

2001

A Crystallographic Map of the Transition from B-DNA to A-DNA

Jeff Vargason

George Fox University, jvargason@georgefox.edu

Keith Henderson

Oregon State University

P. Shing Ho

Oregon State University and Lawrence Berkeley National Laboratory

Follow this and additional works at: http://digitalcommons.georgefox.edu/bio_fac

 Part of the [Biochemistry Commons](#), and the [Biology Commons](#)

Recommended Citation

Previously published in Proceedings of the National Academy of Sciences of the United States of America, 2001, volume 98, pp. 7265-7270 www.pnas.org

This Article is brought to you for free and open access by the Department of Biology and Chemistry at Digital Commons @ George Fox University. It has been accepted for inclusion in Faculty Publications - Department of Biology and Chemistry by an authorized administrator of Digital Commons @ George Fox University.

A crystallographic map of the transition from B-DNA to A-DNA

Jeffrey M. Vargason*, Keith Henderson†, and P. Shing Ho**

*Department of Biochemistry and Biophysics, ALS 2011, Oregon State University, Corvallis, OR 97331; and †Macromolecular Crystallography Facility at the Advanced Light Source, Physical Biosciences Division, MS 6-2100, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Communicated by K. E. van Holde, Oregon State University, Corvallis, OR,

The transition between B- and A-DNA was first observed nearly 50 years ago. We have now mapped this transformation through a set of single-crystal structures of the sequence d(GGCGCC)₂, with various intermediates being trapped by methylating or brominating the cytosine bases. The resulting pathway progresses through 13 conformational steps, with a composite structure that pairs A-nucleotides with complementary B-nucleotides serving as a distinct transition intermediate. The details of each step in the conversion of B- to A-DNA are thus revealed at the atomic level, placing intermediates for this and other sequences in the context of a common pathway.

The transformation from B-DNA in fibers (1) to dehydrated A-DNA (2) was one of the first reversible structural transitions observed in a biomolecule. Although B-DNA is recognized as the standard form of the double helix, the junction between B- and A-DNA appears as a target for the anticancer drug cisplatin (3), and an A-conformation with exaggerated base-pair inclinations is induced by TATA-binding protein, the transcriptional promoter in eukaryotes (4, 5). Consequently, an understanding of how B-DNA converts to A-DNA (the B-A transition) is significant both for its history and for understanding how DNA structure affects cellular function. In the current study, methylated and brominated variants of the sequence d(GGCGCC)₂ were crystallized as standard A-DNA and several intermediate conformations, the most important being a composite structure in which each strand is half A-DNA and half B-DNA, with the A-nucleotides paired with complementary B-nucleotides across the duplex. Thus, along with structures from previous work (6), we have assembled 13 single-crystal structures that define a pathway for the B-A transition in this single sequence.

In a B-A transition, the long and narrow B-duplex—with its 10- to 10.5-bp repeat, 3.4-Å rise, and base pairs stacked at the center of the helix—is converted to the underwound and compact A-DNA structure, characterized by an 11-bp repeat, a 2.6-Å rise, and base pairs that are inclined up to 20° and displaced by ≈4 Å so they essentially wrap around the helix axis (7). The geometries of the deoxyribose sugars are also converted from C2'-endo in B-DNA (with the C2'-carbon puckered above the furanose plane toward the nucleobase) to C3'-endo in A-DNA. Molecular mechanics simulations suggest that many of these structural features are transformed in a concerted manner (8–10), leading to the question of whether a discrete intermediate exists in the B-A transition.

There is a growing number of DNA crystal structures that exhibit properties of both A- and B-DNA. These include d(CCCCGGG)₂, an A-DNA with a relatively extended rise of 3.1 Å (11), and d(CGCCCCGGGCG)₂, a kinked A-DNA structure with B-type helices at the ends (12). More recently, the central six base pairs of d(CATGGGCCATG)₂ were defined as an A/B-intermediate conformation in which the base pairs are displaced but not highly inclined relative to the helix axis (13). Our own studies have shown that methylating or brominating cytosines in d(GGCGCC)₂ induce a unique extended conformation that is neither B- nor A-DNA (6). The sequence d(GGCGCC)₂ crystallized as standard B-DNA, whereas the

