B-DNA structure is intrinsically polymorphic: even at the level of base pair positions

Tatsuya Maehigashi, Chiaolong Hsiao, Kristen Kruger Woods, Tinoush Moulaei, Nicholas V. Hud and Loren Dean Williams*

School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA

ABSTRACT

Increasingly exact measurement of single crystal X-ray diffraction data offers detailed characterization of DNA conformation, hydration and electrostatics. However, instead of providing a more clear and unambiguous image of DNA, highly accurate diffraction data reveal polymorphism of the DNA atomic positions and conformation and hydration. Here we describe an accurate X-ray structure of B-DNA, painstakingly fit to a multistate model that contains multiple competing positions of most of the backbone and of entire base pairs. Two of ten base-pairs of CCAGGCCTGG are in multiple states distinguished primarily by differences in slide. Similarly, all the surrounding ions are seen to fractionally occupy discrete competing and overlapping sites. And finally, the vast majority of water molecules show strong evidence of multiple competing sites. Conventional resolution appears to give a false sense of homogeneity in conformation and interactions of DNA. In addition, conventional resolution yields an average structure that is not accurate, in that it is different from any of the multiple discrete structures observed at high resolution. Because base pair positional heterogeneity has not always been incorporated into model-building, even some high and ultrahigh-resolution structures of DNA do not indicate the full extent of conformational polymorphism.

INTRODUCTION

Increasingly accurate X-ray structures offer highly detailed characterization of macromolecular conformation, hydration and counterion interactions (1–4). However, instead of providing clear and unambiguous structures, accurate diffraction data of B-DNA indicate extensive heterogeneity, in conformation, ion and water molecule positions and occupancies. We describe the structure of a helical turn of B-DNA (Figure 1) determined from diffraction data that extends to atomic resolution. The data reveal an inherent polymorphism in the positions of atoms, in conformation and in molecular interactions, which is not obvious at lower resolution and has not been observed previously to the extent described here. This high-resolution data reveal that intermediate and low-resolution data can give a false sense of homogeneity. Specifically, a comparison with a lower resolution structure of the same B-DNA fragment (5) reveals that lower resolution gives a model in which multiple states are averaged. The average model is not necessarily representative of any of the states observed at high resolution.

Heinemann's structure (5) of CCAGGCCTGG (called here $\text{CCAGGCCTGG}^{1.6}$, 1.6 Å resolution, 2422 unique reflections, PDB entry 1BD1) is a complete turn of B-form double helix with a full complement of Watson–Crick base pairs. This structure was one of the highest resolution and most accurate B-form structures of its era. The quality of the data obtained by Heinemann suggested that this DNA fragment could provide a platform, using modern synchrotron radiation and refinement methods, along with careful model building, and substitution with anomalous scatters, for complete and highly accurate characterization of B-DNA and its saline environment. A resulting high-resolution structure, $\text{CCAGGCCTGG}^{0.96}$ (0.96 Å resolution, 14 269 unique reflections), is isomorphous with $\text{CCAGGCCTGG}^{1.6}$ in that the unit cell, space group and global position of the DNA are conserved. However, new methods allow observation of multiple states of DNA (Figures 1–3) and determination of fractional occupancies, along with anisotropic displacement factors for $\text{CCAGGCCTGG}^{0.96}$. The data indicate substantial polymorphism of the DNA backbone. The phosphate groups of most nucleotides are in multiple, discrete and overlapping positions, as anticipated from the earliest single crystal diffraction studies of DNA (6) and tRNA (7) and from subsequent high resolution work (1–3).

The data indicate multiple positions of entire base pairs (Figure 2C), with base atoms translated by up to 1.8 Å between the two states (Figure 3). $\text{CCAGGCCTGG}^{0.96}$

