to 220		160 to 250	250 to 320	−20 to 70
		180 to 190		140 to 160			
			160 to 200		160 to 200	230 to 300	20 to 70
NDB, *B*-DNA[a]	100 to 230	270 to 330		20 to 80			
			130 to 160		230 to 270	150 to 210	80 to 120

Torsion angle ranges in the A_tB_t, A_2B_2, and A_3B_3 hyperfamilies different from the corresponding values for canonical *A* and *B*-forms (A_1B_1 hyperfamily) are highlighted in boldface. Angles are defined as 0° for planar *cis* arrangements of the designated four atom sequences.

[a] Angular ranges in double helical *A* and *B*-DNA crystal structures (without mismatches and free of bound drugs or proteins) taken from Schneider *et al*. (1997).

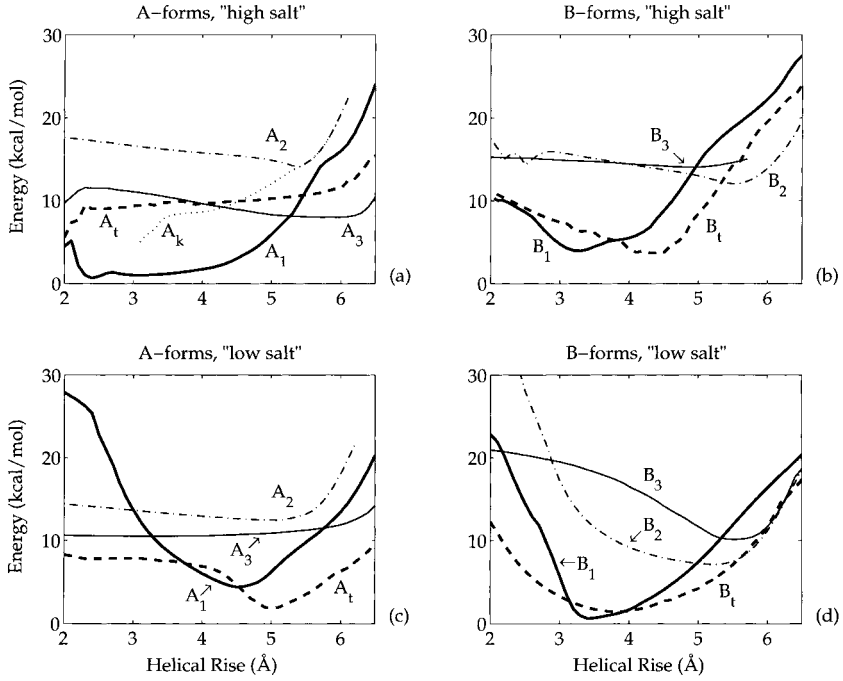

Figure 1. Stretching/compression profiles of energy *versus* helical rise of optimized *A* and *B*-form poly (dG)·poly(dC) duplexes simulated under ((a) and (b)) ''high'' and ((c) and (d)) ''low salt'' conditions. The interchanges between conformational forms occur at points where the profiles intersect. Plot styles distinguish conformational families: ———, A_1B_1; - - - - - -, A_tB_t; – · – · –, A_2B_2; ———, A_3B_3. The thin dotted curves (· · · ·) correspond to the kinked A_k helices found within the A_2 family of structures.

poly(dA)·poly(dT) is roughly identical because of the simplified nature of our homopolymer model (see Methods).

With few exceptions, the torsion angles within each conformational family are fairly constant over the course of longitudinal deformations, with individual angles showing only modest variations (ca. 20-30°) from their characteristic values upon stretching. As a result, the P···P virtual bond distances v (between successive phosphorus atoms

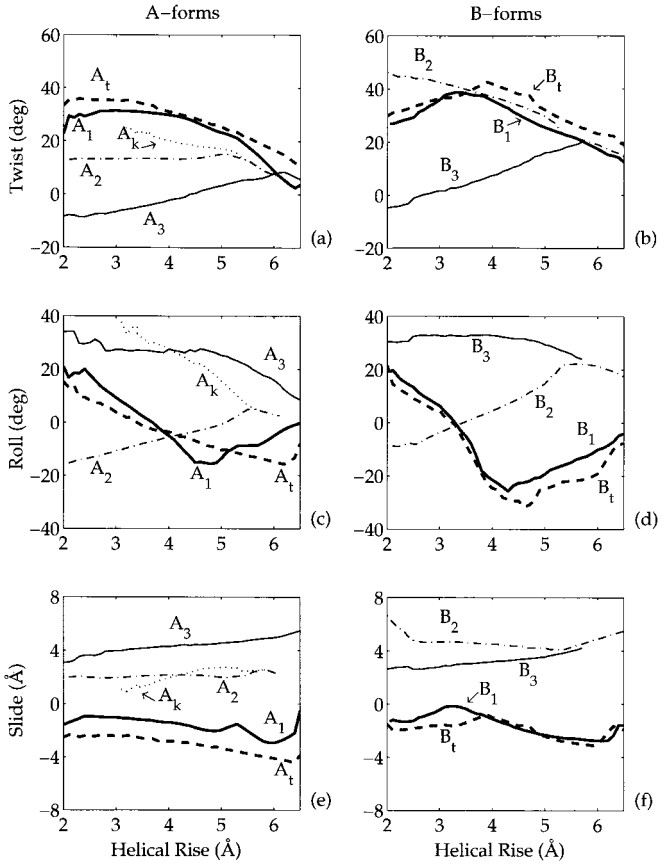

Figure 2. Local (symmetrical) base-pair step parameters: (a) and (b) Twist; (c) and (d) Roll; (e) and (f) Slide; associated with the stretching/compression profiles of *A* and *B* conformational families under the ''high salt'' conditions of Figure 1.

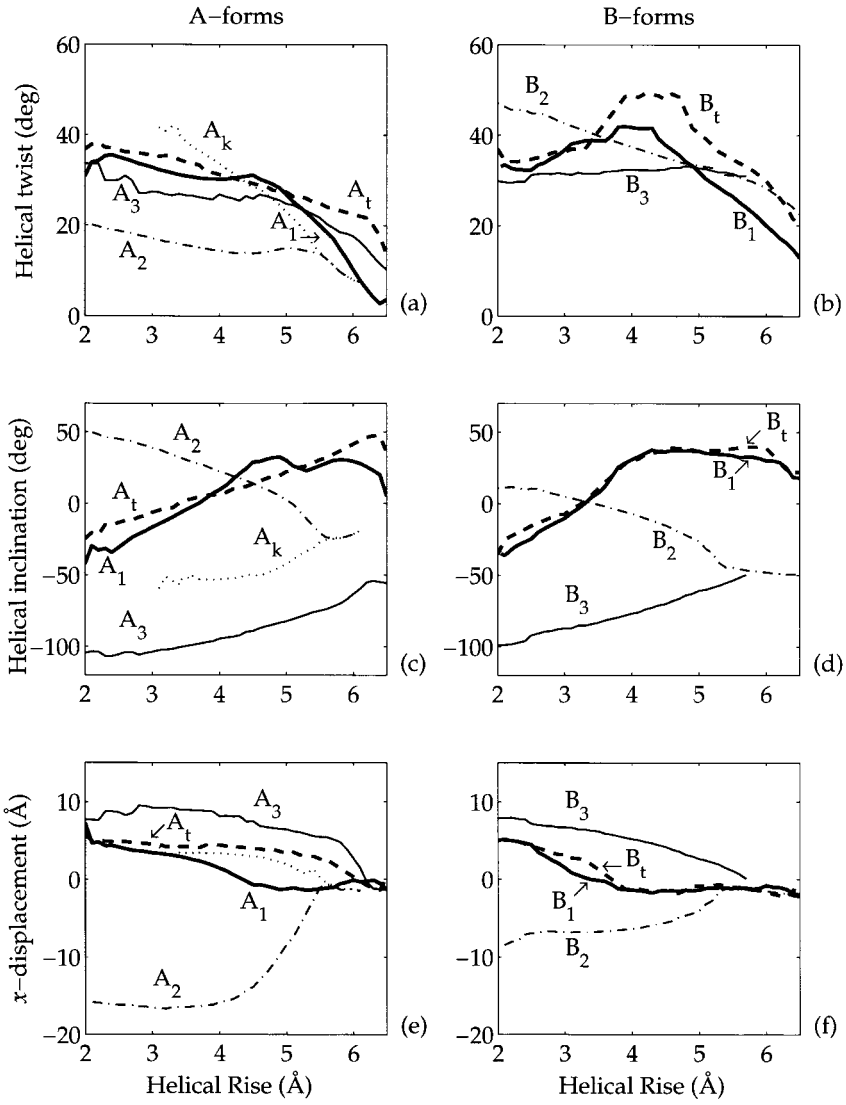

Figure 3. Global helical parameters; (a) and (b) twist; (c) and (d) inclination; (e) and (f) shift (*x*-displacement), characterizing the stretching/compression profiles of *A* and *B* conformational families under the ''high salt'' conditions of Figure 1.

on the same strand) are roughly constant (Figure 4), and the helical radii r (defined by the P atoms of the homopolymer model) follow the theoretically expected dependence on global extension (h), helical twist (Ω_h), and P$\cdots$P distance (v): $r^2 = (v^2 - h^2)/2(1 - \cos\,\Omega_h)$.

A_1B_1 hyperfamily

The lowest energy forms overall, found in what we call the A_1B_1 hyperfamily, correspond to the canonical *A* and *B*-DNA double helices that dominate the X-ray literature (heavy solid curves in Figure 1). Both forms respond to stress in much the same way. Specifically, Twist increases slightly upon stretching from 2.0 to 3.0-3.5 Å/bp, and then decreases substantially (heavy solid curves in Figure 2(a) and (b)). Roll, by contrast, decreases with stretching of the duplex up to helical rise values of 4.5 Å/bp, and subsequently increases with further extension (Figure 2(c) and (d)). These correlated conformational changes mimic the behavior of DNA in solution, stretched fibers, and single crystals. For example, the computed changes in base-pair inclination in the vicinity of the global minimum are the same order of magnitude as the 10-20° deviations from perpendicularity deduced from the linear and circular dichroism spectra of representative DNAs in aqueous salt solution (Nordén *et al.*, 1978; Edmondson & Johnson, 1986). Furthermore, our over-twisted B_1 duplex is characterized by the same negative Roll (Figure 2(d)), positive base-pair inclination (Figure 3(d)), radial narrowing (Figure 4(f)), and minor groove compression (Figure 5(b)) observed in the course of the $B \rightarrow C \rightarrow D$ transition in fibers (Arnott *et al.*, 1974; Selsing *et al.*, 1975).

The variations in local base-pair step and global helical parameters with longitudinal deformations also follow the conformational patterns exhibited by dimer steps in high resolution *A* and *B*-DNA crystal structures (Olson *et al.*, 1998, 1999). Specifically, the local Twist and Rise are positively correlated in nine out of ten dimeric steps, a finding which is in agreement with the present simulations and which serves as further indication of the

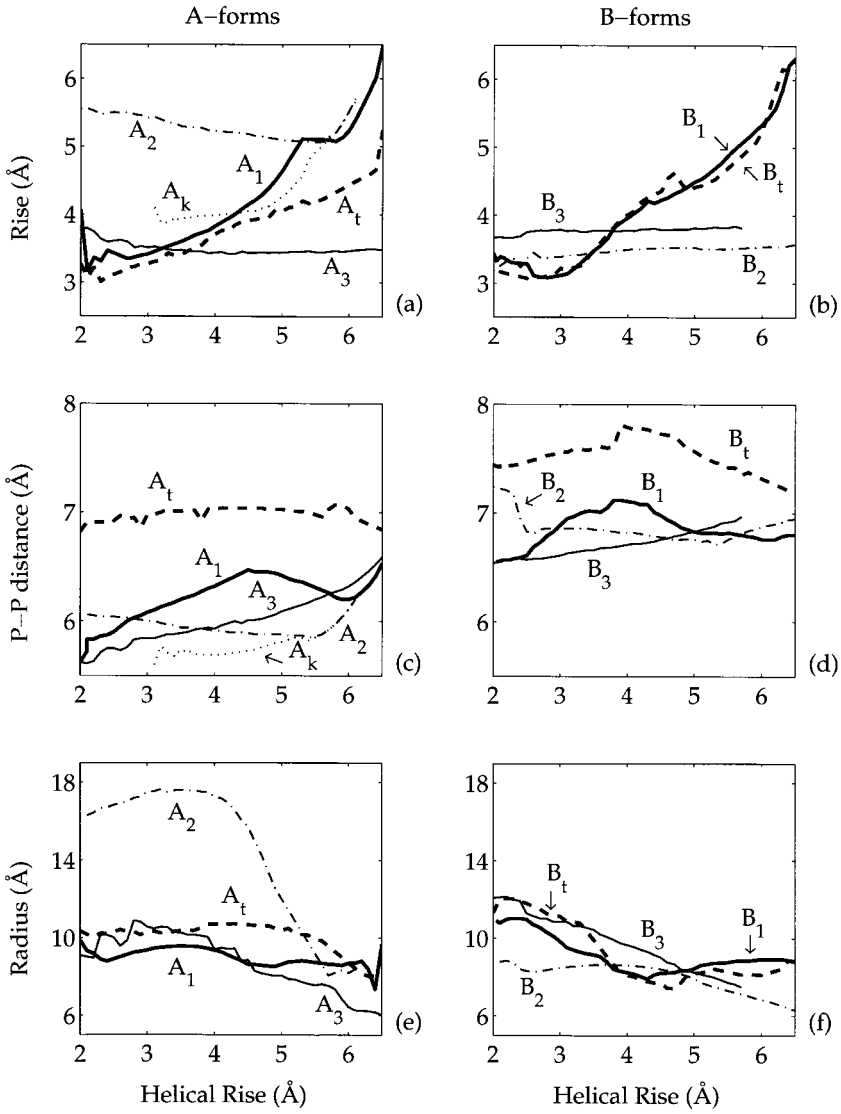

Figure 4. Molecular distances: (a) and (b) base-pair Rise; (c) and (d) P···P virtual bond distance (v); and (e) and (f) phosphorus helical radius (r); characterizing the stretching/compression profiles of A and B conformational families under the ''high salt'' conditions of Figure 1.

sequence independence of this important correlation. In addition, the local Rise of A_1B_1 helices (Figure 4(a) and (b)) increases with duplex stretching along intuitively expected lines, i.e. imposed global extension increases the spacing between base-pair planes. The added base-pair separation presumably contributes to the ''melting'' of the double helix at extreme extension (note the nearly perpendicular orientations, i.e. large propeller twisting, of complementary base-pairs in the overstretched A_1 and B_1 duplexes in Figures 6 and 7). The combined distortions of both base-pair and backbone geometry destabilize the A_1B_1 forms compared to other states at high global extension (see discussion, below). In particular, the difference in groove sizes among helical families is tied to their electrostatic stabilities, and is a key to understanding the nature of the sharp cooperative transition observed in DNA under extreme stretching (see the following section).

Of particular interest is the correspondence between our simulations of B_1-DNA stretching and the conformational patterns associated with localized, protein-induced elongation of the double helix in the high resolution nucleosome core particle structure (Luger *et al.*, 1997). The shorter of the two halves of the nucleosome-bound DNA, the so-called 72 bp fragment, spans the same contour length as its longer (73 bp) counterpart. The added extension of the shorter DNA is concentrated within a 12 bp section, covering the same region of the H3-H4 tetramer as a 13 bp region in the longer half; see so-called sites 2 and −2 in Figure 4(d) of Luger *et al.* (1997). The modest increase in helical rise, which allows the 12 bp fragment to occupy the same space as the 13 bp piece, leads to increases of Twist and base-pair inclination, as well as to narrowing of the minor groove entirely consistent with our simulations (Figures 2(b), 3(d), and 5(d)).