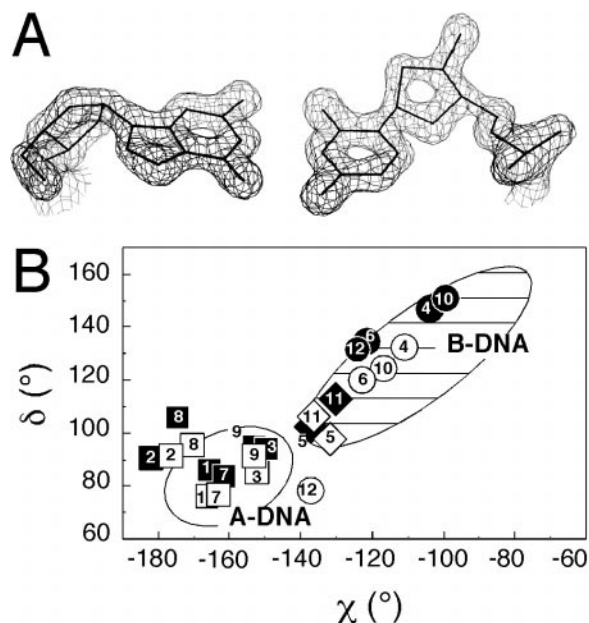


Fig. 1. Single-crystal structures of d(GGBr⁵CGCC)₂ and d(GGm⁵CGCC)₂. (A) Experimental electron density map derived from the 1.6-Å multiwavelength anomalous diffraction phased x-ray diffraction data of d(GGBr⁵CGCC)₂. The electron density, after density modification, of one base pair is shown with the final refined model of this GC base pair included for reference [figure created with BOBSCRIPT (21)]. (B) Comparison of the δ - and χ -torsion angles and the sugar conformations of the nucleotides in d(GGBr⁵CGCC)₂ (open symbols) and d(GGm⁵CGCC)₂ (closed symbols). Nucleotides with C2'-endo type sugars are indicated by circles, with the C3'-endo sugars by squares, and those with intermediate O4'-endo sugars by diamonds. Nucleotides in the two structures are numbered 1–6 for one strand and 7–12 for the complementary strand (both in the 5' to 3' directions). The values of δ and χ for previous single-crystal structures [comprising high resolution—better than 2 Å—crystal structures deposited in the Nucleic Acid Database (22)] of B-DNA are defined by the hashed oval, and those of A-DNA are defined by the open oval [adapted from Lu et al. (23)].

structures of d(GGCGm⁵CC)₂ and d(GGCGBr⁵CC)₂ (m⁵C and Br⁵C are 5-methylcytosine and 5-bromocytosine, respectively) have longer rises than B-DNA and are broader, with a larger slide between base pairs, than A-DNA. All of these conforma-

Abbreviation: z_p, z-displacement of nucleotide phosphates.

Data deposition: The atomic coordinates reported in this paper have been deposited in the Protein Data Bank, www.rcsb.org [PDB ID codes 1IH6 for d(GGBr⁵CGCC)₂, 1IH2 for d(GGBr⁵CGBr⁵CC)₂, 1IH3 for d(GGm⁵Cm⁵CC)₂, 1IH4 for d(GGm⁵CGCC)₂, and 1IH1 for the cobalt structure of d(GGCGCC)₂ as B-DNA].

See commentary on page 6986.

*To whom reprint requests should be addressed. E-mail: hops@ucs.orst.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Data collection and refinement statistics

Sequence: Structure(s):	GGBr ⁵ CGCC f, i, j, l	GGm ⁵ CGCC e, i, k, m	GGm ⁵ CGm ⁵ CC	GGBr ⁵ CGBr ⁵ CC	GGCGCC + Co ³⁺ a
Data collection					
Space Group	<i>P</i> 3 ₂ 21	<i>P</i> 3 ₂ 21	<i>P</i> 3 ₂ 21	<i>P</i> 4 ₃ 2 ₁ 2	<i>P</i> 4 ₁ 22
Unit Cell Dimensions	<i>a</i> = <i>b</i> = 41.1 Å, <i>c</i> = 175.1 Å, γ = 120°	<i>a</i> = <i>b</i> = 41.9 Å, <i>c</i> = 176.7 Å, γ = 120°	<i>a</i> = <i>b</i> = 42.3 Å, <i>c</i> = 179.4 Å, γ = 120°	<i>a</i> = <i>b</i> = 60.7 Å, <i>c</i> = 24.6 Å	<i>a</i> = <i>b</i> = 42.6 Å, <i>c</i> = 63.5 Å
Resolution, Å	20.0 – 1.45	20.0 – 1.9	20.0 – 2.4	20.0 – 2.8	20.0 – 2.0
Total reflections (unique)	628,716 (58,430) including Bijvoets	107,358 (14,930)	49,399 (7,269)	10,777 (1,215)	18,846 (4,207)
Completeness (%)*	99.9 (99.6)	99.0 (96.5)	91.7 (61.5)	95.2 (99.1)	97.5 (95.0)
<i>R</i> _{meas} (%)*†	11.0 (66.8)	4.5 (47.5)	6.3 (40.0)	8.1 (54.4)	6.9 (51.8)
Refinement					
<i>R</i> _{cryst} (<i>R</i> _{free}) (%) [‡]	13.8 (18.7)	18.5 (20.4)	17.3 (20.0)	21.1 (25.7)	22.8 (24.5)
DNA (solvent) atoms	968 (234)	968 (115)	976 (75)	244 (2)	240 (37)
rms deviation bond lengths, Å	0.015	0.003	0.003	0.005	0.003
rms deviation bond angles, °	2.29	0.81	0.82	0.95	1.01

*Values in parentheses refer to the highest-resolution shell.

† $R_{meas} = \sum_{hkl} \sqrt{(n/n - 1) \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i I_{hkl,i}}$, where I_{hkl} is the intensity of a reflection, $\langle I_{hkl} \rangle$ is the average of all observations of this reflection and its symmetry equivalents, and n is the multiplicity (16).

‡ $R_{cryst} = \sum_{hkl} |F_{obs} - kF_{calc}| / \sum_{hkl} |F_{obs}|$. $R_{free} = R_{cryst}$ for 10% of the reflections that were not used in refinement (17).

tions, however, have C3'-endo sugars and therefore are generally considered to be allomorphic forms of A-DNA rather than true transition intermediates. We present here a set of single-crystal structures that represent the missing conformational links required to define a common B-A transition pathway for a single DNA sequence.