---

*To whom correspondence should be addressed. Tel: +1 404 385 6682; Fax: +1 404 894 2295; Email: loren.williams@chemistry.gatech.edu
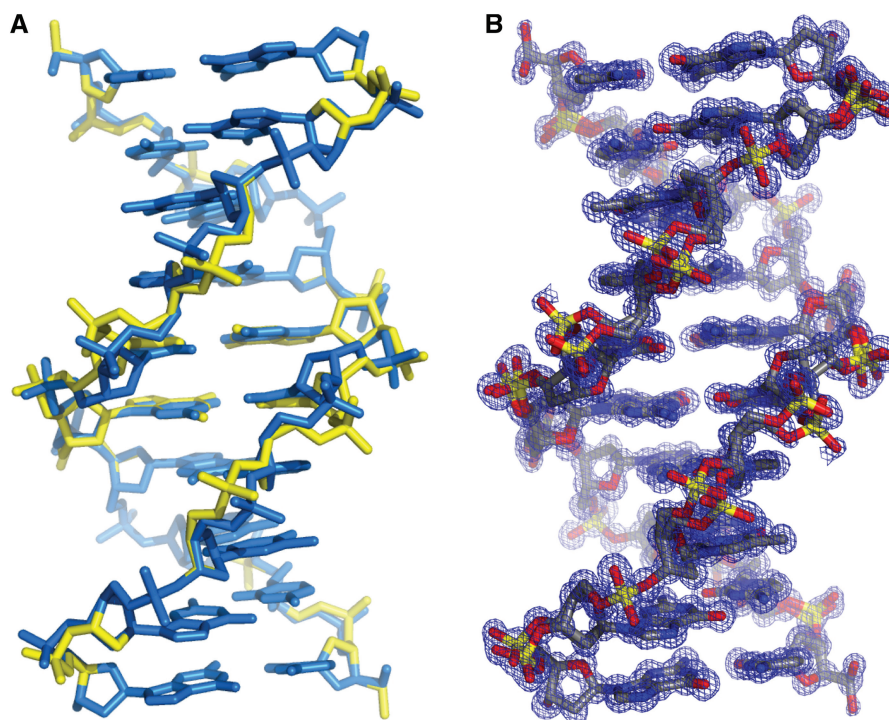
Figure 1. (**A**) The structure of CCAGGCCTGG$^{0.96}$ showing indicating multiple states. State A is blue and state B is yellow. The bases are numbered 1–10 in the first strand, 11–20 in the other. (**B**) The $2F_o − F_c$ electron density map surrounding the DNA only (blue net), contoured at 1σ.

(2 of 10 bp) are in multiple states distinguished primarily by differences in slide. In contrast to previous work (8–10), none of the counterions are observed at single positions. Magnesium ions are found in previously identified regions (9–12), but are located in multiple, discrete and overlapping positions (Figure 4). Similarly, the majority of water molecules are in multiple positions.

The results here suggest that conventional resolution can give an artificial sense of homogeneity in DNA conformation and interactions. The structure of the lower resolution CCAGGCCTGG$^{1.6}$ does not appear to represent any of the multiple structures suggested by our new high-resolution structure. Instead, the conformation of C CAGGCCTGG$^{1.6}$ appears to be a population weighted average of the conformations of CCAGGCCTGG$^{0.96}$. Further, we have found that base pair positional heterogeneity is more frequent than generally appreciated, even in ultra-high resolution structures. Our survey of the database found that base pair positional heterogeneity was incorrectly omitted from at least one previous ultra-high resolution structure of DNA. Re-refinement and careful model building confirms multiple base pairs of that structure. Therefore, even ultra-high resolution structures can suggest an incorrect degree of homogeneity.

MATERIALS AND METHODS

### Crystallization and data collection

Reverse-phase HPLC purified d(CCAGGCCTGG) (Integrated DNA Technology) was annealed by slow cooling. Crystals were grown by hanging drop vapor diffusion, from a drop initially containing 0.23 mM d(CCAGGCCTGG) (strand), 12.5 mM magnesium acetate, 25 mM sodium cacodylate (pH 6.5) and 20% (v/v) 2-methyl-2,4-pentanediol (MPD). The drop was equilibrated at 4°C against a reservoir of 25 mM magnesium acetate, 50 mM sodium cacodylate (pH 6.5) and 40% MPD. Plate-like crystals with dimensions $0.2 \times 0.1 \times 0.05$ mm^3 appeared within 2–3 days.

Crystals were looped and flash frozen in liquid nitrogen. X-ray intensities were collected at beamline 22-ID in the SER-CAT facility at the Advanced Photon Source. A total of 360° of data with 1° oscillation were collected on a MAR 225 CCD detector (Mar Research GmbH, Germany) using 1.0 Å radiation (Table 1). The crystal was maintained at 113 K during data collection. A total of 92 542, reflections were collected, indexed and reduced to 14 269 unique reflections to a maximum resolution of 0.92 Å with the program HKL 2000 (13). This DNA crystallizes in space group C2, with unit cell parameters of $a = 32.14$ Å, $b = 25.17$ Å, $c = 34.09$ Å, α = γ = 90° and β = 116.3° (Table 1).