A_tB_t hyperfamily

Structures in the A_tB_t hyperfamily, distinguished by *trans* (t) arrangements of their P—O5′ and C5′—C4′ (α and γ) backbone torsions, respond

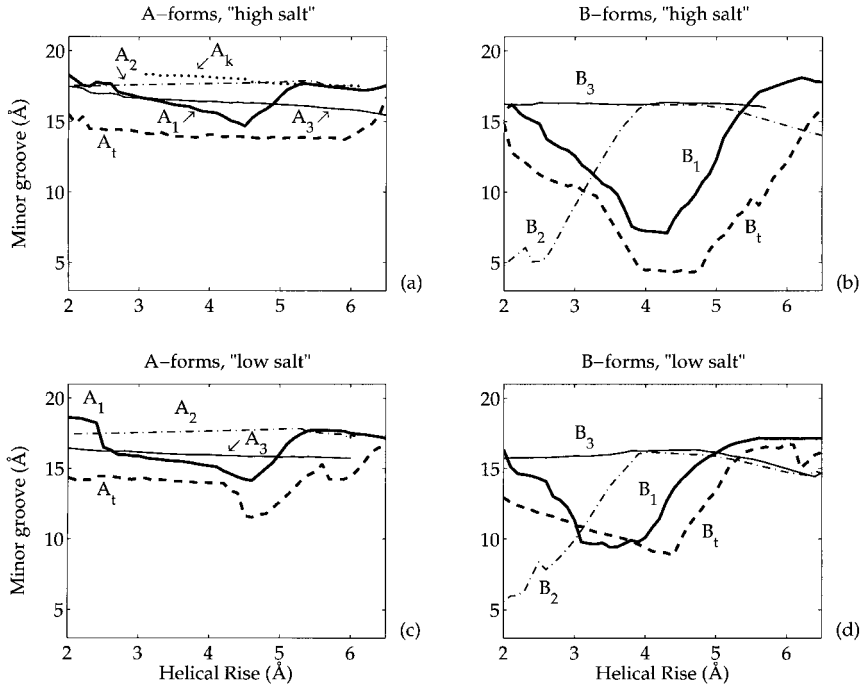

Figure 5. Minor groove widths, measured as the minimal distance between phosphate groups from complementary strands, that characterize the stretching/compression profiles of the *A* and *B* conformational families at (a) and (b) ''high'' and (c) and (d) ''low salt'' in Figure 1.

much like canonical A_1 and B_1-DNA to stretching and compression (compare heavy broken A_tB_t curves with continuous A_1B_1 curves in Figures 2-5). This conformational variant, which was featured in the original Watson-Crick fiber diffraction model (Crick & Watson, 1954) and is seen in ~6% of all dimer steps in currently determined *A*-DNA crystal structures (Schneider *et al.*, 1997), increases P···P backbone distances and reduces glycosyl torsion angles χ compared to their values in the A_1B_1 hyperfamily (Table 1 and Figure 4(c) and (d)). Similar structures were predicted in early conformational energy studies (Olson, 1976; Yathindra & Sundaralingam, 1976; Zhurkin *et al.*, 1978), and also suggested on the basis of fiber diffraction data (Chandrasekaran *et al.*, 1980).

The infrequent occurrence of A_t conformations in the X-ray literature is consistent with the energy difference computed here between unstretched A_t *versus* the more favorably base-stacked A_1 duplex (Figure 1(a), ''high salt''). By contrast, the absence in the X-ray database of B_t conformers, which are fairly close in energy to unstretched B_1 forms under both ''high'' and ''low salt'' conditions (Figure 1(b) and (d)), suggests possible limitations of the present energy analysis. The inclusion of explicit water molecules and counterions in the computations might raise the energy of the unstretched B_t form. We find no significant change in the relative energies of B_1 and B_t conformers, however, in our ''continuous medium'' simulations, when phosphate charges are reduced to levels predicted by counterion condensation theory (Manning, 1978).

A_2B_2 hyperfamily

Helices in the A_2B_2 hyperfamily adopt the conformational switch first characterized in dimer steps at the ends of *B*-DNA dodecamer duplexes (Grzeskowiak *et al.*, 1991) and anticipated in early DNA computations (Zhurkin *et al.*, 1978; Olson, 1982b). We use the subscript 2 to emphasize the conformational similarity of our computed structures with these so-called *B*(II) states where the εζ rotations about the C3′—O3′ and O3′—P bonds adopt g^-t combinations rather than the tg^- arrangements typical of canonical A_1B_1 duplexes, the β angle about the O5′—C5′ bond decreases by 30-70°, and the χ angle orienting the sugar and base rises by 40-60° to the *high anti* range (see Table 1). The combined effect of all four torsion angle changes is a large-scale variation in local and global base-pair geometry. For example, the stretching-induced responses of Roll and helical inclination in the B_2 helices are completely opposite from those in the B_1 family; compare heavy continuous *versus* thin broken curves in Figures 2(d) and 3(d). In further contrast to the canonical B_1 structure, the minor grooves of the B_2 duplexes show little, if any, variation upon extension beyond ~4 Å/bp and a significant decrease upon extreme compression (Figure 5(b)). Slide, Rise, and total energy are similarly unaffected by duplex stretching (Figures 1, 2(f), 4(b)). Importantly, the B_2 conformation becomes energetically more favorable than all other ''high salt'' *B*-forms when stretched to ~5.4 Å/bp.

The A_2 conformation, by contrast, is the least energetically favorable of all *A*-DNA structures under both ''high'' and ''low salt'' conditons (Figure 1(a) and (c)). There are, however, some

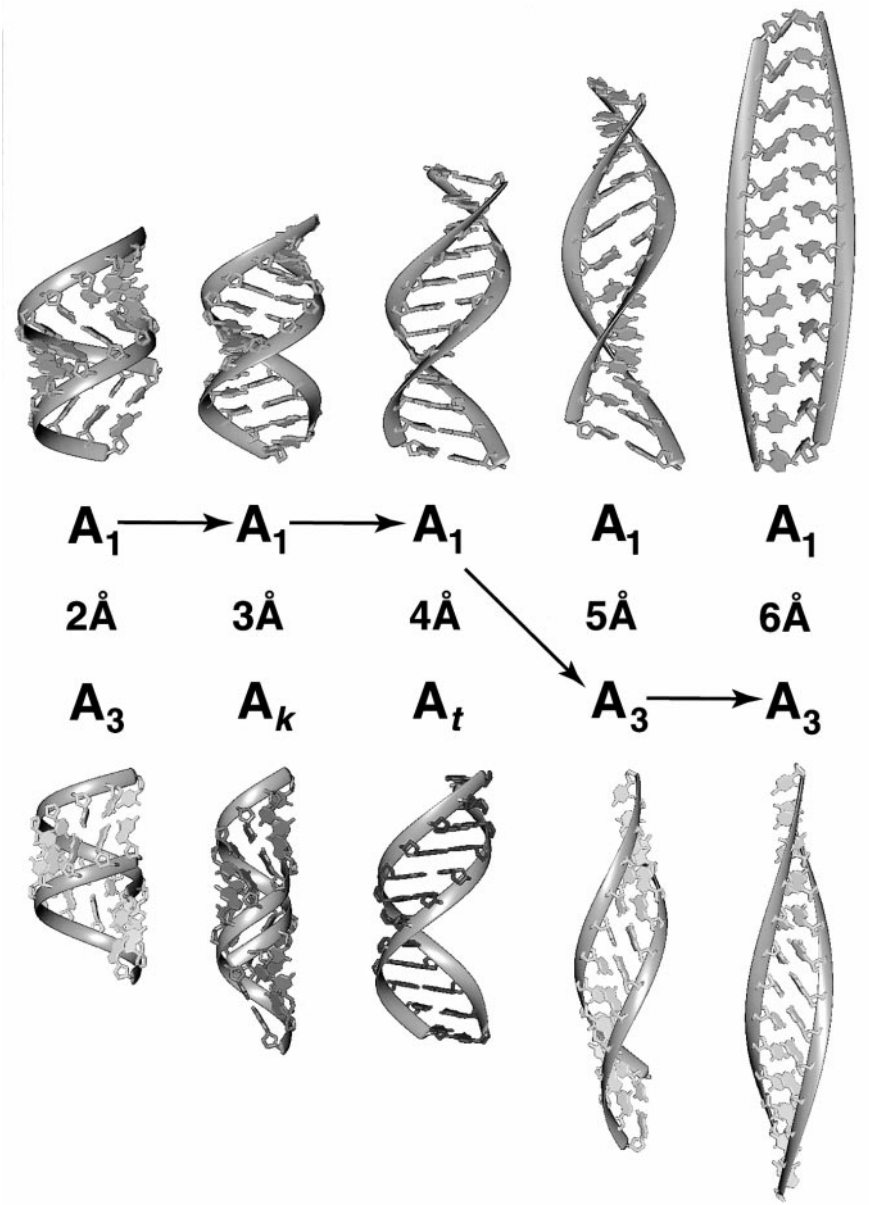

Figure 6. Ribbon representations, generated with `MidasPlus` (Ferrin *et al.*, 1988), of the simulated stretching of *A*-DNA structures. Side views of the lowest energy structures at the specified values of helical rise are shown. The arrows designate the putative conformational transitions suggested by the energy profiles in Figure 1. Chains are oriented so that the minor groove is always shown in the upper part of each duplex. Sugar-phosphate backbones of duplex structures are represented by gray ribbons, and base-pairs are color-coded according to conformational family: cyan, canonical A_1 and B_1 structures; magenta, A_3 and B_2 forms; ochre, A_k duplex; green, B_t conformations.

relatively stable, but highly kinked *A*-like forms with $\varepsilon\zeta$ angles confined to the g^-t domain which bear close resemblance to the dimer steps of DNA complexed to the TATA-box binding protein (see discussion of these A_k states in Table 2 in a later section).

A_3B_3 hyperfamily

The A_3B_3 structures, though similar at the backbone level to the A_1B_1 hyperfamily, differ dramatically from all other duplexes in terms of global base-pair orientation and the associated responses to stretching and compression. Large-scale changes in the glycosyl torsions χ to the *high anti* range completely untwist neighboring residues, significantly incline and displace base-pairs with respect to the global helical axis, and appreciably widen the minor groove (thin continuous curves in Figures 2(a), (b), 3(c)-(f) and 5(a), (b)). These unusual ''vertical helices'', anticipated in early theoretical work (Olson & Dasika, 1976; Olson, 1977), preserve both the hydrogen bonding between complementary residues and the overlap of successive base-pairs during the entire range of global extension (see Slide and Rise values in Figures 2(e), (f) and 4(a), (b)). The A_3B_3 forms resemble the A_2B_2 structures by

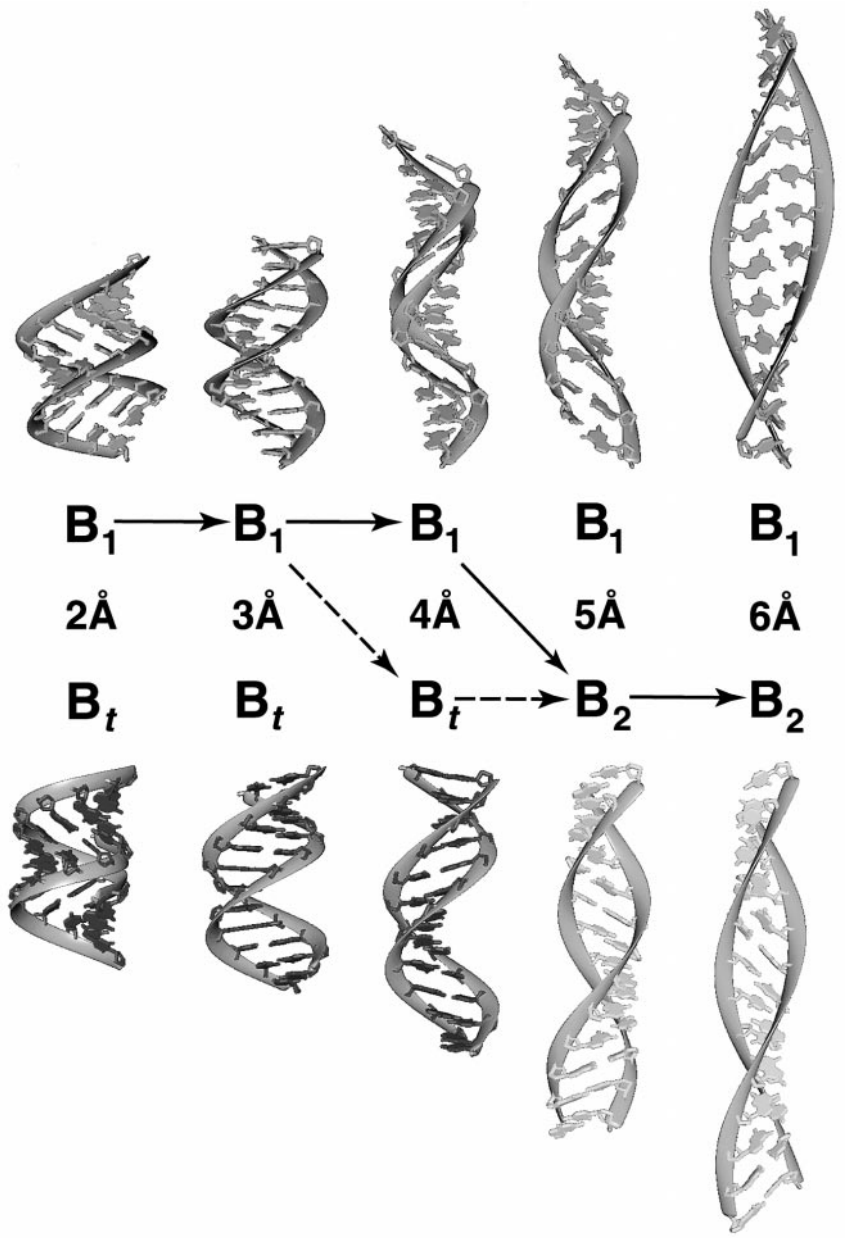

Figure 7. Ribbon representations of simulated *B*-DNA stretching. See the legend to Figure 6.

stretching *via* Twist and Roll with limited variations in Slide, Rise, and minor groove width (Figures 2(a)-(f), 4(a) and (b), and 5(a) and (b)) and in becoming energetically advantageous only at high extension (Figure 1(a) and (d)).

Over-stretched A_2B_2 and A_3B_3 duplexes are quite different from extended A_1B_1 and A_tB_t forms with undistorted and negatively inclined (positively rolled) base-pairs in the former cases and partially ''melted'' and positively inclined (negatively rolled) base-pairs in the latter (Figures 2(c), (d), 3(c), (d), 6, and 7). All four hyperfamilies converge in the limit of maximal extension to a common ladder-like backbone with comparable values of helical twist, *x*-displacement, P$\cdots$P virtual bond distance, global radius, and minor groove width but with altered base-pair geometry.