Methods

All deoxyoligonucleotides were synthesized by using phosphoramidite chemistry on an Applied Biosystems DNA synthesizer in the Center for Gene Research and Biotechnology at Oregon State University. The oligonucleotides were filtered through a Sephadex G-25 (Sigma-Aldrich) column, lyophilized, redis-

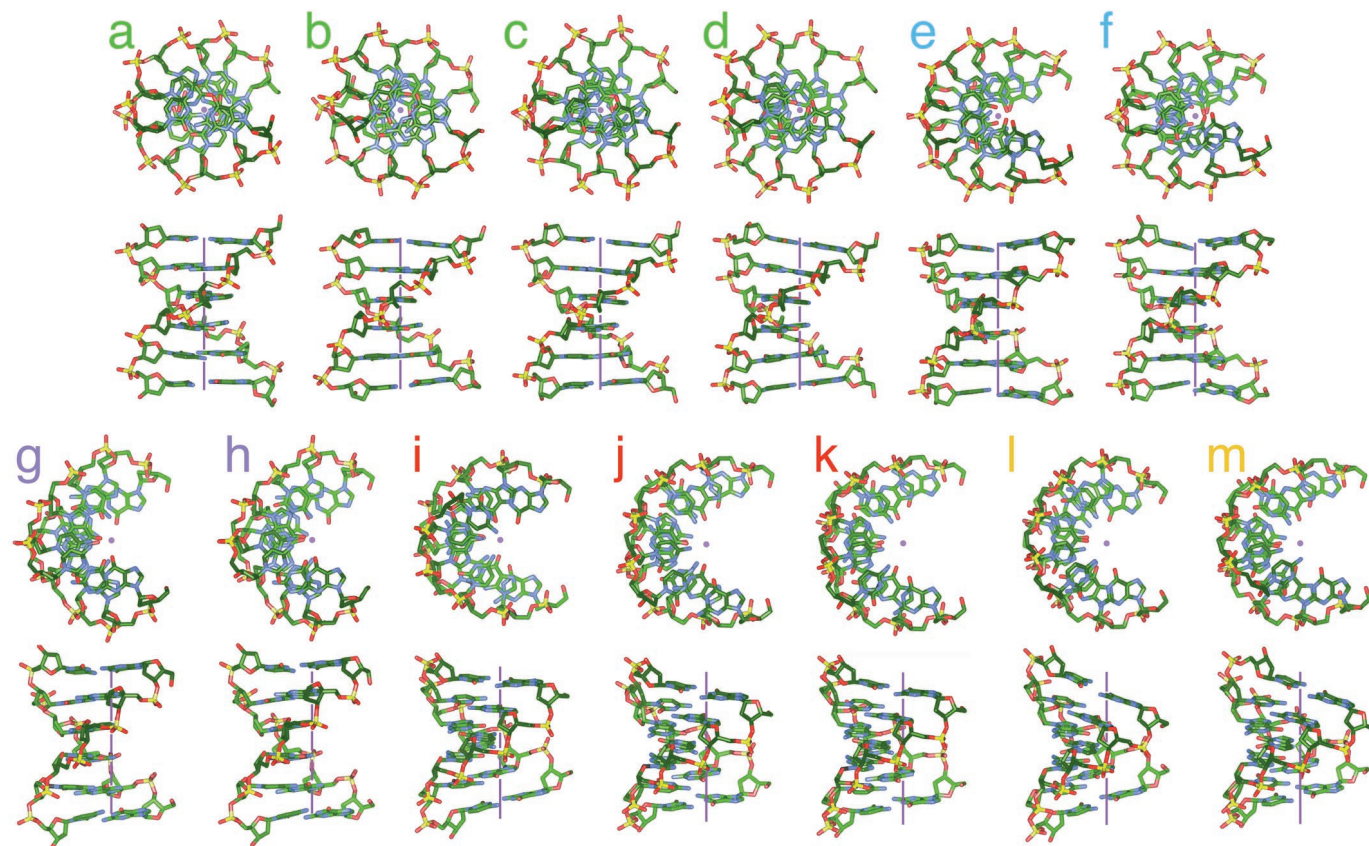


Fig. 2. The 13 unique conformations (a to m) seen in the single-crystal structures of d(GGCGCC)₂, d(GGm⁵CGCC)₂, d(GGBr⁵CGCC)₂, d(GGCGm⁵CC)₂, and d(GGCGBr⁵CC)₂ arranged as described in the text and Table 2. Each structure is viewed down the helix axis (purple dot), and into the helix axis (purple line). Conformations a–d are B-type helices (green labels), e and f are composite helices (blue labels), g and h are extended intermediates (purple labels), i–m are A-DNA allomorphs, starting with A/B-intermediates (red labels) and progressing continuously to standard A-DNA (orange labels). The overall transition through each structure is shown in Movie 1, which is published as supplemental data on the PNAS web site, www.pnas.org.

Table 2. Crystallized sequences and duplex conformations

Sequence	Resolution, Å	Conformation	Structure	Rise/Patterson rise*, Å	Twist, °
GGCGCC		Fiber B-DNA	B	3.38	36.0
GGCGCC + Co ³⁺	2.0	B-DNA	a	3.38 ± 0.08/3.5	35.0 ± 2.3
GGCGCC + spermine ⁴⁺	2.7	B-DNA	b	3.33 ± 0.28/3.4	33.9 ± 1.1
			c	3.35 ± 0.28/3.5	35.9 ± 4.7
			d	3.17 ± 0.47/3.5	33.6 ± 2.8
GGm ⁵ CGCC	1.9	Composite	e	3.36 ± 0.46/3.1	31.7 ± 4.9
		A/B	i	2.93 ± 0.24	31.3 ± 2.5
		A/B	k	3.06 ± 0.11	30.4 ± 2.2
		A-DNA	m	2.68 ± 0.27	30.9 ± 3.7
GGBr ⁵ CGCC	1.4	Composite	f	3.07 ± 0.30/3.2	31.9 ± 3.4
		A/B	i	2.78 ± 0.26	31.5 ± 3.0
		A/B	j	3.02 ± 0.11	31.0 ± 0.9
		A-DNA	l	2.63 ± 0.33	31.0 ± 2.9
GGCGBr ⁵ CC	2.25	Extended	g	3.57 ± 0.40/3.7	30.0 ± 4.7
GGCGm ⁵ CC	2.2	Extended	h	3.55 ± 0.35/3.7	28.4 ± 4.9
GGCGCC		Fiber A-DNA	A	2.56	32.7

The crystal structures of d(GGCGCC)₂, d(GGCGm⁵CC)₂, and d(GGCGBr⁵CC)₂ have been described (6). We have resolved the Co³⁺ form of d(GGCGCC)₂ here to higher resolution than before.

*Values for Patterson rise (the rise between base planes in the Patterson maps calculated from the refined structures) are included as measures of an intrinsic helical parameter (29) for comparison with the values determined by CURVES (30).

solved in 15 mM sodium cacodylate buffer (pH 7.0), and used for crystallization without further purification. Crystals were grown at room temperature by vapor diffusion in sitting drop setups with initial solutions containing 0.7 mM DNA, 25 mM sodium cacodylate buffer (pH 6), 0.8 mM MgCl₂, and 0.1–1.2 mM spermine tetrahydrochloride, and were equilibrated against a reservoir of 15% (vol/vol) 2-methyl-2,4-pentanediol. The crystal of d(GGCGCC)₂ was grown, as above, with the addition of 2 mM Co(NH₃)₃³⁺. These conditions are similar to those of previous studies (6).

X-ray diffraction data for the Co³⁺ form of d(GGCGCC)₂, d(GGm⁵CGm⁵CC)₂, d(GGm⁵CGCC)₂, and d(GGBr⁵CGBr⁵CC)₂ were collected at room temperature by using CuK_α radiation from a RUH3R generator and R-AXIS IV image plate detector (Rigaku, Tokyo). Multiwavelength x-ray diffraction data for d(GGBr⁵CGCC)₂ were collected on beamline 5.0.2 at the Advanced Light Source in Berkeley, CA, at liquid nitrogen temperature by using an ADSC Quantum IV detector (San Diego). Three wavelengths corresponding to the peak of bromine absorption (0.92052Å), the inflection (0.92065Å), and a remote (0.90836Å) were used in CNS (14) to determine the phases for d(GGBr⁵CGCC)₂. Diffraction data collected at the peak wavelength were used in SHELXL-97 (15) for refinement.

The crystals of d(GGm⁵CGCC)₂ and d(GGm⁵CGm⁵CC)₂ were isomorphous with those of d(GGBr⁵CGCC)₂; consequently, their structures were solved by molecular replacement by using the four unique duplexes of the d(GGBr⁵CGCC)₂ structure as the starting model. The bromine atoms were removed and methyl groups added to the appropriate cytosines. The initial values of R_{cryst} and R_{free} were 42.6 and 41.6%, respectively, for d(GGm⁵CGCC)₂, and 51.2 and 51.0% for d(GGm⁵CGm⁵CC)₂. The structures were then refined in CNS (14) by using the maximum likelihood target while maintaining the same crossvalidation set as d(GGBr⁵CGCC)₂, yielding the final refinement statistics in Table 1. The structure of d(GGBr⁵CGBr⁵CC)₂ was solved by molecular replacement by using the previously solved extended structure of d(GGCGBr⁵CC)₂ (6) and refined by using the same technique. The structure of the Co³⁺-form of d(GGCGCC)₂ was refined with higher-resolution data by using the previously solved and refined structure (6). All datasets were reduced by using the programs DENZO and SCALEPACK from the HKL package (18) and structures, with the exception of d(GGBr⁵CGCC)₂, were

refined with CNS (14) incorporating nucleic acid specific parameters (19).

Results

Single-Crystal Structures of d(GGBr⁵CGCC)₂ and d(GGm⁵CGCC)₂. For the current study, we have grown crystals of the sequences d(GGm⁵CGm⁵CC)₂, d(GGm⁵CGCC)₂, and d(GGBr⁵CGCC)₂, which are all isomorphous to each other but differ from all previous oligonucleotide sequences (Table 1). The crystal structures of these sequences were seen to include a conformation that is a composite of A- and B-DNA. The structure of d(GGBr⁵CGCC)₂ was solved to 1.6-Å resolution by using multiwavelength anomalous diffraction phasing with bromine as the anomalous scatterer (20). The resulting experimental electron density maps allow unbiased interpretations of four independent DNA double helices in the asymmetric unit of the crystal, as well as the detailed conformational features of the individual nucleotides (Fig. 1A). This structure was subsequently used to determine the structures of d(GGm⁵CGm⁵CC)₂ and d(GGm⁵CGCC)₂ by molecular replacement (Table 1). The DNAs of these sequences are arranged with three duplexes forming a planar canopy in the *a*-*b* plane, whereas a fourth duplex extends perpendicularly from near the center of this planar assembly. One duplex in the plane is a classic A-DNA structure, characterized by C3'-*endo* sugar puckers, and highly inclined and displaced base pairs. The remaining two helices in the plane have C3'-*endo* sugars and base pairs that are displaced from but not highly inclined relative to the helix axis, similar to the A/B intermediate of d(CATGGGCCCATG)₂ (13).