Stretching-induced cooperative transitions

The treatment of idealized DNA homopolymers undertaken here precludes detailed analysis of the cooperative transitions between different families of helical forms, such as the $B \leftrightarrow A$ transition (Ivanov *et al.*, 1974), characterized by irregular high-energy ''boundaries'' between interchanging conformational fragments. On the other hand, the variation of energy with helical rise within isolated families (as analyzed here) has the following advantage. As seen in Figure 1, the conformational states of lowest energy depend upon the degree of stretching. Thus, one can imagine how the DNA backbone interconverts from one conformational family to another as the energy profiles intersect. These large-scale transitions, including interchanges between extended helices (Figures 6 and

Table 2. Range of conformational parameters computed for duplex stretching and observed in TBP-bound DNA crystal complexes

Family	P (deg.)	χ (deg.)	Twist (deg.)	Helical twist (deg.)	Roll (deg.)	Helical inclination (deg.)	Slide (Å)	Helical x-displacement (Å)	Rise (Å)	Helical rise (Å)
A_3[a]	−10 to 15	99 to 105	−8 to −2	27 to 33	26 to 34	−107 to −95	3.0 to 4.3	7.7 to 9.5	3.4 to 3.8	2.0 to 4.0
B_3[a]	140 to 160	133 to 135	−5 to 7	29 to 32	30 to 33	−77 to −99	2.6 to 3.2	5.2 to 7.9	3.7 to 3.8	2.0 to 4.0
TATA	3 to 109[c]	76 to 100[c]								
Unkinked	96 to 214[d]	103 to 143[d]	1 to 27	20 to 41	7 to 31	−87 to −16	−0.6 to 2.4	−3.0 to 9.2	2.9 to 3.9	1.1 to 4.4
A_k[b]	24 to 25	32 to 43	20 to 25	34 to 42	28 to 38	−60 to −53	1.0 to 2.0	3.3 to 4.1	3.9 to 4.1	3.0 to 4.0
TATA	6 to 179[c]	17 to 113[c]								
Kinks	137 to 176[d]	79 to 130[d]	14 to 31	40 to 51	38 to 53	−70 to −49	−1.5 to 0.7	4.7 to 6.3	4.3 to 5.4	1.1 to 3.3

Parameters derived from crystal structures taken from the Nucleic Acid Database (Berman *et al.*, 1992) and described (Y. Kim *et al.*, 1993; Kim & Burley, 1994; Juo *et al.*, 1996; Nikolov *et al.*, 1996; Tan *et al.*, 1996).

[a] A_3 and B_3 structures in the Table are restricted to stretching in the range 2.0-4.0 Å/bp helical rise, rather than their complete stability ranges (see the text).

[b] "Kinked" structures (A_k) are limited to helical rise values between 3.0 and 4.0 Å/bp , whereas the overall range of stable A_k structures corresponds to 3.0-5.5 Å/bp helical rise.

[c] Limits of variation for the yeast (Y. Kim *et al.*, 1993; Tan *et al.*, 1996), *Arabidopsis* (Kim & Burley, 1994), and 1.9 Å resolution human (Nikolov *et al.*, 1996) TBP-DNA complexes.

[d] Limits of variation for the 2.9 Å resolution human TBP-DNA complex (Juo *et al.*, 1996), except for one thymine deoxyribose where $P = 266^\circ$.

7), are reminiscent of those implicated by direct stretching experiments (Bensimon, 1996; Cluzel *et al.*, 1996; Smith *et al.*, 1996; Baumann *et al.*, 1997).

Conformational interchanges

At the starting point of the calculations where the double helix is fixed in the canonical *A* and *B*-DNA forms (with respective helical rise of 3.0 and 3.3 Å/bp in poly(dG)·poly(dC) and of 2.9 and 3.2 Å/bp in poly(dA)·poly(dT)), the A_1B_1 energy is the global minimum. Upon stretching, the energies of the canonical helices increase while the energies of other conformational forms, such as the B_t duplex, decrease (see Figure 1). The *B*-DNA backbone apparently converts to the Watson-Crick-like B_t form at 3.8-4.0 Å/bp helical rise, when the energy of the B_t structure becomes lower than that of the B_1 form. As noted above, the computed preference for B_t forms could reflect our simplified computational approach which excludes explicit treatment of water molecules and counterions. This unusual over-twisted (B_t) helix with low helical radius and narrow minor groove is stabilized only under high ionic strength conditions (see Figures 1(b), 5(b), and 7). The minor groove is appropriately increased when the low salt model is considered (Figure 5(d)).

As we continue to stretch the DNA, the energy of the B_2 form approaches that of the B_1 and B_t families, and at ~5.5 Å/bp the B_2 form becomes energetically favored over the B_t helix (Figures 1(b), (d) and 7). The energy profiles thus predict a two-step $B_1 \rightarrow B_t \rightarrow B_2$ transition, where each step entails certain rearrangements of conformational parameters. The $B_1 \rightarrow B_t$ transition involves the correlated switching of α and γ torsions (Table 1) and little, if any, reorientation of neighboring base-pairs (Figure 2(b), (d), and (f)). The $B_t \rightarrow B_2$ step, by contrast, entails relatively minor changes in backbone torsions (ε, ζ, β), but large (*anti* to *high anti*) variations in the glycosyl rotation χ and accompanying sharp jumps in base-pair Roll and Slide (~40° and ~4-5 Å, respectively), which are expected to introduce an extremely high energy barrier at the borders between the two equi-potential forms (Olson *et al.*, 1999; see Figure 7). We therefore hypothesize that the "energetically expensive" switch into the B_2 conformation may account for the sharp phase transition observed at 60-70 % DNA extension (Cluzel *et al.*, 1996; Smith *et al.*, 1996). In other words, the concerted ~40° rotations of base-pairs about their long axes in the $B_1/B_t \rightarrow B_2$ transition offer an appropriate mechanism to account for the "domino effect" implied by the cooperative force-dependent stretching of DNA. Note that our data further predict that the B_2 helix becomes more favorable than all other *B*-type structures at 5.35 Å/bp helical rise under "high salt" conditions, and at 5.7 Å/bp at "low salt" (Figure 1(b) and (d)), degrees of extensions closely corresponding to the degree of DNA stretching at the observed transition.

On the other hand, if we consider the lack of crystallographic examples of B_t duplexes noted above as an argument against the calculated energetic preferences of B_t states in the range 4.0-5.5 Å/bp, we conclude that the stretching of the canonical B_1 duplex can also occur directly *via* a $B_1 \rightarrow B_2$ transition at ~5.0 Å/bp. In this case the calculated helical rise at the transition point is ~10 % less than the observed extension, 5.4-5.6 Å/bp (Cluzel *et al.*, 1996) or ~5.8 Å/bp (Smith *et al.*, 1996). The conformational mechanism, however, remains the same: large rearrangements of sugars and bases about the glycosyl linkages.

One of the factors leading to the steep increase in energy of B_1 and B_t forms upon stretching is the electrostatic repulsion between duplex strands (note the minima in minor groove widths at 4.0-

5.0 Å/bp under high salt conditions in Figures 5(b) and 7). As mentioned above, the decrease in groove size is linked to the negative Roll (-25° to -30°) in B_1 and B_t forms (Figure 2(d)). The reorientation of base-pairs in the extended B_2 duplex results in a positive Roll (Figure 2(d)), which widens the minor groove (Figure 5(b) and 7) and accordingly diminishes electrostatic interactions. Thus, the transition is expected to depend on environmental conditions which modulate the polyelectrolyte character of DNA (see below).

The calculations predict a somewhat different scenario for the stretching-induced transitions of *A*-type DNA. The A_t form is never optimal under ''high salt'' conditions. Thus, the A_1 form dominates over the stretching interval between 2.0 and 5.1-5.3 Å/bp (Figure 1(a)) until the A_3 form becomes lower in energy. Like the *B*-type DNA stretching-induced transition, the $A_1 \rightarrow A_3$ conversion also entails significant rearrangements of the glycosyl torsion χ from the *anti* to *high anti* range (Table 1). Furthermore, the extended A_3 duplex has the same strong positive Roll, and negative helical inclination as the over-stretched B_2 duplex (Figures 2(c), (d) and 3(c), (d)), meaning that the collective rotation of base-pairs also encounters a high energy barrier during the $A_1 \rightarrow A_3$ transition (Figure 6). Finally, the larger minor groove of the extended A_3 phosphodiester backbone relieves electrostatic repulsions brought about by the slight, stretching-induced narrowing of the A_1 minor groove (Figure 5(a) and (c)).

Thus, the stretching of both *A* and *B*-type helices simulated here is in principle consistent with the extreme cooperativity observed in single molecule manipulation experiments (Cluzel *et al*., 1996; Smith *et al*., 1996). At present, it is impossible to say whether the final stretched DNA states belong to *A* or *B*-type structures. The observed decrease in cooperativity of the transition in the presence of alcohol (Smith *et al*., 1996) may indicate the involvement of *A*-helices in the process, but this kind of reasoning is inconclusive and further experiments are needed to clarify the issue (see the section on electrostatic and solvation effects, below).

Comparison with published studies

As pointed out above, highly stretched A_2B_2 and A_3B_3 structures are similar to one another, but significantly different from extended A_1B_1 and A_tB_t helices. The switch from A_1B_1 and A_tB_t structures to the extended A_2B_2 and A_3B_3 states introduces a conformational barrier, much like a temperature-induced melting transition or a solvent-induced $B \rightarrow A$ or $B \rightarrow Z$ transformation, that must be crossed before the DNA can be extended further. Importantly, the stretched DNA double helices reported in previous computer simulations-the ''S-ribbon'' from the Lavery group (Cluzel *et al*., 1996; Lebrun & Lavery, 1996) with bases roughly perpendicular to the helical axis, the ''S-ladder'' by Konrad & Bolonick (1996) with negatively inclined bases, and Lavery's narrow fiber also with negatively inclined bases (Lebrun & Lavery, 1996), share common features with our extended forms. For example, base-phosphate contact distances (between N1 of purine or N3 of pyrimidine and the 5′-P of the same nucleotide) in our stretched B_2 and A_3 structures correspond closely with numerical values reported by Konrad & Bolonick (1996) for the ''S-ladder''. These extended duplexes, illustrated in Figures 6 and 7, adopt novel, energy stabilized cross-strand stacking interactions with a given base on one strand associated with the complementary base of the next residue, i.e. base (i) on one strand interacts with the complement of base ($i-1$). These unusual contacts dominate the energy and are responsible, in part, for the shallow minima in the A_2B_2 and A_3B_3 energy profiles at 5.5-6.0 Å/bp (Figure 1).

At the same time, there is an important difference between our results and those obtained by others. The smooth profiles of energy *versus* helical rise presented here reveal potential pathways of cooperative conformational interconversion undetected in previous work. In distinction to Lebrun & Lavery (1996) and Konrad & Bolonick (1996), we make no assumptions as to how the level of chain extension is achieved so that all states with the same helical rise are directly comparable. In other words, based on our energies we can conclude that B_t is the best among all possible structures of 5.2 Å/bp extension, whereas B_2 is the best for 5.4 Å/bp (Figure 1(b)). This comparison is made possible primarily because of our selection of the helical rise, a natural stretching coordinate with clear ''macroscopic'' meaning, as an independent conformational variable.

By contrast, it is not clear whether duplexes stretched by different means (e.g. by pulling on 5′-5′ *versus* 3′-3′ ends) can be directly compared and related to experiment. For example, consider the two forms of poly(dGC)·poly(dGC) chains pulled *via* 3′ or 5′ ends in Figure 2(a) of Lebrun & Lavery (1996). These two families of structures, one with positive and the other with negative base-pair inclination, have equivalent deformation energies at relative DNA extensions of 1.7, 1.9, and 2.1. The roughly parallel increase in energy of these two forms with stretching argues against any phase transitions.

Our identification of multiple conformational forms at a given degree of duplex extension stems from the fine (0.05 Å or less) incremental variation of helical rise used in our calculations. The larger (0.5 Å) translational steps taken by Lebrun & Lavery (1996) in energy minimization yield jagged plots of energy *versus* chain extension that miss the distinctions that we find among conformational families (compare their Figure 2 with our Figure 1).

The stretching-induced transition with an abrupt increase in internal energy calculated by Konrad & Bolonick (1996) also bears some resemblance to the cooperative transition found here. Dimer steps in their over-stretched model have large positive

Slide values much like the steps in our extended A_3 and B_2 states. Their transition state further corresponds to approximately the same degree of extension computed here (~1.75 times the contour length of their unstressed duplex). (Local and helical parameters were computed by us with the CompDNA software (Gorin *et al.*, 1995), i.e. with the same definitions used in our energy simulations, from the coordinates available on the GeneVue, Inc. web page.)

On the other hand, there is a crucial difference between our model and that by Konrad & Bolonick (1996). We interpret the observed phase transition in over-stretched DNA as an abrupt large-scale change in the orientations of base-pairs (similar to a ''domino effect''). This kind of reorientation, associated with the conversion of the double helix from the $B_1/B_t/A_1$ conformation to the B_2/A_3 form, implies that dozens, if not hundreds, of base-pairs must rotate simultaneously to avoid formation of the high-energy ''boundaries'' between different helical states. The ''mechanistic'' reason behind the cooperative transition predicted by Konrad & Bolonick, however, is unclear. The DNA in their simulations is pulled at the 3′-ends during the complete course of stretching, and intermediate forms appear to have the same sign of base-pair inclination (see Figure 2 of Konrad & Bolonick, 1996). The duplex seemingly remains within the same family of conformations during the computer simulations.

Further analysis of the over-stretched model available on the GeneVue web site reveals significant deviations in torsional parameters compared to values observed in crystal forms. In particular, the sugar puckering lies in a prohibited range ($P = 290\text{-}330^\circ$) not even observed in the simplest mononucleosides (Gelbin *et al.*, 1996). Furthermore, most of the dimer steps in the model assume backbone linkages like the B_t form (i.e. *trans* $\alpha\gamma$ values), even though the inclination of base-pairs resembles that in the B_2 family. This mixed conformational picture is suggestive of an extremely high energy state located far from the structural transition ''path'' induced by DNA stretching. It is also unlikely that the correlated changes of α and γ angles computed in the Konrad & Bolonick study can account for the large cooperative effect observed experimentally (Bensimon, 1996; Cluzel *et al.*, 1996; Smith *et al.*, 1996; Baumann *et al.*, 1997), especially since such rotational combinations can occur independently in different nucleotides (Olson, 1981).

Electrostatic and solvation effects

The predicted stretching-induced transitions of *B*-DNA duplexes follow the same qualitative patterns under simulated low and high salt conditions (Figure 1(b) and (d)). That is, the *B*-form duplexes preferentially adopt canonical (B_1) conformations over the range of normal chain extension and convert upon added stretching to the B_t and B_2 forms. Indeed, the $B_1 \rightarrow B_t$ and $B_t \rightarrow B_2$ transitions occur at roughly the same degree of extension under the two ionic extremes. The decrease in ionic strength, however, changes the relative energies of the different hyperfamilies, enhancing their convergence to a common conformational form at extreme stretching. Compression of *B*-form DNA under ''low salt'' conditions also favors the B_t over the B_1 form. This preference is apparently influenced by the greater local spacing between successive phosphate groups in the B_t family characterized by extended *trans, trans* conformations of the $\alpha\gamma$ phosphodiester linkage (Zhurkin *et al.*, 1978).

The change in ionic conditions has a more pronounced effect on the stretching and compression of *A*-DNA. The variation in salt concentration shifts the relative energies of the A_1 and A_t forms, interchanging their conformational preferences at low and high salt (Figure 1(a) and (c)). Whereas the *A*-type helices preferentially adopt the canonical A_1 form at high salt, the A_t state with more widely separated phosphate groups dominates over nearly the complete range of stretching at low salt. The A_1 state is favored at low ionic strength over only a narrow range of helical rise (4.0-4.5 Å/bp). Both below and above this range, the A_t family is energetically preferable. In this respect, the *A* and *B*-forms behave similarly at ''low salt'' (Figure 1(c) and (d)).