The fourth duplex that bridges these planes of A-like structures is a composite double-helical conformation. The three nucleotides at the 5'-end of each strand have A-type C3'-*endo* sugars (Fig. 1B), whereas the last three nucleotides have C2'-*endo* and C1'-*exo* sugars, which fall within the B-type C2'-*endo* family, and an O4'-*endo* sugar, which is the lowest-energy intermediate between the C2'- and C3'-*endo* conformations. In addition, the δ and χ dihedral angles show the three nucleotides at the 5'-end of each strand to be A-DNA and those of the three 3'-nucleotides to be B-DNA. Thus, the A-type nucleotides at the 5'-end of each strand are paired with B-type nucleotides at the 3'-end of the complementary strand. This mixture of A- and B-DNA base pairing is similar to the A-RNA/B-DNA chimera seen by fiber diffraction (24), but differs from the A-RNA

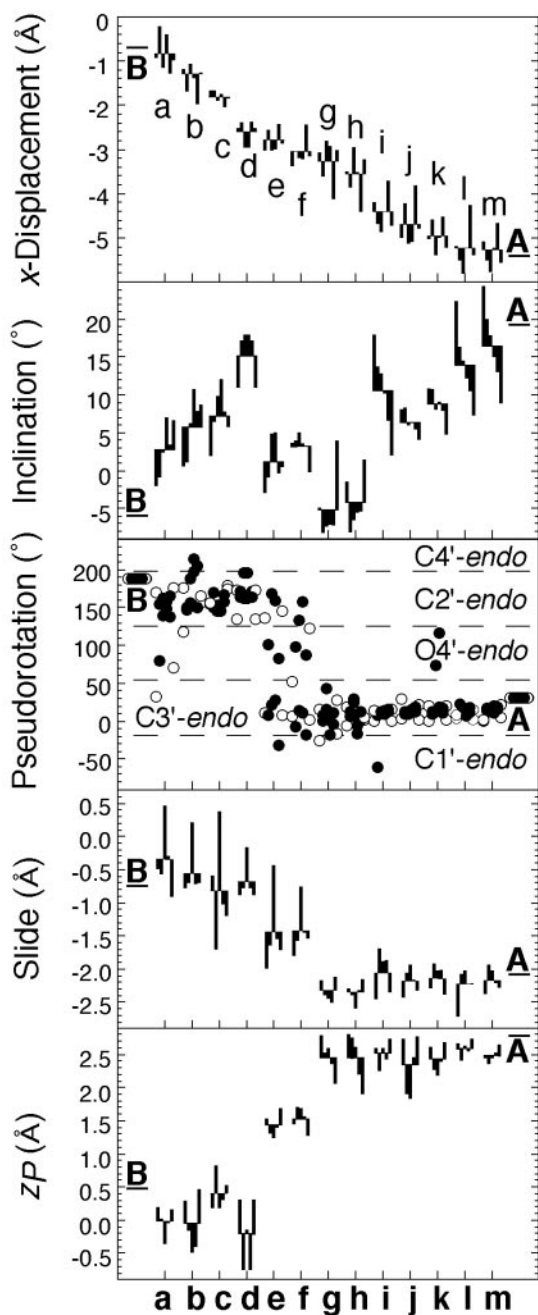


Fig. 3. Helical parameters of the crystal structures of $d(\text{GGCGCC})_2$ and its methylated and brominated analogs. The x -displacement, inclination of the base pairs, pseudorotation phase angle, base-pair slide, and z_P are plotted, from top to bottom, for the 13 unique helical conformations (a–m) arranged along a common transition pathway according to their x -displacement. Helical parameters were calculated with CURVES 5.2 (30), with the exception of z_P , which was calculated by $x_3\text{DNA}$ (23). The average values for the parameters are shown for each structure, with the bars extending above and below the average reflecting the values for the individual base pairs or dinucleotide steps. Values for ideal B- and A-DNA in each plot are labeled B and A, respectively. Negative x -displacements reflect shifts of the base pairs away from the helix axis toward the minor groove; negative inclination angles arise from base pairs being tipped away from perpendicular of the helix axis and toward the 3'-end of each strand, and slide has the stacked base pairs displaced relative to each other along their long axes (30). Pseudorotation phase angles reflect the conformation of the deoxyribose sugars (labeled according to their respective *endo*-type families), with open circles representing the pseudorotation angle of the end nucleotides and closed circles the internal nucleotides of each strand. The parameter z_P measures the displacement of the phosphates along each strand away from the midpoint between two stacked base pairs (31).

structures of DNA/RNA hybrids (25, 26). The composite conformation seen here is analogous to the junction between B- and A-DNA first proposed by Arnott and colleagues (27), and to the intermediate for the B-A transition in poly(dG)·poly(dC) suggested from Raman spectroscopy (28).

We have also shown that the sequence $d(\text{GGBr}^5\text{CGBr}^5\text{CC})_2$ is nearly identical to the previously solved structures of $d(\text{GGCGm}^5\text{CC})_2$ and $d(\text{GGCGBr}^5\text{CC})_2$, both of which had extended helical rises compared with both A- and B-DNAs (6). Finally, we have resolved the B-DNA structure of $d(\text{GGCGCC})_2$ in its cobalt form to a higher (2.0-Å) resolution than reported (6).

The B-A Transition Pathway. The structures of $d(\text{GGm}^5\text{CGCC})_2$ and $d(\text{GGBr}^5\text{CGCC})_2$, along with the previous B-conformations of $d(\text{GGCGCC})_2$ and extended conformations of $d(\text{GGCGm}^5\text{CC})_2$ and $d(\text{GGCGBr}^5\text{CC})_2$, form a set of 13 unique structures (designated a to m, Table 2) that define a common transition pathway (Fig. 2). The structures were sorted according to their increasingly negative x -displacement, starting with the structure that most closely resembles fiber B-DNA (a), and ending with the structure that most closely resembles fiber A-DNA (m) (Fig. 3). This yields a continuous and monotonic transformation of duplexes with progressively deeper major grooves. Central to this transition pathway are the composite intermediates (e and f) that bridge the gap between the B- and A-type conformations. The structures arranged in this way also show a continuous progression of the conformations away from standard B-DNA. The rms deviation of the atoms for each conformation relative to standard B-DNA (a) systematically increases as the structures proceed from a to m. Sorting these structures according to other parameters, for example by increasing inclination of the base pairs, results in nonlogical pathways that intermix B- and A-DNAs throughout the transition.

The pathway as we have defined it is consistent with a cooperative B-A transition according to the sugar puckers (as measured by the pseudorotation phase angle), the slide between base pairs, and the displacement of the phosphates reflected in the parameter z_P (31) (z_P , z -displacement of nucleotide phosphates) (Fig. 3). These three helical parameters, considered to be reliable discriminators between the B- and A-DNA forms (23), are highly correlated. All of the B-type duplexes (a to d) have C2'-*endo* type sugar puckers, small base-pair slides, and phosphates positioned equidistant between the stacked base pairs of each dinucleotide step ($z_P \approx 0$). The A-type structures (g to m) have C3'-*endo* sugars, large negative base-pair slides, and phosphates pushed toward the 3'-end of each dinucleotide ($z_P \approx 2.5$ Å). The composite conformations (e and f) are truly transition intermediates, mixing both C2'-*endo* and C3'-*endo* type sugars in the same duplex, and having intermediate slides. The intermediate values of z_P reflect the phosphates being pushed toward the 3'-end of the A-type nucleotides, whereas the phosphates of the B-nucleotides remain at the midpoint between base pairs.

There is a simple relationship between displacement and inclination for the B-DNA structures (a to d): as the base pairs become displaced from the helix axis, they are more inclined. In the extreme case, the structure d is nearly halfway to A-DNA in terms of the displacement and inclination of its base pairs. A conformation with A-type base stacking but B-type sugars had been proposed as the intermediate for the B-A transition of $d(\text{CCCCGGGG})_2$ in solution (32).

Displacement and inclination are also correlated in the A-type structures (g to m), but the trend is not simply an extension from the B-DNA structures. The first set of A-type structures (g and h) show no base-pair inclination. Indeed, the opposite trend is seen as the structures actually switch from B- to A-DNA (d to h), with the base pairs reset to the inclination of fiber B-DNA. From g to m, we see the DNA base pairs becoming more displaced, inclined, and compact in a near continuous manner,