Finally, our data suggest a putative interpretation of the observed dependence of the stretching-induced transition on hydration. The reduced water activity in 60 % ethanol almost eliminates the cooperativity of the transition (Baumann *et al.*, 1997). Based on our data, the effect of ethanol can be interpreted in two ways, neither of which excludes the other. First, the decrease in water activity facilitates the $B \rightarrow A$ transition (Ivanov *et al.*, 1974). Therefore, reduction in the cooperativity of DNA stretching in the presence of ethanol may be tied to the relative advantage of extended *A*-like forms with C3′-*endo* puckered sugars under these conditions (Figure 1(a)). Second, in the presence of alcohols the monovalent cations may stabilize B_1-duplexes with narrower minor grooves than the corresponding forms in aqueous solution (Ivanov *et al.*, 1973). In such cases, the reduced cooperativity can be linked to an ethanol-induced decrease in the barriers introduced by unfavorable inter-phosphate interactions in the B_1 and B_t-forms at 4.0-5.0 Å/bp stretching (Figures 1(b) and 7). Evidently, further experiments and theoretical studies are needed to clarify this issue.

TBP-bound DNA

The deformations of DNA brought about by stretching or compression can be compared with the extreme distortions of the double helix induced by the binding of proteins such as the TATA-box binding protein (TBP). Recent interpretations (Lebrun *et al.*, 1997) of the crystal complex suggest that the TATA box is ''locally stretched and unwound''

upon association with TBP. Our analysis, however, shows that the TATA-box conformation is not adequately described in such simple and intuitively defined terms. In fact, depending upon the choice of structural parameters, the DNA appears to be either stretched or compressed.

Local conformational distortions

The severe distortion of the TATA-box region has been described in terms of the distances between P atoms in different strands (Guzikevich-Guerstein & Shakked, 1995; Lebrun *et al.*, 1997). Specifically, the distances that are used to measure the major and minor groove widths of *B*-DNA, the 20.3 Å separation of 5′-P atoms on either side of a four base-pair double helical stretch and the corresponding 13.8 Å distance between 3′-P atoms, are found to change asymmetrically upon TBP binding. The distance between the 3′ termini of DNA in the TBP-complex increases dramatically to 28.8 Å, whereas the 5′-P···5′-P separation retains the *B*-form value of 20.3 Å, suggesting a "local" or asymmetric stretching of DNA in the TATA-box region (Guzikevich-Guerstein & Shakked, 1995; Lebrun *et al.*, 1997).

Interestingly, increased P-P distances between 3′-P atoms are characteristic features of the A_3B_3 hyperfamily (see Figure 5 and the above discussion of the unusually wide minor grooves in these helices). These distances, which remain nearly constant (30-32 Å) over the complete range of global stretching (2.0-7.0 Å/bp), along with the corresponding 5′-P···5′-P distances (19-22 Å), reveal the close resemblance of the A_3B_3 hyperfamily to TBP-bound DNA. Apparently, the duplex does not necessarily have to be stretched to reproduce the crystallographically observed increase in 3′-P···3′-P distances.

Helical deformations

Further examination of DNA dimer steps in the published TBP complexes (J. L. Kim *et al.*, 1993; Y. Kim *et al.*, 1993; Kim & Burley, 1994; Nikolov *et al.*, 1995, 1996; Juo *et al.*, 1996; Tan *et al.*, 1996) in terms of the helical rise uncovers a compression, rather than a stretching of base-pair steps. This parameter corresponds to the distance between consecutive base-pairs along the helical axis of the regular polymeric structure that is generated by successive repetition of a given dimer step (Miyazawa, 1961; Rosenberg *et al.*, 1976; see Methods). The helical rise computed with the `CompDNA` software (Gorin *et al.*, 1995) varies from fractions of an Ångstrom unit to slightly over 4.0 Å/bp in the seven dimer steps of the TATA box contacted by protein (see Figure 8). Indeed, some steps are even more globally compressed than those in the canonical *A*-DNA duplex.

Among the seven DNA steps involved in direct contact with TBP, steps 1 and 7 are strongly kinked into the major groove, as opposed to the five less

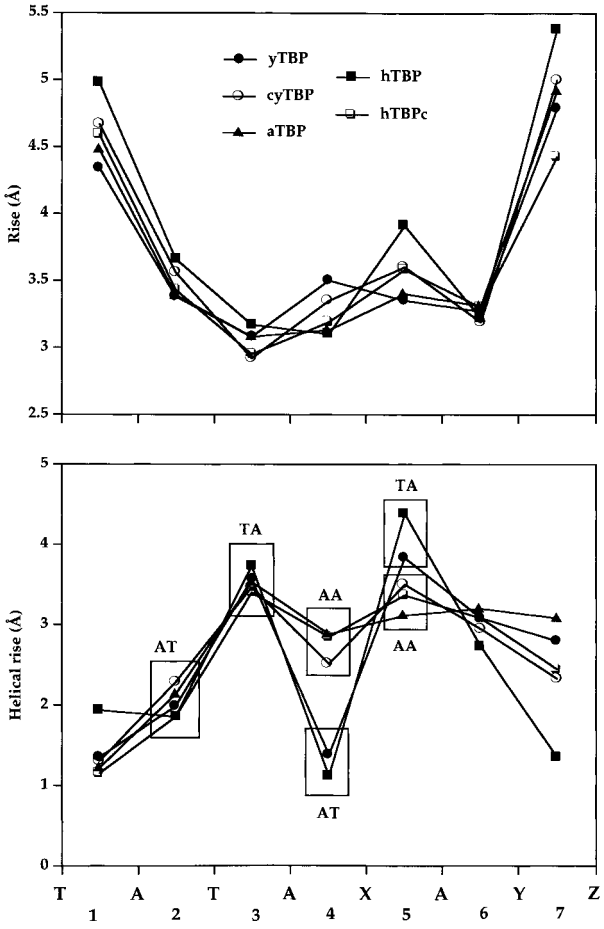

Figure 8. Variation of (a) local and (b) helical rise in five TBP-DNA complexes (Y. Kim *et al.*, 1993; Kim & Burley, 1994; Juo *et al.*, 1996; Nikolov *et al.*, 1996; Tan *et al.*, 1996). In those cases where two different structures of a given complex are given (Y. Kim *et al.*, 1993; Kim & Burley, 1994), the average values are presented. Notations: filled circles, yTBP (Y. Kim *et al.*, 1993); open circles, cyTBP (Tan *et al.*, 1996); filled triangles, aTBP (Kim & Burley, 1994); filled squares, hTBP (Juo *et al.*, 1996); open squares, hTBPc (Nikolov *et al.*, 1996). The sequences in positions 5, 6 and 8 are as follows: X = A for cyTBP, aTBP, and hTBPc, X = T for yTBP and hTBP; Y = T for hTBP and Y = A for others; Z = A for yTBP and hTBP; Z = C for cyTBP, Z = G for aTBP and hTBPc. The rectangles illustrate the sequence-dependence of the helical rise for steps 2-5: AT < AA < TA.

severely deformed intermediate steps 2-6. We thus consider these two groups of steps separately. Table 2 compares selected conformational parameters of TBP-bound DNA dimer steps against corresponding values found in our simulated A_3B_3 and A_k structures (where the A_k stands, as noted above, for the kinked structures in the A_2 family). The data for the computed structures are reported in terms of the ranges of conformational parameters accompanying an increase of helical rise between 2.0 and 4.0 Å/bp, an interval corresponding approximately to the variation in helical rise observed in the TBP-DNA co-crystals. The calcu-

lated A_3B_3 and A_k homopolymers span a slightly narrower range of helical rise compared to individual TBP-bound DNA steps. Steric repulsions prevent compression of the idealized A_3B_3 structures below 2 Å/bp helical rise, and the A_k homopolymers are only stable in the 3.0-5.5 Å/bp range (see thin dotted curves in Figure 1(a)).

Importantly, the DNA steps between TBP-induced kinks follow conformational trends found upon stretching the A_3B_3 hyperfamily. These characteristics include: (1) roughly constant local Rise (Figure 4(a) and (b)); (2) a decrease in Roll linked to an increase in base-pair inclination with respect to the helical axis (Figures 2(c), (d) and 3(c), (d)); (3) an increase in Slide tied to a decrease in helical *x*-displacement (see Figures 2(e), (f) and 3(e), (f), and discussion below); and (4) an increase in local Twist but almost no change in helical twist as the chain is extended (Figures 2(a), (b) and 3(a), (b)).

In most TBP-DNA complexes resolved to date, the intervening TATA steps are predominantly *A*-like in that they tend to adopt C3′-*endo* sugar puckering, although the puckering at some steps is C4′-*exo* (an intermediate pseudorotational state between the C3′-*endo* and C2′-*endo* puckers). Thus, the A_3 duplex with its *high anti* χ rotation and local untwisting appears to be an extreme version of TATA-box DNA, for which χ is *anti* at most steps (see Table 2).

The only exception to this assessment is the 2.9 Å resolution human TBP-DNA complex (Juo *et al*., 1996) where the steps between the two kinks exhibit the C2′-*endo* sugar pucker characteristic of *B*-like DNA. The glycosyl angle χ in this structure is relatively high compared to that of ''standard'' *B*-DNA, placing this form close to the B_3 family. The variation of base-pair step parameters in the human TATA-box DNA also resembles that of the calculated B_3 structures (see Table 2). Therefore, we conclude that both subsets of the calculated A_3B_3 hyperfamily, A_3 with C3′-*endo* sugars and B_3 with C2′-*endo* sugars, are close relatives of the experimentally observed TATA-box structures.

Sequence effects

Figure 8 further illustrates the effects of sequence on the structure of DNA in the TATA box. The helical rise values increase in the order: AT < AA < TA (see the boxes in Figure 8). Interestingly, a similar sequence dependence has been observed for local Twist (Juo *et al*., 1996). In particular, the central fourth step is most compressed and unwound when it is AT, with a helical rise of 1.0-1.3 Å/bp (Y. Kim *et al*., 1993; Juo *et al*., 1996). The corresponding helical twist is close to 20° (local Twist $\approx 0°$) and the base-pair (helical) inclination is $-80°$ to $-90°$ (data not shown), values consistent with the behavior of the A_3B_3 structures (Table 2). On the other hand, when the fourth step is AA, the dimer looks more like a canonical *A*-like form with helical rise 2.5-3.0 Å/bp in the yeast TBP-DNA structures (Y. Kim *et al*., 1993; Tan *et al*., 1996).

The unusually low values of helical rise for the AT steps correlate with the extremely high (7.7-9.5 Å) values of helical *x*-displacement (see Table 2). This parameter is identical to the displacement *D* used in earlier notations (Arnott, 1970), where a positive *D* value indicates the shifting of base-pairs toward the minor groove. Because of the tendency of the sugar-phosphate backbone to retain its optimal geometry (Gorin *et al*., 1995), an increase in *D* is linked with the decreases in helical rise and helical twist that accompany the *B*-to-*A* transition (Olson, 1976; Yathindra & Sundaralingam, 1976; Zhurkin *et al*., 1978). Notice, however, that the displacement *D* is about 4.5-5.0 Å/bp in the canonical *A*-form, whereas here *D* is as high as ~9.5 Å at the central AT step in the human TBP-DNA complex (Juo *et al*., 1996). In other words, the AT steps appear to be highly exposed in the minor groove. This pronounced sequence dependence in the TATA-box parameters is probably functionally significant for mutual protein-DNA fit, i.e. optimization of TBP-DNA interactions at the interface surface.

As for the kinks at either end of the protein-bound DNA fragment, the local Rise is appreciably high compared to the intermediate dimer steps (Table 2 and Figure 8). However, the corresponding helical rise is rather low, especially at the first step, which thus bears some similarity to extremely compressed conformations. In fact, local and helical base-pair parameters observed for the extreme terminal kinks closely match their calculated range and variation in A_k structures. For example, the high (over 40°) helical twist at these steps appears to reduce the duplex radius *r* and to induce large negative helical inclination.

Mechanics of DNA distortion

Finally, we conclude that despite the simplicity of our homopolymer model, which (i) ignores the mixed sugar puckering found at kinked TBP-DNA steps and accordingly biases the choice of backbone and glycosyl torsion angles; (ii) sets the Buckle angle to zero, although it varies noticeably in the TBP-DNA structures, especially at kinked dimer steps; and (iii) uses a different sequence, we closely mimic the structure of DNA in contact with TBP at both non-kinked (A_3B_3 forms) and kinked (A_k subfamily) steps. The close resemblance of the co-crystal DNA structures to our calculated forms (both stretched and compressed) further corroborates our conclusion that a simple ''stretching and unwinding'' mechanism of TATA-box recognition (Lebrun *et al*., 1997) does not reflect the complicated mechanics of DNA distortions observed in the complex with TBP. In fact, an intricate interplay between various base parameters leads to the unusual conformation of TATA-box DNA, which is neither *A* nor *B*. Furthermore, the nature of the conformational change depends on the structural

parameters used to describe the observed deformations: while P···P distances are suggestive of local chain extension (Lebrun *et al*., 1997), the two rise values reported here indicate that either the chain length remains practically constant, or is at most moderately stretched (local Rise), or that the structure is noticeably compressed (helical rise).

Our interpretation of the TATA-box structure also reflects differences in our definition of helical rise compared to values reported by others (Lebrun *et al*., 1997) using the Lavery-Sklenar `Curves` algorithm (Lavery & Sklenar, 1989). Our helical rise corresponds to the distance between consecutive base-pair centers along the axis of the mini-helix generated by these two residues (see Methods). By contrast, the inter-base-pair distance in `Curves` (Lavery & Sklenar, 1989) is measured along the normals to the base-pairs (with the rise an average of the distances measured along the two normal vectors). To compare the two sets of parameters, consider an ''ideal *B*-DNA'' with its helical axis passing through the base-pair centers, and local Roll = Tilt = 0°, Twist = 36°, Rise = 3.4 Å/bp. Now, if each pair is rotated about its long axis by the same angle (e.g. 20°), our helical rise does not change, since the base-pair centers and helical axis remain in exactly the same places. The rise computed with `Curves`, however, increases noticeably since the normal vectors are no longer parallel to the helical axis (see Lu & Olson (1999) for further details). This simple consideration accounts for much of the increase in rise found with `Curves` at highly kinked DNA steps compared to other analysis schemes (Werner *et al*., 1996; Fernandez *et al*., 1997; Lu *et al*., 1997; Lu & Olson, 1999), including the `CompDNA` routines (Gorin *et al*., 1995) used here.

RecA-stretched DNA

Our results can also be tied to the RecA-driven process of homologous recombination, where both single-stranded (ssDNA) and double-stranded (dsDNA) molecules are severely extended and underwound (Stasiak *et al*., 1981; Stasiak & Di Capua, 1982). During the exchange of genetic information the dsDNA from one chromosome recognizes and swaps strands with the homologous ssDNA fragment from a sister chromosome (for reviews, see West, 1992; Kowalczykowski & Eggleston, 1994). RecA, one of many known recombination proteins, binds ssDNA and forms a helical filament that facilitates ssDNA-dsDNA interactions and the search for homology (Egelman & Stasiak, 1986). One of the most intriguing characteristics of the RecA-DNA complex is its extreme stretching (helical rise ≈ 5.1 Å/bp) and unwinding (helical twist ≈ 20°) (Stasiak *et al*., 1981; Stasiak & Di Capua, 1982). Remarkably, the DNA rise and twist in the recombination filaments are nearly species-independent (Ogawa *et al*., 1993), suggesting that these distortions are functionally meaningful. Molecular models of ssDNA-dsDNA interactions (Zhurkin *et al*., 1994a,b) indicate that the stretching and unwinding may be advantageous for DN*A*-DNA homology recognition. In particular, the increased separation of bases, i.e. rise, reduces the chances of mismatch formation between single strand and duplex.

Stretched duplexes

In view of these data, it is interesting to consider whether our calculated stretched forms offer any clues to the recognition mechanisms involved in homologous recombination. The energetically optimal dsDNA forms corresponding to the RecA-induced helical rise of 5.1 Å/bp include: the A_1 and B_t duplexes at high salt and the A_t and B_t forms at low salt (Figure 1). We assume that our ''high salt'' models more closely mimic the environmental conditions in the center of the RecA filament, where there are numerous positively charged amino acids, especially in the vicinity of the L2 loop (Story *et al*., 1992). We further choose the *A*-like (A_1) over the *B*-like (B_t) form in the recombination model detailed below, given that a C2′-*endo* to C3′-*endo* conformational transition is detected in FTIR measurements of the RecA-three-stranded complex (Dagneaux *et al*., 1995). The reasoning behind our protein-DNA model, however, is applicable to any of the aforementioned conformations: A_1, A_t, B_1 and B_t, all of which share similar global features.

The energetically optimal A_1 form extended to 5.1 Å/bp, is characterized by positive helical inclination and negative Roll (Figures 2, 3, and 6). That is, the bases are oriented in the same way as in the classic *C*- and *D*-DNA fiber forms (Arnott *et al*., 1974; Selsing *et al*., 1975). The extremely wide major grooves of these helices (Figure 6) are expected to facilitate favorable interactions with ssDNA. An additional functional advantage of our extended A_1 model is related to the orientation of base-pairs in the duplex and the inclination of the duplex itself as it slides along the RecA filament in search of the homologous single strand (Figure 9; Howard-Flanders *et al*., 1987). The inclination of base-pairs in the extended A_1 duplex effectively cancels the inclination of the A_1 duplex itself with respect to the RecA filament, so that the base-pairs of dsDNA are nearly perpendicular to the overall RecA axis and nearly parallel to the bases of ssDNA (Figure 9). Such an orientation is expected to be functionally advantageous, facilitating base-base interactions, specific recognition, and strand exchange between double and single-stranded DNAs (Zhurkin *et al*., 1994a,b).

Recombination model

Consider in more detail two critical assumptions in the model presented in Figure 9: (i) the bases in ssDNA are nearly perpendicular to the RecA filament; (ii) the interaction between double and single-stranded helices occurs in the major groove

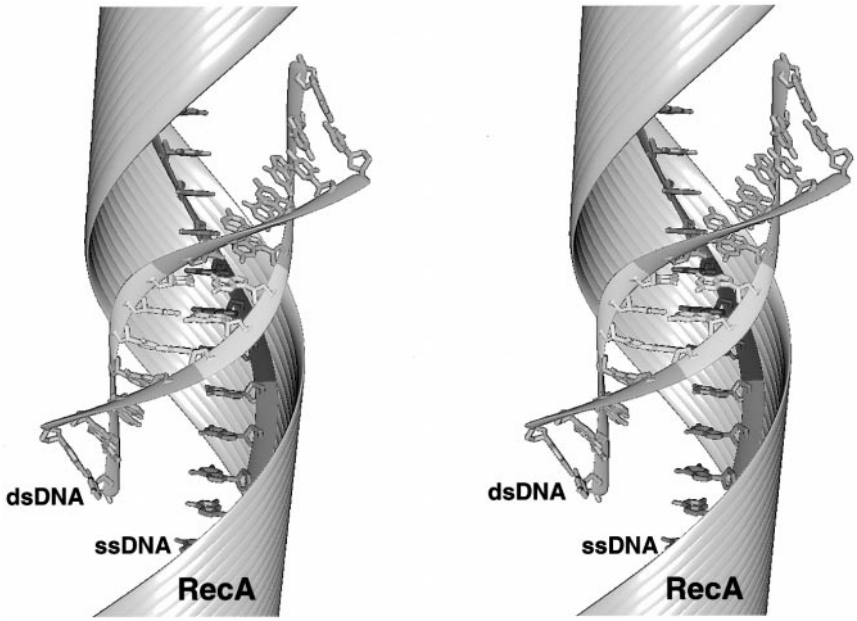

Figure 9. Model, in stereo, of stretched dsDNA interacting with the extended ssDNA localized in the center of the RecA filament (Egelman & Stasiak, 1986; Howard-Flanders *et al*., 1987). The RecA filament is shown schematically by gray ribbons. Notice that the bases in the duplex are inclined so as to facilitate interactions with the ssDNA bases known to be nearly perpendicular to the filament axis (Nordén *et al*., 1992, 1998). The large separation between the bases helps to avoid mispairing. The duplex is a member of the A_1 family extended to 5.1 Å/bp (Figure 7). The three base-pairs shown in magenta are in register with the three nucleotides of the single strand (green), facilitating ssDNA-dsDNA recognition in the major groove (Howard-Flanders *et al*., 1984) and subsequent strand exchange (Zhurkin *et al*., 1994a,b).

of the duplex. The nearly perpendicular orientation of ssDNA bases is consistent with the linear dichroism of single-stranded DNA bound to RecA (Nordén *et al*., 1992, 1998). The most unusual feature of the ssDNA is the large separation between the bases, on one hand, large enough to weaken the stacking interactions and allow for the swapping of bases between ssDNA and dsDNA and on the other, too small to allow water molecules to penetrate between the bases (Zhurkin *et al*., 1994b).

Importantly, this base positioning has been corroborated by recent NMR structures of several short ssDNA oligomers bound to RecA (Nishinaka *et al*., 1997, 1998), where the partially unstacked bases are separated by distances of ~5 Å along the base normals. The loss in stacking interactions in the NMR structure is partially neutralized by favorable sugar-base interactions (Figure 10), but is also accompanied by an unusually high helical twist ~50° (a value calculated by us with the `CompDNA` software (Gorin *et al*., 1995) from the coordinates deposited in the Protein Data Bank (Bernstein *et al*., 1977), PDB entry 3REC). While such over-twisting is apparently favorable in selected single-stranded DNA fibers (Leslie & Arnott, 1978), it contradicts the well known unwinding of long ssDNA in the RecA filament (Stasiak & Di Capua, 1982). The extreme overtwisting of bases in the NMR structure may stem from the very short fragments of ssDNA bound to RecA in that study (Nishinaka *et al*., 1997, 1998). Indeed, only one or two RecA monomers can form a complex with the three to six nucleotide ssDNA characterized by NMR. The association of multiple RecA proteins is apparently necessary for stabilization of the underwound DNA structure operative in recombination. Therefore, we do not incorporate the NMR-derived information directly in our modeling, but rather use a ssDNA computer model (Zhurkin *et al*., 1994a) consistent with the 5.1 Å/bp stretching and 20° twisting of RecA-ssDNA complexes (Stasiak *et al*., 1981).

This single-stranded structure, illustrated in Figure 10, matches the duplex DNAs found in this work and is moderately stabilized by sugar-sugar interactions. The sugars associate *via* an unusually short (3.7 Å) C2′···O4′ contact, where the V-shaped CH_2 group at the 2′-carbon tightly ''envelopes'' the oxygen atom of the adjacent deoxyribose. Such advantageous interactions are well known in the DNA crystallographic literature (Wahl *et al*., 1996; Mandel-Gutfreund *et al*., 1998). Geometrical analyses further show that this sugar-sugar contact is more favorable when the sugar pucker is C3′-*endo* (in accordance with FTIR measurements (Dagneaux *et al*., 1995)), and the sugars are deoxyriboses (in accordance with the low affinity of RecA for RNA (McEntee *et al*., 1981)).

The second assumption of the model in Figure 9, the recognition of ssDNA *via* the major groove of the duplex (Lacks, 1966; Howard-Flanders *et al*., 1984), is consistent with recent data obtained from the cleavage of RecA-bound DNA by ^{125}I-emitted

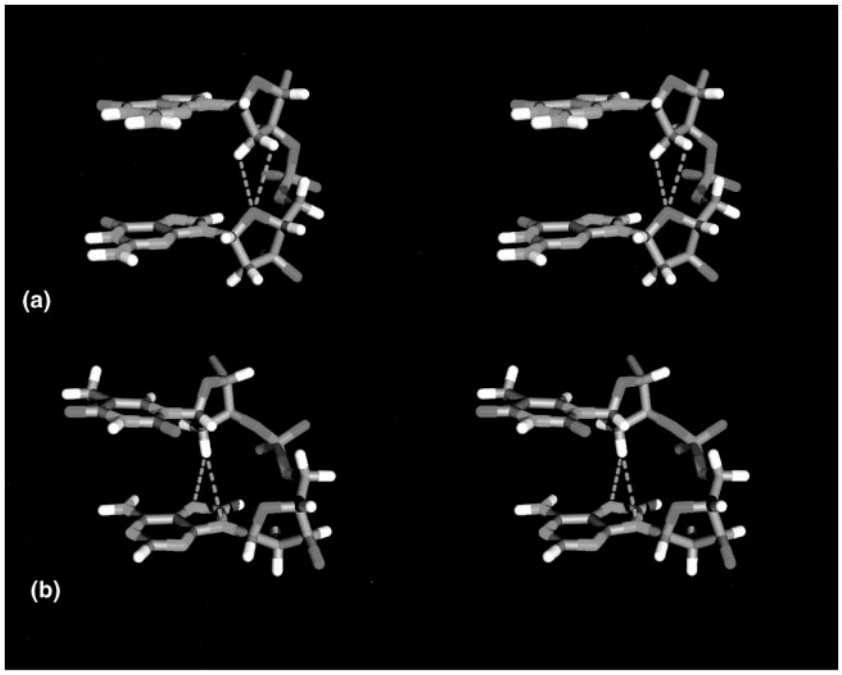

Figure 10. Comparison, in stereo, between (a) our energy optimized ssDNA dimer (Zhurkin *et al*., 1994a) and (b) the NMR-based model by Nishinaka *et al*. (1997, 1998). The yellow broken lines designate favorable contacts between the CH_2 group at the 2′-carbon and the O4′ atom in (a) and the (2′-carbon) $CH_2\cdots$base interaction in (b). Notice that the helical twist is 20° in (a) and 50° in (b). The helical rise is ~5 Å/bp in both cases (see the text for details).

Auger electrons (Malkov *et al*., 1999). This scheme also agrees with our findings for extended duplexes. As evident from the view into the minor groove in Figure 9, the right-handedness of the RecA-DNA filament (Ogawa *et al*., 1993) implies that the base-pairs must be rotated counter-clockwise around the dyad axis to form planar associations with the bases of ssDNA approximately perpendicular to the protein filament axis (Nordén *et al*., 1992, 1998). That is, to facilitate ssDNA-dsDNA recognition in the major groove, the base-pair inclination has to be positive. Importantly, such base-pair inclination is energetically favorable in the extended A_1B_1 and A_tB_t duplexes (Figure 3). By contrast, if the recognition were to occur through the minor groove, as suggested by others (Baliga *et al*., 1995; Podyminogin *et al*., 1995; Zhou & Adzuma, 1997), the base-pairs would have to be inclined negatively, which is energetically unfavorable for stretched canonical duplexes, but feasible for higher energy A_2B_2 and A_3B_3 helices (Figures 1 and 3).

Thus, our energy-optimized duplex model helps to tie together a number of previously known structural features of the RecA-DNA complex: (i) the RecA filament is right-handed (EM observation); (ii) the ssDNA and dsDNA in the RecA complex are stretched and unwound by 50% (EM data); (iii) the bases in ssDNA are nearly perpendicular to the RecA helical axis (linear dichroism measurements); (iv) the base-pairs in the energetically favorable extended duplex are inclined positively (Figures 1 and 3); (v) the recognition occurs in the major groove of the duplex (Howard-Flanders *et al*., 1984; Malkov *et al*., 1999). Taken together, these features yield a self-consistent stereochemical model for the ssDNA-dsDNA recognition brought in register by RecA proteins.

Conclusions

This study has attempted to find the intrinsic structural characteristics of double-helical DNA under extreme stretching and compression. These ''activated'' conformations, energetically unfavorable under ''normal'' conditions, are observed in complexes with proteins and other ligands and can be induced by novel micromanipulation techniques (Bensimon, 1996; Cluzel *et al*., 1996; Smith *et al*., 1996; Baumann *et al*., 1997). The critical assumption made here is that the intrinsic energy preferences of the double helix *per se* are used by the cellular machinery involved in DNA metabolism, and therefore, these natural preferences deserve extensive study. More specifically, we presume that DNA binding proteins take advantage of the natural DNA ''energy landscape'', and that when such proteins deform the duplex, they do so in an ''energy economical'' way. This assumption has already been vindicated by numerous crystallographic examples of regulatory proteins bending and twisting the DNA to which they bind (for reviews, see Travers, 1989; Steitz, 1990; Suzuki & Yagi, 1995; Dickerson, 1998; Olson *et al*., 1998). Here we have extended this assumption to the ''longitudinal'' deformations of DNA.

Although usually ignored, the longitudinal elasticity of double-helical DNA is an important structural factor facilitating its mutual fit with proteins. One of the best known examples is the variable (142-149 bp) length of DNA on the nucleosome core particle (Satchwell *et al*., 1986), an observation

possibly linked to local stretching and compression of the double helix in sequence-dependent wrapping around the histone octamer (see the detailed comparison in Results and Discussion of our calculations with the localized stretching of DNA observed in the crystallized nucleosome (Luger *et al*., 1997)). Here we wish to emphasize that the interdependence between DNA stretching and twisting revealed in our simulations may be important for protein-DNA recognition in general. Proteins usually recognize DNA sequences containing several ''consensus boxes'' separated by variable spacers. Therefore, assuming that bending deformations are limited, the duplex has to deform both in terms of twisting and stretching to facilitate effective protein binding to different DNA sites belonging to the same class of response elements. In this regard, it is important that our simulations predict positive correlations between *B*-DNA twisting and stretching, at least within the limit of 10-15% of the equilibrium values (Figure 2(b)). This correlation implies that the stretching of a duplex automatically leads to its over-twisting, making it relatively easy to fit such a fragment into the same space normally occupied by a longer sequence, e.g. the terminal base-pairs of a stretched 10-mer will easily superimpose on the end residues of an 11 bp chain.

Another manifestation of the longitudinal ''breathing'' of DNA is the variability in length of *Escherichia coli* promoters. The DNA spacer between the '' -35'' and the '' -10 boxes'' varies from 15 to 19 bp (Auble & deHaseth, 1988), a finding which suggests that the helical rise of DNA may vary inversely with the length of the spacer (to secure the best fit between promoter and RNA polymerase). A third example is the cyclic AMP receptor protein (CRP), which binds to consensus CACA and TGTG tetramers separated by distances varying from eight to ten base-pairs (Barber *et al*., 1993; Ivanov *et al*., 1995). At least part of this variability may stem from the longitudinal flexibility of the duplex. Notably, when CRP binds to a sequence with a longer spacer, the DNA undergoes a *B*-to-*A*-like transition (Ivanov *et al*., 1995). In other words, the double helix is compressed and unwound to improve the mutual fit between protein and DNA.

In principle, these ''activated'' forms of DNA cannot be analyzed by traditional all-atom computational approaches, i.e. Monte Carlo or molecular dynamics simulations of a macromolecule and the surrounding solvent environment, since the generation of these states requires high temperatures that would break base-pairs and rearrange the backbone. As a consequence, an alternative approach is needed to simulate the imposed high energy states.