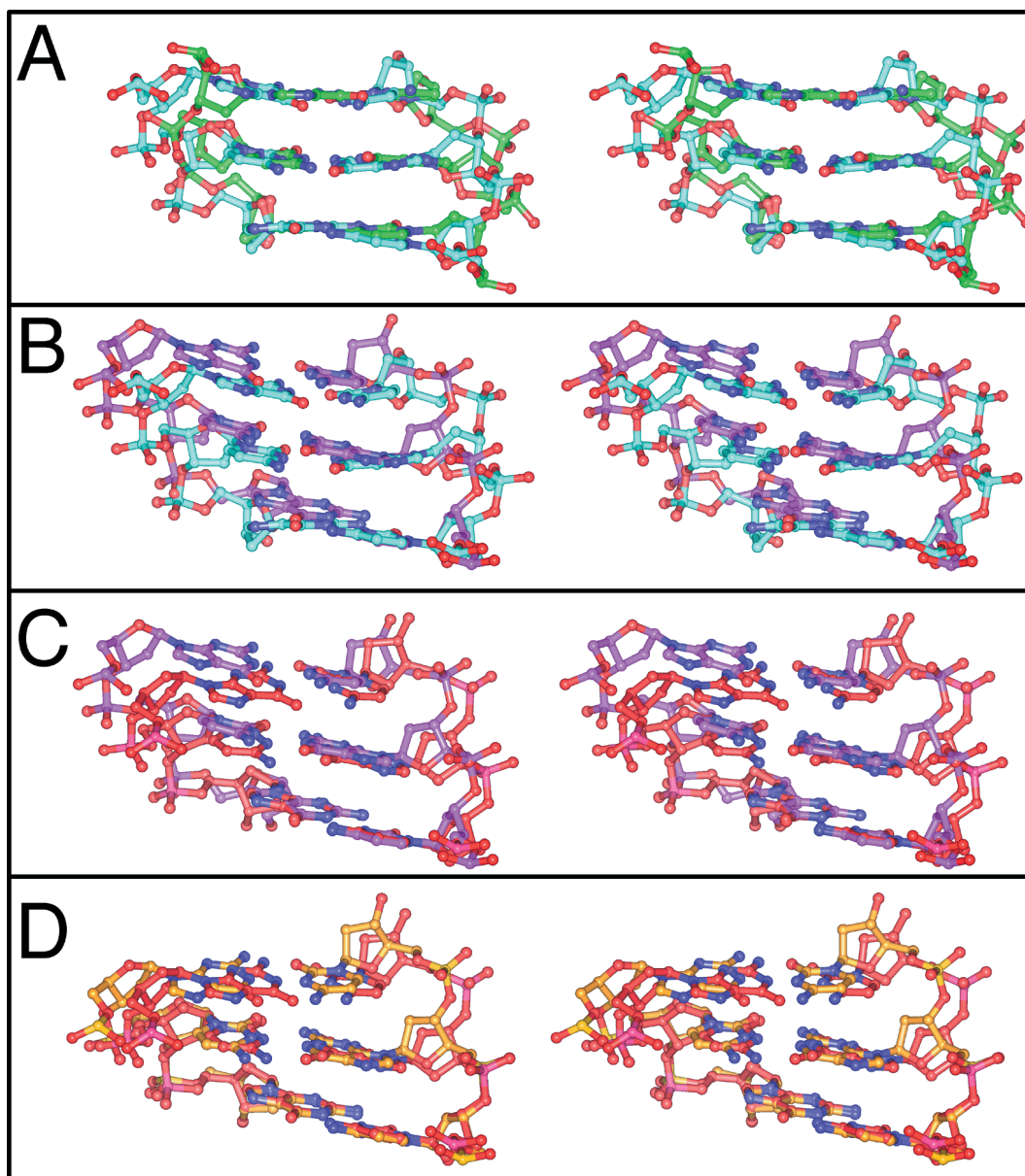


Fig. 4. Stereoview of the individual steps of the B- to A-DNA transition at the central base pairs of the double helix. *A–D* show the view into the major groove of the three base pairs (G2-C11, C3-G10, and G4-C9) from each conformational class superimposed on base pairs from the next conformational class along the transition pathway. The atoms of G4-C9 (lower base pairs) serve as the common reference to superimpose the structures. The carbons and phosphorus atoms of the base pairs are colored to distinguish between the conformations. (A) The transformation of standard B-DNA structure (d, green) to the composite intermediate (e, blue) shows the effect of changing half the sugars from C2'-*endo* to C3'-*endo*. (B) The transformation from e to the extended intermediate (g, purple) involves the complete conversion of the sugar puckers, resulting in the increase in the rise, slide, and unwinding of base pairs. (C) The transition to an A/B intermediate (j, red) results in the reduction of the rise and slide, but with very little inclination of the base pairs. (D) The final conversion to canonical A-DNA (m, orange) shows the compression of the major groove resulting from the inclination of the base pairs. These details for the individual base pairs are shown in Movie 2, which is published as supplemental data on the PNAS web site, www.pnas.org.

passing through a set of conformations that are similar to the A/B-intermediates of $d(\text{CATGGGCCCATG})_2$ (13).

All variants of this sequence were crystallized from nearly identical solutions, indicating that the conformations are associated with modifications to the cytosine bases and are not artifacts of crystallization. This provides the rationale for placing these structures along a common pathway. The coexistence of the composite and A/B-intermediates, and standard A-DNA within the crystals of $d(\text{GGm}^5\text{CGCC})_2$ and $d(\text{GGBr}^5\text{CGCC})_2$, indicates that these three forms are energetically related. The crystallization of both $d(\text{GGm}^5\text{CGCC})_2$ and $d(\text{GGm}^5\text{CGm}^5\text{CC})_2$ as the composite intermediate and $d(\text{GGCGm}^5\text{CC})_2$ as the extended intermediate shows

that the two conformations depend only on methylation patterns; therefore, they are energetically similar and should be readily interconvertible. The extended conformation of $d(\text{GGCGm}^5\text{CC})_2$ had previously been shown to be a kinetically trapped intermediate that leads ultimately to the formation of A-DNA (6). Finally, it is evident that $d(\text{GGCGCC})_2$ and its variants crystallize in conformations that assume several possible crystal lattices, and therefore the conformations are not defined explicitly by crystal packing.

Discussion

We describe here the stepwise transition between two distinct conformations of a macromolecule, in this case from B- to

A-DNA. This B-A transition shows many of the features first proposed by Drew and Calladine for this mechanism (33). Sugar pucker and slide are highly correlated throughout the transition, whereas inclination is highly variable. Indeed, the base pairs can be highly inclined in B-type DNAs, but they must be reset back to their noninclined geometries before continuing through the transformation to A-DNA. This is generally consistent with the Drew and Calladine mechanism, in which slide precedes inclination. However, the intermediate structures seen here also show that the change in sugar pucker is associated with an extension and unwinding of the helix.

The transition mapped through structures **a** to **m** incorporates a set of distinct conformational intermediates. We were able to trap these intermediates because the d(GGCGCC)₂ sequence is resistant to forming A-DNA. This sequence was predicted from A-DNA propensity energies (34) to favor B-DNA, whereas a permutation of the sequence—(GCCGGC)₂—was predicted to be and was crystallized only as canonical A-DNA (35).

In this transition pathway, B-DNA is first transformed to the composite intermediates in which half of each strand in the duplex is A-DNA (Fig. 4A). The C3'-endo sugars are associated with a slide between base pairs and extension of the backbone of the A-type nucleotides at the 5'-end of each strand. The complementary B-type nucleotides, however, oppose these perturbations; consequently, the composite A-B base pairs have only intermediate values for slide and are buckled to accommodate the backbone extension.

In the next series of steps, the transformation goes through the extended intermediates (**g** and **h**). With all nucleotides now adopting C3'-endo sugar conformations, the slide and extension between base pairs are no longer opposed, and both reach their maximum values (Fig. 4B). Although **g** and **h** could be considered A-type conformations, they are extended [Patterson rise, which measures the intrinsic rise between base pairs = 3.7 Å (29); Table 2], underwound (helical repeat ≥ 12 bp/turn), and have exaggerated base-pair slides; therefore, they should be considered to be at the

interface between the transition intermediates and true A-DNA structures. It is not unreasonable that the transition between conformational states of DNA goes through an intermediate that bears little resemblance to either the starting or ending states. From this point, the DNA duplex becomes more compact, with a shortened rise, and rewinds itself to an 11-bp/turn repeat (Fig. 4C). In the final stages, the transition to the canonical A-DNA structure is nearly continuous, passing through the A/B-type intermediates along the way (Fig. 4D).