To study such a complex system, the number of degrees of freedom must be reduced to a minimum and the independent parameters must be chosen so that they reflect the property of interest. For this purpose, we have used the generalized coordinates of bases as independent parameters, both in local and global (helical) coordinate frames, an approach which has proven to be effective in previous computer simulations of DNA bending and twisting (Zhurkin *et al*., 1991). The interchange between global and local representations is straightforward and offers a distinct advantage for computer simulations, as well as for visualization and interpretation of the obtained structures. Here we use global helical rise, the natural measure of the length of a stretched DNA molecule, as an independent variable, and study the conformational changes that accompany imposed chain extension and compression. This contrasts with previous work where chain extension is brought about by arbitrarily imposed forces on the sugar-phosphate ends of the molecule (Konrad & Bolonick, 1996; Lebrun & Lavery, 1996). Our choice of a natural ''reaction coordinate'' avoids problems associated with the correct parameterization of applied forces and the apparent dependence of chain conformation on the sites of pulling (Cluzel *et al*., 1996; Lebrun & Lavery, 1996). The simplifications introduced in our model, in particular the imposed symmetry constraints on homopolymers, clearly minimize sequence-dependent structural features of the duplex. The energetic and structural responses to stretching and compression are therefore very similar for poly(dG)·poly(dC) and poly(dA)·poly(dT) deformed under ''low'' and ''high salt'' conditions. Remarkably, the energy optimization yields most of the right-handed forms of DNA found in crystal and fiber structures of double helices and their complexes with proteins and drugs. Thus, a *comprehensive* classification of known double-helical forms becomes possible.

Stretching transition models

Our simulations reveal a conformational ''domino effect'' that accounts for the stretching-induced phase transition found at ~1.6 times the normal extension of *B*-DNA (Bensimon, 1996; Cluzel *et al*., 1996; Smith *et al*., 1996; Baumann *et al*., 1997). The model draws upon three principal findings. First, there are several hyperfamilies of duplex forms characterized by qualitatively different responses to imposed stretching at the base-pair level. In particular, the base-pair inclination decreases upon DNA stretching in the A_1B_1 and A_tB_t duplexes, but increases in the A_2B_2 and A_3B_3 forms (Figure 3(c) and (d)). Second, the first set of hyperfamilies (A_1B_1 and A_tB_t) dominates at ''normal'' degrees of DNA extension, whereas the latter set (A_2B_2 and A_3B_3) is preferable at ''extreme'' levels of deformation. Third, the two classes of structures are energetically equivalent at ''intermediate'' stretching (Figure 1). The substantial differences in base-pair inclination between these equipotential forms introduce a steric barrier at the boundaries between the different conformations (Olson *et al*., 1999). The considerable magnitude of this barrier compared to the energy of base stacking is consist-

ent with the sharp cooperative transition observed at this level of DNA extension.

Thus, the over-stretching of single DNA molecules can be described (to a first approximation) by the one-dimensional Ising model that has also been used to account for the $B \rightarrow A$ and $B \rightarrow Z$ transitions of the double helix (Ivanov *et al*., 1974, 1983). According to our calculations, the rearrangement of base-pairs caused by DNA stretching is greater than that in the $B \rightarrow A$ transition, but less than those needed to effect a $B \rightarrow Z$ transition (Ivanov *et al*., 1983; Olson *et al*., 1983).

The present calculations, however, incorporate an implicit model of the solvent environment and a two-well potential for sugar puckering, which preclude treatment of a possible stretching-induced switch of sugar conformation. Molecular dynamics or Monte Carlo simulations treating counterions and water molecules explicitly (Cheatham & Kollman, 1996; Yang & Pettitt, 1996; Sprous *et al*., 1998), may be able to determine whether the unusual A_tB_t structures found here over the range of 4.0-5.0 Å/bp are indeed favored (provided that the helical rise is appropriately restrained). Open questions include whether a *B*-to-*A*-like transition is facilitated upon stretching, and what role non-canonical forms may play in reducing the cooperativity of DNA stretching in the presence of dehydrating agents (see Results and Discussion). Clearly, further experiments and theoretical analyses are needed.

Significantly, comparable conformational ordering is found in simulations to be reported elsewhere of the longitudinal deformations of the alternating poly(dTA)·poly(dTA) and poly (dCG)·poly(dCG) copolymers. While the pyrimidine-purine dimer steps are slightly more responsive than the purine-pyrimidine steps to imposed global stretching and compression, the over-stretching transition occurs at an equivalent level of extension and entails the identical conformational interconversions, i.e. $A_1 \rightarrow A_3$ and $B_1 \rightarrow B_t \rightarrow B_2$, reported here. Base sequence thus plays the same sort of subtle role in the DNA over-stretching transition as it does in guiding the natural folding and deformability of the unstressed double helix (Olson *et al*., 1998). Moreover, the over-stretching transition occurs at nearly the same degree of extension in different chains, e.g. EMBL3 λ DNA (Cluzel *et al*., 1996) and λ bacteriophage DNA (Smith *et al*., 1996), but the amount of force needed to effect the transformation depends on sequence, e.g. ~65 pN for λ DNA and poly(dCG)·poly(dCG) *versus* ~35 pN for poly(dTA)·poly(dTA) (Rief *et al*., 1999). Our energy calculations reveal a common ''domino effect'' in the over-stretching of different sequences, but do not directly probe the conformational barriers separating stretched helices of equivalent internal energy.

Extended states in the complex with RecA protein

The extended DNA observed in recombination filaments (Stasiak & Egelman, 1988) is arguably the best known example of DNA stretching in biological systems. Remarkably, the degree of DNA extension in the complex, 5.1 Å/bp, is close to the helical extension (5.3-5.6 Å/bp) corresponding to the phase transition observed by DNA micromanipulations (Bensimon, 1996; Cluzel *et al*., 1996; Smith *et al*., 1996; Baumann *et al*., 1997), and to the interval (4.0-5.0 Å/bp) of inter-family $B_1 \rightarrow B_t$ and $B_1 \rightarrow A_1/A_t$ transitions predicted here. Apparently, DNA extended to ~5 Å/bp is characterized by an increased conformational lability: various helical forms coexist in the same region of DNA conformation space, thereby enabling interconversions between equi-potential forms (Figure 1(a) and (b)). The conformational variability appears to be critical for the mechanistically complicated process of homology recognition and strand exchange. In terms of detailed functional implications, the unwinding and displacement of base-pairs, linked in this study to DNA stretching, enhance contacts with the DNA interior and reduces the probability of mispairing between double and single strands (Figure 9).

Earlier, one of us has predicted a *B*-to-*A*-like transition in DNA stretched under the restrictions imposed by the geometry of the RecA-DNA filament in the so-called ''extended *R*-form'' (Zhurkin *et al*., 1994a,b). This transition, subsequently confirmed by the partial C2′-*endo* to C3′-*endo* switching of sugar puckering in FTIR and NMR studies of DNA-RecA complexes (Dagneaux *et al*., 1995; Nishinaka *et al*., 1997, 1998), also agrees with the present study: A_1-forms are favored but B_1-forms are disfavored at ~5 Å/bp stretching (Figure 1(a) and (b)). Like the earlier extended *R*-form model (Zhurkin *et al*., 1994b), the DNA stretched here has an increased hydrophobic surface compared to canonical *A* and *B*-form DNA. In principle, this feature could be an additional factor stabilizing the *B*-to-*A* switch in RecA-bound DNA, provided that water activity is decreased inside the RecA filament cavity. On the other hand, the decreased cooperativity of ''physically manipulated'' DNA stretching detected in the presence of ethanol (Baumann *et al*., 1997), can be related to the same *B*-to-*A* switch (see above). Thus, the RecA-DNA complex is perhaps another fascinating example of the synergism between the energy preferences observed for an isolated DNA molecule, and the ''design'' of the protein machinery that takes advantage of these preferences.

Compressed states and TATA-box deformations

Systematic compression of DNA homopolymers yields structures similar in terms of both dimeric and helical parameters to some of the deformed

steps observed in crystal complexes of DNA with TATA-box binding proteins (J. L. Kim *et al*., 1993; Y. Kim *et al*., 1993; Kim & Burley, 1994; Nikolov *et al*., 1995, 1996; Juo *et al*., 1996; Tan *et al*., 1996). Paradoxically, the seven DNA steps contacted by protein appear to be globally but not locally compressed, in addition to being strongly bent and unwound, as described earlier (J. L. Kim *et al*., 1993). The conformations of the two highly kinked DNA steps at the ends of the sequence resemble the A_k structures generated here as well as a compressed poly(dTA)·poly(dTA) copolymer to be described elsewhere. The intermediate non-kinked steps closely match compressed A_3 and B_3 structures which feature unusually wide minor grooves.

Despite the overall difference between RecA and TBP-induced deformations of DNA, the two systems have features in common that may be functionally important: base-pairs are accessible for recognition and single bases are easily flipped out of the duplex core much like the single bases in DNA-methyltransferase complexes (Roberts & Cheng, 1998). In the case of RecA, the flipping of bases is an essential part of the strand exchange process (Howard-Flanders *et al*., 1984; Zhurkin *et al*., 1994a,b; Nishinaka *et al*., 1997, 1998). In the case of TBP, the situation is somewhat more complicated. Here we must assume, in the absence of detailed structural information, that the distortions of DNA in the TBP complex are relevant to the interactions of DNA with RNA polymerase (see also Kim *et al*., 1997; Robert *et al*., 1998). According to conventional thinking (von Hippel, 1998), DNA base-pairs are somehow ''melted'' during RNA synthesis so that the RNA can form Watson-Crick pairs with the DNA bases. The disruption of DNA base-pairing may not necessarily follow the same stretching-induced pathway as in the RecA-DNA complex where the duplex is untwisted, both grooves are widened, and the bases are easily flipped out. DNA in the cell, especially in eukaryotes, is tightly packed, and longitudinal stretching is extremely expensive in terms of space: a stretched ''bubble'' containing ~20 base-pairs would be ~100 Å long if the base-pair rise were increased to that in the RecA filament.

Evolution has apparently chosen an alternate way to open the double helix: DNA is sharply bent, perhaps as in the TBP complex, and such deformations have two important consequences: (1) individual bases are readily flipped out of the base-pair stack (Ramstein & Lavery, 1988); (2) the large kink increases the distance between adjacent bases. RNA-DNA mismatches can thus be avoided along the same lines as in the RecA-stretched DNA model (Zhurkin *et al*., 1994a). Therefore, TBP-like manipulations of DNA, in principle, can achieve the same two goals which are secured by DNA extension, but without the waste of space. The paradoxical deformations, i.e. local extension *versus* global compression, of TBP-bound DNA described here perhaps reflect a novel mechanism of information readout from the double helix. Such a process is spatially economical, kinetically efficient, and error-resistant. Notably, while direct structural data on DNA in the active center of RNA polymerase are still missing, the deformations in DNA bound to the reverse transcriptase (Ding *et al*., 1998; Huang *et al*., 1998) are consistent with the above scenario. In particular, the DNA duplex undergoes a local $B \rightarrow A$ transition, bending sharply toward the major groove and thereby exposing base-pairs in the minor groove and facilitating their opening.

Finally, the ''activated'' forms of DNA analyzed in this study mimic the protein-induced deformations of the double helix in known complexes and are presumably tied to other biochemical processes. The distortions of DNA in such cases may lead to either complete or partial disruption of Watson-Crick base-pairs. In the absence of proteins, physical manipulations may yield these high energy ''activated'' states and thereby suggest as new models for the design of DNA-binding pharmaceuticals.

Methods

Conformational variables

Nucleic acid polymers are described in terms of generalized coordinates that define the positions and orientations of bases and sugar-phosphate backbones. The configurations of successive base-pairs are specified by six independent ''pair'' parameters: Propeller Twist, Buckle, Opening rotations and Shear, Stretch, Stagger displacements; and six ''step'' variables, Twist, Tilt, Roll angles and Shift, Slide, Rise translations (see Dickerson *et al*., 1989) for qualitative definitions). Coordinates of backbone atoms are found using a standard chain-closure algorithm (Zhurkin *et al*., 1978), which requires two additional variables per nucleoside unit (the pseudorotational phase angle P and the glycosyl torsion angle χ) and an initial estimate of backbone geometry (e.g. *trans versus gauche*$^+$ forms of the exocyclic O5′—C5′—C4′—C3′ backbone torsion γ). In addition to energy minimization, this parameter set has proved to be efficient in Monte Carlo simulations of DNA helices based on the `DNAminiCarlo` software package (Zhurkin *et al*., 1991).

Regular homopolymers composed of identical base-pair units are used to reduce the number of independent variables. Residues in complementary strands are further assumed to adopt the same sugar puckering and glycosyl angles. These simplifications make it possible to carry out a detailed search of conformation space, and also provide an easy way to control the length of the molecule. Indeed, for a regular DNA molecule it is possible to calculate a global axis and to determine all helical parameters with respect to this axis (see below). The magnitude of the helical rise serves as the natural measure of chain stretching or compression.

Helical parameters

The helical parameters used as independent degrees of freedom in the calculations are expressed in terms of the global helical axis relating neighboring base-

pairs. While finding such an axis is non-trivial in the case of a deformed (asymmetric) structure (Lavery & Sklenar, 1990), identification is straightforward for a regularly repeating polymer (Miyazawa, 1961; Rosenberg *et al.*, 1976). Here we present formulas relating the "*helical parameters*" of a DNA homopolymer, helical twist (Ω_h), helical shift or x-displacement (D_x), helical slide or y-displacement (D_y), and step height (h), to the "*local parameters*" describing the geometry of successive base-pairs: Twist (Ω), Tilt (τ), Roll (ρ), Shift (Δx), Slide (Δy), and Rise (Δz).

We start by embedding coordinate frames B_i and B_{i+1} in the planes of neighboring base-pairs and defining S as the global helical axis (Figure 11). We introduce two additional coordinate frames, A_i and A_{i+1}, positioned at their origins along S such that transformations between pairs of coordinate axes are defined by:

$$\begin{aligned}
\mathbf{r}_{A_i} &= \mathbf{M}\mathbf{r}_{B_i} + \begin{bmatrix} D_x \\ D_y \\ 0 \end{bmatrix}, \\
\mathbf{r}_{A_{i+1}} &= \mathbf{M}\mathbf{r}_{B_{i+1}} + \begin{bmatrix} D_x \\ D_y \\ 0 \end{bmatrix}, \\
\mathbf{r}_{A_i} &= \mathbf{Z}(\Omega_h)\mathbf{r}_{A_{i+1}} + \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix}, \\
\mathbf{r}_{B_i} &= \mathbf{T}(\Omega, \tau, \rho)\mathbf{r}_{B_{i+1}} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix},
\end{aligned} \tag{1}$$

Here the $\mathbf{r}_X$ are vectors in coordinate system X, $\mathbf{M}$ is the transformation matrix that orients base-pairs in the global coordinate frame, $\mathbf{Z}(\Omega_h)$ is the matrix defining rotation about the helical axis by angle Ω_h, $\mathbf{T}(\Omega,\tau,\rho)$ is the matrix that relates neighboring base-pairs, (D_x, D_y, h) are the components of helical displacement, and (Δx, Δy, Δz) are the components of local dimer displacement.

Rearrangement of the preceding expressions leads to the following relationships between local and helical parameters:

$$\mathbf{T}(\Omega, \tau, \rho) = \mathbf{M}^{-1}\mathbf{Z}(\Omega_h)\mathbf{M}, \tag{2}$$

$$\mathbf{M}\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \mathbf{Z}(\Omega_h)\begin{bmatrix} D_x \\ D_y \\ 0 \end{bmatrix} - \begin{bmatrix} D_x \\ D_y \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix} \tag{3}$$

Local parameters are obtained from global parameters in two steps. First, the transformation matrix $\mathbf{T}$ relating neighboring base-pairs is constructed from $\mathbf{M}$ and $\mathbf{Z}$ using the given helical inclination η, tip θ, and twist Ω_h in equation (2). Here we take $\mathbf{M} = \mathbf{X}(\theta)\mathbf{Y}(\eta)$, where $\mathbf{X}$ and $\mathbf{Y}$ designate matrices for rotations by the designated angles about the local x and y-coordinate axes. Thus, matrix $\mathbf{T}$ is calculated as $\mathbf{T}(\Omega,\tau,\rho) = \mathbf{Y}(-\eta)\mathbf{X}(-\theta)\ \mathbf{Z}(\Omega_h)\ \mathbf{X}(\theta)\mathbf{Y}(\eta)$. Symmetric base-pair step parameters can then be extracted from selected elements of $\mathbf{T}$. For example, with $\mathbf{T}$ expressed as the symmetrized matrix product $\mathbf{Z}(\Omega/2 + \alpha)\mathbf{Y}(\beta)\mathbf{Z}(\Omega/2 - \alpha)$ (Zhurkin *et al.*, 1979), Twist (Ω), Tilt (τ), and Roll (ρ) are given by:

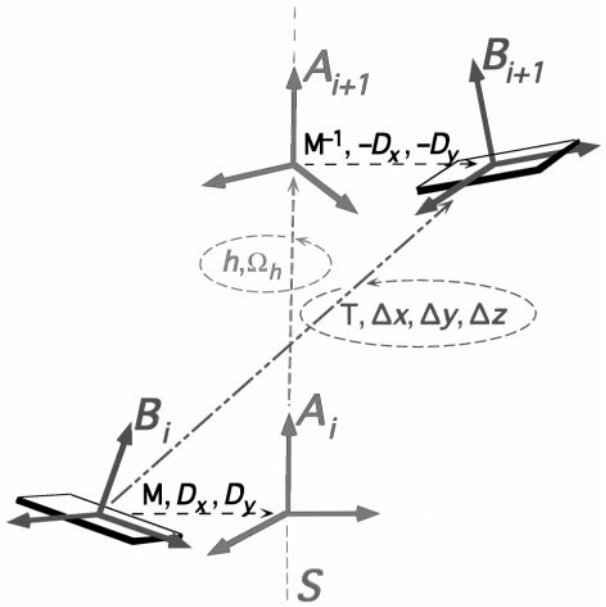

Figure 11. Schematic of coordinate frames and parameters used to relate local and helical parameters of dimer steps. Frames A_i and A_{i+1} lie along the helical axis S, while frames B_i and B_{i+1} are embedded in the planes of neighboring base-pairs.

$$\tan\Omega = \frac{T_{21} - T_{12}}{T_{11} + T_{22}} \tag{4a}$$

$$\tau = \beta \sin\alpha \tag{4b}$$

$$\rho = \beta \cos\alpha \tag{4c}$$

Here β is the net bending angle, and α the phase angle of bending, defined by:

$$\cos\beta = T_{33} \tag{5a}$$

$$\tan\alpha = \frac{T_{32} - T_{23}}{T_{31} - T_{13}} \tag{5b}$$

Other definitions of $\mathbf{T}$, e.g. (Babcock *et al.*, 1994; El Hassan & Calladine, 1995), yield related expressions for symmetric base-pair step parameters. Local parameters used here follow the defintions of the `CompDNA` software package (Gorin *et al.*, 1995).

Rearrangement of equation (3) similarly yields local translational parameters in terms of helical rise h, and the D_x,D_y components of helical shift (x and y displacements):

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \mathbf{M}^{-1}\begin{bmatrix} (\cos\Omega_h - 1) & -\sin\Omega_h & 0 \\ \sin\Omega_h & (\cos\Omega_h - 1) & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} D_x \\ D_y \\ h \end{bmatrix}. \tag{6}$$

Alternatively, helical parameters, Ω_h, θ, and η, can be determined from local parameters. The angular com-

ponents are extracted from the elements of **T** constructed from a given set of local dimer rotations (Ω, τ, ρ):

$$\cos\Omega_h = \frac{1}{2}(T_{11} + T_{22} + T_{33} - 1) \tag{7a}$$

$$\tan\theta = \frac{T_{23} - T_{32}}{T_{12} - T_{21}} \tag{7b}$$

$$2\sin\eta\sin\Omega_h = T_{13} - T_{31} \tag{7c}$$

Substitution of these values and the local dimer displacement (Δx, Δy, Δz) in equation (5) then yields the components of helical displacement:

$$D_x = \frac{P_x(\cos\Omega_h - 1) + P_y\sin\Omega_h}{2(1 - \cos\Omega_h)} \tag{8a}$$

$$D_y = \frac{P_y(\cos\Omega_h - 1) - P_x\sin\Omega_h}{2(1 - \cos\Omega_h)} \tag{8b}$$

$$h = Pz \tag{8c}$$

where:

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \mathbf{M}\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \tag{9}$$

Energies

Potential energies are calculated using standard all-atom potential functions (Zhurkin *et al.*, 1981; Poltev & Shulyupina, 1986). Notably, these potentials reproduce the observed enthalpies of base stacking (Zhurkin *et al.*, 1981) and planar base-base interactions (Poltev & Shulyupina, 1986). A sigmoidal, distance-dependent dielectric constant of the form (Lavery, 1988; Ramstein & Lavery, 1988),

$$\varepsilon(r_{ij}) = \varepsilon_\infty - (\varepsilon_\infty - 1)\left(\frac{1}{2}(sr_{ij})^2 + sr_{ij} + 1\right)\exp(-sr_{ij}) \tag{10}$$

is chosen to account for solvent effects. We use two extreme versions of this function: a steeper version (defined by the controlling parameter $s = 0.356$) combined with neutralized phosphate groups, which we term ''high salt'' conditions, because of the masking of most electrostatic interactions, and a version with lower slope ($s = 0.16$) combined with full (-1) charges on phosphate groups, which we call ''low salt''. We set ε_∞ to 78.3, the dielectric constant of water, so that ε ranges from 8.2 to 21.4 at high salt and 2.0 to 4.7 at low salt as r_{ij}, the non-bonded distance between atoms i and j, increases from 3 to 5 Å. We thus can model a broad range of ionic conditions as well as see the influence of short and long-range electrostatic interactions on DNA conformation.

Rather than simulating the interconversion between *A* and *B* helical forms, we restrict the furanose rings to puckered states in either the C2′-*endo* or C3′-*endo* conformational range. This is accomplished by replacing the standard (i.e. fixed amplitude, $\tau_m = 38^\circ$) pseudorotational energy profile (Olson, 1982a; Zhurkin *et al.*, 1982) by a harmonic expression fit to the computed energy minima
at 298 K. Specifically, the energy is increased by $RT/2$ when the phase angle P deviates from the minimum energy rest state, $P_o = 171^\circ$ for C2′-*endo* and 9° for C3′-*endo* forms, by the root-mean-square fluctuation, $\langle\Delta P^2\rangle^{1/2}$. We consider conformational asymmetry by introducing force constants that reflect positive *versus* negative pseudorotational changes, i.e. $\langle\Delta P_+^2\rangle^{1/2} = 12^\circ$ *versus* $\langle\Delta P_-^2\rangle^{1/2} = 14^\circ$ for C2′-*endo* and $\langle\Delta P_+^2\rangle^{1/2} = 14^\circ$ *versus* $\langle\Delta P_-^2\rangle^{1/2} = 15^\circ$ for C3′-*endo* states. We thus force all sugars to be either *B* or *A*-like and to mimic the preferred fluctuations in *P*.

Minimization

The minimization procedure uses the method of Rosenbrock (1960) with a linear search, modified according to the ''numerical recipes'' provided by Fletcher & Reeves (1964). Recent analyses of the mutual correlations of base-pair step parameters in DNA crystal structures (Gorin *et al.*, 1995; Olson *et al.*, 1998) suggest ways to perform the conformational search more efficiently. Thus, for a given polymer stretch (i.e. fixed helical rise), minimization in the (inclination, helical twist) plane, two of the most deformable helical variables, usually leads to an energetically tolerable structure. Stable minima are also obtained with the corresponding local parameters, Roll and Twist, used as independent variables. Indeed, we have found that we can carry out minimizations interchangeably in local and helical parameter space and identify the same low energy states. This further increases the effectiveness of the energy optimization procedure. Here, in addition, independent parameters known to span a limited range in DNA crystal structures, Tilt (or helical tip), local Shift (or helical *y*-displacement), Buckle, and Opening, are set to zero. Each of the optimized structures is represented by a point on the profiles of energy *versus* helical rise in Figure 1 and the plots of local and helical parameters in Figures 2-5. These plots are obtained from calculations at increments in helical rise of 0.05 Å or less.

Acknowledgements

We are grateful to Drs Andrzej Stasiak, Mikhail A. Livshits, and Suse Broyde for useful discussions and to Drs Takehiko Shibata and Taro Nishinaka for providing atomic coordinates of their model of single-stranded DNA bound to RecA prior to public release. Support of this work through USPHS grant GM20861 and the National Cancer Institute, under contract N10-CO-74102, is gratefully acknowledged. K.M.K. also acknowledges fellowship support from the Program in Mathematics and Molecular Biology based at Florida State University. Computations were carried out at the Rutgers University Center for Computational Chemistry and through the facilities of the Nucleic Acid Database project (NSF grant DBI 9510703). This work was taken in part from the dissertation of K.M.K. written in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Rutgers University, 1998.

References

Arnott, S. (1970). The geometry of nucleic acids. *Prog. Biophys. Mol. Biol.* **21**, 265-319.

Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C. & Watts, L. (1974). Structural details observed for DNAs containing alternating purine and pyrimidine sequences. *J. Mol. Biol.* **88**, 523-533.

Auble, D. T. & deHaseth, P. L. (1988). Promoter recognition by *Escherichia coli* RNA polymerase. Influence of DNA structure in the spacer separating the −10 and −35 regions. *J. Mol. Biol.* **202**, 471-482.

Babcock, M. S., Pednault, E. P. D. & Olson, W. K. (1994). Nucleic acid structure analysis. Mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* **237**, 125-156.

Baliga, R., Singleton, J. W. & Dervan, P. B. (1995). RecA·oligonucleotide filaments bind in the minor groove of double-stranded DNA. *Proc. Natl Acad. Sci. USA,* **92**, 10393-10397.

Barber, A. M., Zhurkin, V. B. & Adhya, S. (1993). CRP-binding sites: evidence for two structural classes with 6 bp and 8 bp spacers. *Gene,* **130**, 1-8.

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

Bensimon, D. (1996). Force: a new structural control parameter? *Structure,* **4**, 885-889.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751-759.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Calladine, C. R. & Drew, H. R. (1984). A base-centered explanation of the *B*-to-*A* transition in DNA. *J. Mol. Biol.* **178**, 773-782.

Chandrasekaran, R., Arnott, S., Banerjee, A., Campbell-Smith, S., Leslie, A. G. W. & Puigjaner, L. (1980). Some new polynucleotide structures and some new thoughts about old structures. In *Fiber Diffraction Methods. ACS Symposium Series 141* (French, A. D. & Gardner, K. H., eds), pp. 483-502, American Chemical Society, Washington, DC.

Cheatham, T. E., III & Kollman, P. A. (1996). Observation of the *A*-DNA to *B*-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* **259**, 434-444.

Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J.-L., Chatenay, D. & Caron, F. (1996). DNA: an extensible molecule. *Science,* **271**, 792-794.

Crick, F. H. C. & Watson, J. D. (1954). The complementary structure of deoxyribonucleic acid. *Proc. Roy. Soc. ser. A,* **223**, 80-96.

Dagneaux, C., Porumb, H., Liquier, J., Takahashi, M. & Taillandier, E. (1995). Conformations of three-stranded DNA structures formed in presence and in absence of the RecA protein. *J. Biomol. Struct. Dynam.* **13**, 465-470.

Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtures of normality. *Nucl. Acids Res.* **26**, 1906-1926.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M. & Tung, C.-S., *et al.* (1989). Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* **208**, 787-791.

Ding, J., Das, K., Hsiou, Y., Sarafianos, S. G., Clark, A. D., Jr, Jacobo-Molina, A., Tantillo, C., Hughes, S. H. & Arnold, E. (1998). Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **284**, 1095-1111.

Edmondson, S. P. & Johnson, W. C., Jr (1986). Base tilt of *B*-form poly[d(G)]-poly[d(C)] and the *B*- and *Z*-conformations of poly[d(GC)]-poly[d(GC)] in solution. *Biopolymers,* **25**, 2335-2348.

Egelman, E. & Stasiak, A. (1986). Structure of helical RecA-DNA complexes. Complexes formed in the presence of ATP-γ-*S* or ATP. *J. Mol. Biol.* **191**, 677-697.

El Hassan, M. A. & Calladine, C. R. (1995). The assessment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme with an appendix. *J. Mol. Biol.* **251**, 648-664.

Fernandez, L. G., Subirana, J. A., Verdaguer, N., Pyshnyi, D., Campos, L. & Malinina, L. (1997). Structural variability of *A*-DNA in crystals of d(pCpCpCpGpCpGpG). *J. Biomol. Struct. Dynam.* **15**, 151-163.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The `MIDAS` display system. *J. Mol. Graphics,* **6**, 13-27.

Fletcher, R. & Reeves, C. M. (1964). Function minimization by conjugate gradients. *Comput. J.* **7**, 149-154.

Franklin, R. E. & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature,* **171**, 740-741.

Gelbin, A., Schneider, B., Clowney, L., Olson, W. K. & Berman, H. M. (1996). Geometric parameters in nucleic acids: sugar and phosphate constituents. *J. Am. Chem. Soc.* **118**, 530-540.

Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995). *B*-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**, 34-48.

Grzeskowiak, K., Yanagi, K., Privé, G. G. & Dickerson, R. E. (1991). The structure of *B*-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. *J. Biol. Chem.* **266**, 8861-8883.

Guzikevich-Guerstein, G. & Shakked, Z. (1995). A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nature Struct. Biol.* **3**, 32-37.

Howard-Flanders, P., West, S. C. & Stasiak, A. (1984). Role of RecA protein spiral filaments in genetic recombination. *Nature,* **309**, 215-219.

Howard-Flanders, P., West, S. C., Cassuto, E., Hahn, T.-R., Egelman, E. & Stasiak, A. (1987). Structure of RecA spiral filaments and their role in homologous pairing and strand exchange in genetic recombination. In *DNA Replication and Recombination* (Kelly, T. & McMacken, R., eds), pp. 609-617, Alan R. Liss, Inc., New York.

Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science,* **282**, 1669-1675.

Ivanov, V. I., Minchenkova, L. E., Schyolkina, A. K. & Poletayev, A. I. (1973). Different conformations of double-stranded nucleic acid in solution as revealed by circular dichroism. *Biopolymers,* **12**, 89-110.

Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D. & Schyolkina, A. K. (1974). The *B* to *A* transition of DNA in solution. *J. Mol. Biol.* **87**, 817-833.

Ivanov, V. I., Minchenkova, L. E., Minyat, E. E. & Schyolkina, A. K. (1983). Cooperative transitions in DNA with no separation of strands. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 243-250.