The detailed pathway is explicitly defined here for the B-A transition induced by methylation or bromination, but the common occurrences of a B-type conformation with A-like features and of the A/B intermediates in the current study and in previous unrelated sequences (13, 32) suggest that this pathway may generally be applicable to B-A transitions in other systems. For example, the A-DNA region of the cisplatin-bound A-B junction is located at the 5'-end of the crosslinked DNA strand (3), suggesting that A-type features propagate backwards from the point where the B-DNA helix is perturbed. Comparing d(GGm⁵CGCC)₂, which has A-type nucleotides at the 5'-end of each strand, with d(GGCGm⁵CC)₂, which has entirely A-type nucleotides, suggests that the A-conformation is initiated at the point of methylation and is propagated back toward the 5'-end of the strands. Finally, the transition to the conformation induced by TATA-binding protein requires only that the pathway be extended to even more exaggerated base-pair inclinations [from ≈ 20° for standard A-DNA to ≈ 40° for the promoter induced structure (36)].

We thank P. A. Karplus and Z. A. Wood for a critical reading of this manuscript and Drs. X.-J. Lu and W. K. Olson at Rutgers University for discussion on the X3DNA analysis program. This work was supported by the National Science Foundation (MCB-0090615) and the National Institutes of Environmental Health Sciences (ES00210). X-ray facilities were funded in part by the M. J. Murdock Charitable Trust and the Proteins and Nucleic Acids Research Core of the Environmental Health Sciences Center at Oregon State University.

1. Watson, J. D. & Crick, F. H. C. (1953) *Nature (London)* **171**, 737–738.
2. Franklin, R. E. & Gosling, R. G. (1953) *Nature (London)* **171**, 740–741.
3. Takahara, P. M., Rosenzweig, A. C., Frederick, C. A. & Lippard, S. J. (1995) *Nature (London)* **377**, 649–652.
4. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993) *Nature (London)* **365**, 512–520.
5. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993) *Nature (London)* **365**, 520–527.
6. Vargason, J. M., Eichman, B. F. & Ho, P. S. (2000) *Nat. Struct. Biol.* **7**, 758–761.
7. Arnott, S. (1999) in *Oxford Handbook of Nucleic Acid Structure*, ed. Neidle, S. (Oxford Univ. Press, New York), pp. 1–38.
8. Cheatham, T. E., III, & Kollman, P. A. (1996) *J. Mol. Biol.* **259**, 434–444.
9. Cheatham, T. E., III, & Kollman, P. A. (1997) *Structure (London)* **5**, 1297–1311.
10. Cheatham, T. E., III, Crowley, M. F., Fox, T. & Kollman, P. A. (1997) *Proc. Natl Acad. Sci. USA* **94**, 9626–9630.
11. Wang, A. H.-J., Fujii, S., van Boom, J. H. & Rich, A. (1982) *Proc. Natl Acad. Sci. USA* **79**, 3968–3972.
12. Malinina, L., Fernandez, L. G., Huynh-Dinh, T. & Subirana, J. A. (1999) *J. Mol. Biol.* **285**, 1679–1690.
13. Ng, H.-L., Kopka, M. L. & Dickerson, R. E. (2000) *Proc. Natl Acad. Sci. USA* **97**, 2035–2039. (First published February 25, 2000; 10.1073/pnas.040571197)
14. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998) *Acta Crystallogr. D* **54**, 905–921.
15. Sheldrick, G. M. & Schneider, T. R. (1997) *Methods Enzymol.* **277**, 319–343.
16. Diederichs, K. & Karplus, P. A. (1997) *Nat. Struct. Biol.* **4**, 269–275.
17. Brünger, A. T. (1992) *Nature (London)* **355**, 472–475.
18. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307–326.
19. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996) *Acta Crystallogr. D* **52**, 57–64.
20. Hendrickson, W. A. & Ogata, C. M. (1997) *Methods Enzymol.* **276**, 494–523.
21. Esnouf, R. M. (1999) *Acta Crystallogr. D* **55**, 938–940.
22. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992) *Biophys. J.* **63**, 751–759.
23. Lu, X.-J., Shakked, Z. & Olson, W. K. (2000) *J. Mol. Biol.* **300**, 819–840.
24. Arnott, S., Chandrasekaran, R., Millane, R. P. & Park, H.-S. (1986) *J. Mol. Biol.* **188**, 631–640.
25. Horton, N. C. & Finzel, B. C. (1996) *J. Mol. Biol.* **264**, 521–533.
26. Egli, M., Usman, N., Zhang, S. & Rich, A. (1992) *Proc. Natl Acad. Sci. USA* **89**, 534–538.
27. Selsing, E., Wells, R. D., Alden, C. J. & Arnott, S. (1979) *J. Biol. Chem.* **254**, 5417–5422.
28. Nishimura, Y., Torigoe, C. & Tsuboi, M. (1986) *Nucleic Acids Res.* **14**, 2737–2748.
29. Vargason, J. M. & Ho, P. S. (2001) *Nat. Struct. Biol.* **8**, 107–108.
30. Lavery, R. & Sklenar, H. (1989) *J. Biomol. Struct. Dyn.* **6**, 655–667.
31. El Hassan, M. A. & Calladine, C. R. (1997) *Philos. Trans. R. Soc. London A* **355**, 43–100.
32. Trantírek, L., Štefl, R., Vorlíčková, M., Koča, J., Sklenář, V. & Kypr, J. (2000) *J. Mol. Biol.* **297**, 907–922.
33. Calladine, C. R. & Drew, H. R. (1984) *J. Mol. Biol.* **178**, 773–782.
34. Basham, B., Schroth, G. P. & Ho, P. S. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6464–6468.
35. Mooers, B. H. M., Schroth, G. P., Baxter, W. W. & Ho, P. S. (1995) *J. Mol. Biol.* **249**, 772–784.
36. Guzikovich-Guerstein, G. & Shakked, Z. (1996) *Nat. Struct. Biol.* **3**, 32–37.