Ivanov, V. I., Minchenkova, L. E., Chernov, B. K., McPhie, P., Ryu, S., Garges, S., Barber, A. M., Zhurkin, V. B. & Adhya, S. (1995). CRP-DNA complexes: inducing the *A*-like form in the binding sites with an extended central spacer. *J. Mol. Biol.* **245**, 228-240.

Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. & Dickerson, R. E. (1996). How proteins recognize the TATA-box. *J. Mol. Biol.* **261**, 239-254.

Kim, J. L. & Burley, S. K. (1994). 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Struct. Biol.* **1**, 638-653.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature,* **365**, 520-527.

Kim, T.-K., LaGrange, T., Wang, Y.-H., Griffith, J. D., Reinberg, D. & Ebright, R. H. (1997). Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. *Proc. Natl Acad. Sci. USA,* **94**, 12268-12273.

Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature,* **365**, 512-520.

Konrad, M. W. & Bolonick, J. I. (1996). Molecular dynamics simulation of DNA stretching is consistent with the tension observed for extension and strand separation and predicts a novel ladder structure. *J. Am. Chem. Soc.* **118**, 10989-10994.

Kowalczykowski, S. C. & Eggleston, A. K. (1994). Homologous pairing and DNA strand-exchange proteins. *Annu. Rev. Biochem.* **63**, 991-1043.

Lacks, S. (1966). Integration efficiency and genetic recombination in pneumococcal transformation. *Genetics,* **53**, 207-235.

Lavery, R. (1988). Junctions and bends in nucleic acids: a new theoretical modeling approach. In *Structure & Expression: DNA Bending and Curvature* (Olson, W. K., Sarma, M. H., Sarma, R. H. & Sundaralingam, M., eds), vol. 3, pp. 191-211, Adenine Press, Guilderland, NY.

Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynam.* **6**, 655-667.

Lavery, R. & Sklenar, H. (1990). A quantitative description of the conformation of biological molecules. In *Structure and Methods, DNA Protein Complexes and Proteins* (Sarma, R. H. & Sarma, M. H., eds), vol. 2, pp. 215-235, Adenine Press, Schenectady, NY.

Lavery, R., Zakrzewska, K. & Sklenar, H. (1995). JUMNA (junction minimisation of nucleic acids). *Comput. Phys. Commun.* **91**, 135-158.

Lebrun, A. & Lavery, R. (1996). Modelling extreme stretching of DNA. *Nucl. Acids Res.* **24**, 2260-2267.

Lebrun, A., Shakked, Z. & Lavery, R. (1997). Local DNA stretching mimics the distortion caused by the TATA box-binding protein. *Proc. Natl Acad. Sci. USA,* **94**, 2993-2998.

Leslie, A. G. W. & Arnott, S. (1978). Structure of the single-stranded polyribonucleotide poly(2′-O-methylcytidylic acid). *J. Mol. Biol.* **119**, 399-414.

Lu, X. & Olson, W. K. (1999). Resolving the discrepancies among nucleic acid conformational analyses. *J. Mol. Biol.* **285**, 1563-1575.

Lu, X. J., El Hassan, M. A. & Hunter, C. A. (1997). Structure and conformation of helical nucleic acids: analysis program (SCHNAaP). *J. Mol. Biol.* **273**, 668-680.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature,* **389**, 251-260.

Malkov, V. A., Panyutin, I. G., Neumann, R. D., Zhurkin, V. B. & Camerini-Otero, R. D. (1999). The arrangement of the three DNA strands within the RecA protein mediated synaptic complexes. *Biophys. J.* **76**, A130.

Mandel-Gutfreund, Y., Margalit, H., Jernigan, R. L. & Zhurkin, V. B. (1998). A role for CH···O interactions in protein-DNA recognition. *J. Mol. Biol.* **277**, 1129-1140.

Manning, G. (1978). The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quart. Rev. Biophys.* **11**, 179-246.

McEntee, K., Weinstock, G. M. & Lehman, I. R. (1981). Binding of the RecA protein of *Escherichia coli* to single- and double-stranded DNA. *J. Biol. Chem.* **256**, 8835-8844.

Miyazawa, T. (1961). Molecular vibrations and structure of high polymers. II. Helical parameters of infinite polymer chains as functions of bond lengths, bond angles, and internal rotation angles. *J. Polymer Sci.* **55**, 215-231.

Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Disatake, K., Lee, D. K., Roeder, R. G. & Burley, S. K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature,* **377**, 119-128.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffmann, A., Roeder, R. G. & Burley, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl Acad. Sci. USA,* **93**, 4862-4867.

Nishinaka, T., Ito, Y., Yokoyama, S. & Shibata, T. (1997). An extended DNA structure through deoxyribose-base stacking induced by RecA protein. *Proc. Natl Acad. Sci. USA,* **94**, 6623-6628.

Nishinaka, T., Shinohara, A., Ito, Y., Yokoyama, S. & Shibata, T. (1998). Base-pair switching by interconversion of sugar puckers in DNA extended by proteins of RecA-family: a model for homology search in homologous genetic recombination. *Proc. Natl Acad. Sci. USA,* **95**, 11071-11076.

Nordén, B., Seth, S. & Tjerneld, F. (1978). Renaturation of DNA in ethanol-methanol solvent induced by complexation with methyl green. *Biopolymers,* **17**, 523-525.

Nordén, B., Elvingson, C., Kubista, M., Sjoberg, B., Ryberg, H., Ryberg, M., Mortensen, K. & Takahashi, M. (1992). Structure of RecA-DNA complexes studied by combination of linear dichroism and small-angle neutron scattering measurements on flow-oriented samples. *J. Mol. Biol.* **226**, 1175-1191.

Nordén, B., Wittung-Stafshede, P., Ellouze, C., Kim, H.-K., Mortensen, K. & Takahashi, M. (1998). Base orientation of second DNA in RecA·DNA filaments.

Analysis by combination of linear dichroism and small angle neutron scattering in flow-oriented solution. *J. Biol. Chem.* **273**, 15682-15686.

Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. (1993). Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science,* **259**, 1896-1899.

Olson, W. K. (1976). The spatial configuration of ordered polynucleotide chains. I. Helix formation and stacking. *Biopolymers,* **15**, 859-878.

Olson, W. K. (1977). Spatial configuration of ordered polynucleotide chains: a novel double helix. *Proc. Natl Acad. Sci. USA,* **74**, 1775-1779.

Olson, W. K. (1981). Understanding the motions of DNA. In *Biomolecular Stereodynamics* (Sarma, R. H., ed.), vol. 1, pp. 327-343, GuilderlandAdenine Press, New York.

Olson, W. K. (1982a). How flexible is the furanose ring? 2. An updated potential energy estimate. *J. Am. Chem. Soc.* **104**, 278-286.

Olson, W. K. (1982b). Theoretical studies of nucleic acid conformation: potential energies, chain statistics, and model building. In *Topics in Nucleic Acid Structures* (Neidle, S., ed.), part 2 edit., pp. 1-79, Macmillan Press, London.

Olson, W. K. & Dasika, R. D. (1976). The spatial configuration of ordered polynucleotide chains. III. Polycyclonucleotides. *J. Am. Chem. Soc.* **98**, 5371-5380.

Olson, W. K., Srinivasan, A. R., Marky, N. L. & Balaji, V. N. (1983). Theoretical probes of DNA conformation: examining the *B*-to-*Z* conformational transition. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 229-241.

Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M. & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA,* **95**, 11163-11168.

Olson, W. K., Kosikov, K. M., Colasanti, A., Gorin, A. A. & Zhurkin, V. B. (1999). Pulling and pushing the DNA double helix. In *Proceedings of the Third Symposium on Biological Physics* (Frauenfelder, H., ed.), American Institute of Physics , Woodbury, NY **In the press**.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III, DeBolt, S., Ferguson, D., Seibel, G. & Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1-41.

Podyminogin, M. A., Meyer, R. B. & Gamper, H. B. (1995). Sequence-specific covalent modification of DNA by cross-linking oligonucleotides. Catalysis by RecA and implication for the mechanism of synaptic joint formation. *Biochemistry,* **34**, 13098-13108.

Poltev, V. I. & Shulyupina, N. V. (1986). Simulation of interactions between nucleic acid bases by refined atom-atom potential functions. *J. Biomol. Struct. Dynam.* **3**, 739-765.

Ramstein, J. & Lavery, R. (1988). Energetic coupling between DNA bending and base-pair opening. *Proc. Natl Acad. Sci. USA,* **85**, 7231-7235.

Rich, A. (1959). An analysis of the relation between DNA and RNA. *Ann. NY Acad. Sci.* **81**, 709-722.

Rief, M., Clausen-Schaumann, H. & Gaub, H. E. (1999). Sequence-dependent mechanics of single DNA molecules. *Nature Struct. Biol.* **6**, 346-349.

Robert, F., Douzlech, M., Forget, D., Egly, J.-M., Greenblatt, J., Burton, Z. F. & Coulombe, B. (1998). Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol. Cell,* **2**, 341-351.

Roberts, R. J. & Cheng, X. (1998). Base flipping. *Annu. Rev. Biochem.* **67**, 181-198.

Rosenberg, J. M., Seeman, N. C., Day, R. O. & Rich, A. (1976). RNA double helices generated from crystal structures of double helical dinucleoside phosphates. *Biochem. Biophys. Res. Commun.* **69**, 979.

Rosenbrock, H. H. (1960). An automated method for finding the greatest or least value of a function. *Comput. J.* **3**, 175-184.

Satchwell, S. C., Drew, H. R. & Travers, A. A. (1986). Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659-675.

Schneider, B., Neidle, S. & Berman, H. M. (1997). Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers,* **42**, 113-124.

Selsing, E., Arnott, S. & Ratliff, R. L. (1975). Conformations of poly(d(A-T-T))-poly(d(A-A-T)). *J. Mol. Biol.* **98**, 243-248.

Smith, S. B., Cui, Y. & Bustamante, C. (1996). Overstretching *B*-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science,* **271**, 795-799.

Sprous, D., Young, M. A. & Beveridge, D. L. (1998). Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: theoretical considerations of the $A \leftrightarrow B$ transition. *J. Phys. Chem. B,* **102**, 4658-4667.

Stasiak, A. & Di Capua, E. (1982). The helicity of DNA in complexes with RecA protein. *Nature,* **299**, 185-186.

Stasiak, A. & Egelman, E. (1988). Structure of helical RecA-DNA complexes. III. The structural polarity of RecA filaments and functional polarity in the RecA-mediated strand-exchange reaction. *J. Mol. Biol.* **202**, 659-662.

Stasiak, A., Di Capua, E. & Koller, T. (1981). Elongation of duplex DNA by RecA protein. *J. Mol. Biol.* **151**, 557-564.

Steitz, T. A. (1990). Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Quart. Rev. Biophys.* **23**, 205-280.

Story, R. M., Weber, I. T. & Steitz, T. A. (1992). The structure of the *E. coli* RecA protein monomer and polymer. *Nature,* **355**, 318-325.

Suzuki, M. & Yagi, N. (1995). Stereochemical basis of DNA bending by transcription factors. *Nucl. Acids Res.* **23**, 2083-2091.

Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature,* **381**, 127-134.

Travers, A. A. (1989). DNA conformation and protein binding. *Annu. Rev. Biochem.* **58**, 427-452.

von Hippel, P. H. (1998). An integrated model of the transcription complex in elongation, termination, and editing. *Science,* **281**, 660-665.

Wahl, M. C., Rao, S. T. & Sundaralingham, M. (1996). The structure of r(UUC(GCG)) has 5′-UU overhang exhibiting Hoogsten-like trans U·U base-pairs. *Nature Struct. Biol.* **3**, 24-31.

Watson, J. D. & Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. *Nature,* **171**, 737-738.

Werner, M. H., Gronenborn, A. M. & Clore, G. M. (1996). Intercalation, DNA kinking, and the control of transcription. *Science,* **271**, 778-784.

West, S. C. (1992). Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.* **61**, 603-640.

Wilkins, M. H. F., Gosling, R. G. & Seeds, W. E. (1951). Nucleic acid: an extensible molecule?. *Nature,* **167**, 759-760.

Yang, L. & Pettitt, B. M. (1996). *B* to *A* transition of DNA on the nanosecond time scale. *J. Phys. Chem.* **100**, 2564-2566.

Yathindra, N. & Sundaralingam, M. (1976). Analysis of the possible helical structures of nucleic acids and polynucleotides. Application of (*n-h*) plots. *Nucl. Acids Res.* **3**, 729-747.

Young, M. A., Jayaram, B. & Beveridge, D. L. (1997). Intrusion of counterions into the spine of hydration in the minor groove of *B*-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.* **119**, 59-69.

Zhou, X. & Adzuma, K. (1997). DNA strand exchange mediated by the *Escherichia coli* RecA protein initiates in the minor groove of double-stranded DNA. *Biochemistry,* **36**, 4650-4661.

Zhurkin, V. B., Lysov, Y. P. & Ivanov, V. I. (1978). Different families of double-stranded conformations of DNA revealed by computer calculations. *Biopolymers,* **17**, 377-412.

Zhurkin, V. B., Lysov, Y. P. & Ivanov, V. I. (1979). Anisotropic flexibility of DNA and the nucleosomal structure. *Nucl. Acids Res.* **6**, 1081-1096.

Zhurkin, V. B., Poltev, V. I. & Florent'ev, V. L. (1981). Atom-atom potential functions for conformational calculations of nucleic acid. *Mol. Biol. USSR (English Edit.),* **14**, 882-895.

Zhurkin, V. B., Lysov, Y. P., Florentiev, V. L. & Ivanov, V. I. (1982). Torsional flexibility of *B*-DNA as revealed by conformational analysis. *Nucl. Acids Res.* **10**, 1811-1830.

Zhurkin, V. B., Ulyanov, N. B., Gorin, A. A. & Jernigan, R. L. (1991). Static and statistical bending of DNA evaluated by Monte Carlo simulations. *Proc. Natl Acad. Sci., USA,* **88**, 7046-7050.

Zhurkin, V. B., Raghunathan, G., Ulyanov, N. B., Camerini-Otero, R. D. & Jernigan, R. L. (1994a). A parallel DNA triplex as a model for the intermediate in homologous recombination. *J. Mol. Biol.* **239**, 181-200.

Zhurkin, V. B., Raghunathan, G., Ulyanov, N. B., Camerini-Otero, R. D. & Jernigan, R. L. (1994b). Recombination triple helix, *R*-form DNA. A stereochemical model for recognition and strand exchange. In *Structural Biology: The State of the Art* (Sarma, R. H. & Sarma, M. H., eds), vol. 2, pp. 43-66, Adenine Press, Schenectady, NY.

Note added in proof

Two studies published after this paper was accepted for publication are consistent with the mechanism of information readout suggested here. First, model building based on the crystal structure of T7 RNA polymerase with promoter DNA (Cheetham, G. M. T., Jeruzalmi, D. & Steitz, T. A. (1999). *Nature* **399**, 80-83.) implies that the DNA template strand is sharply kinked at the active center of polymerase. Second, analysis of the Mn(II) binding to ribonucleotides (Garland, C. S., Tarien, E., Nirmala, R., Clark, P., Rifkind, J. & Eichhorn, G. L. (1999). *Biochemistry* **38**, 3421-3425.) suggests that both the RNA and DNA strands are noticeably kinked at the elongation site of *E. coli* RNA polymerase. Importantly, this kink increases the distance between adjacent bases along the DNA chain and thereby diminishes the probability of misincorporation in the RNA chain.

Edited by I. Tinoco

(Received 27 October 1998; received in revised form 15 April 1999; accepted 15 April 1999)