Structure solution and refinement

Phase determination was carried out by molecular replacement using the coordinates from CCAGGCCTGG$^{1.6}$ with CNS version 1.1 (15). After several rounds of simulated annealing and refinement of the DNA alone using the parameters of Berman and co-workers (16,17), the model was transferred to the program REFMAC5 (18), preserving the R-free flags. Hydrogen atoms were added in
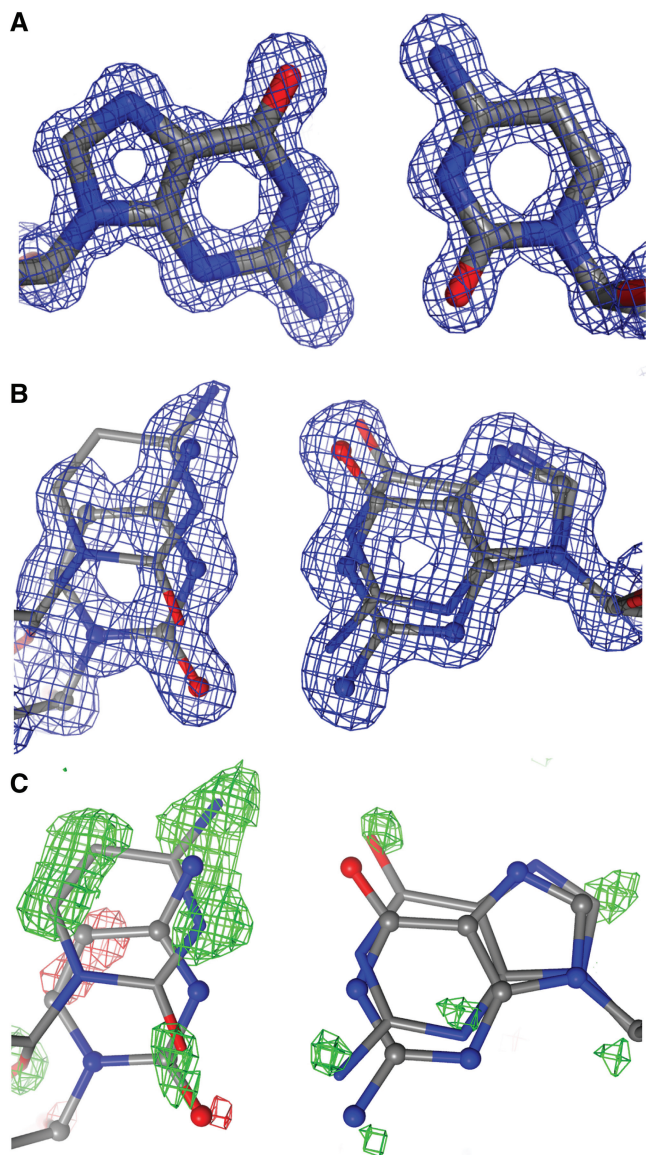
Figure 2. (**A**) $2F_o - F_c$ electron electron density map (sum map) showing a base pair [C(1)–G(11)] that is found in a single position. (**B**) Electron density showing two discrete positions of the central G(5)–C(16) base pair. This map is unbiased in that phases were calculated from the initial, single-position model even though atoms of both states are shown. (**C**) Difference electron density surrounding the central G(5)–C(16) base pair. The positive difference density is green, indicating where additional base atoms should added to give a better fit of the model to the data. These difference maps is unbiased, with phases calculated from the initial, single-position model. The atoms of the final multiple-position model are shown but were not used for phase calculation. The negative difference density is red, indicating where atoms should be removed from the model to give a better fit. For all three panels, the sum maps are contoured at contoured at 1σ. Difference maps are contoured at 2.5σ.

the riding positions during the refinement process, with anisotropic refinement of all atoms. The refinement was continued to convergence, with *R*-free and thermal ellipsoids monitored to avoid over refinement. The models were manipulated manually with the program Coot (19). Thermal ellipsoids were computed with ORTEP-3 for Windows (20).
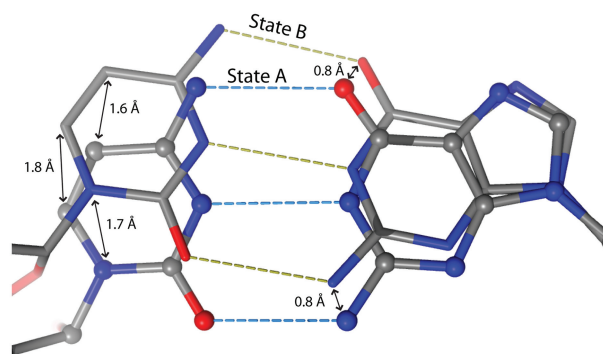
Figure 3. The G(5)–C(16) base pair is in two discrete positions.
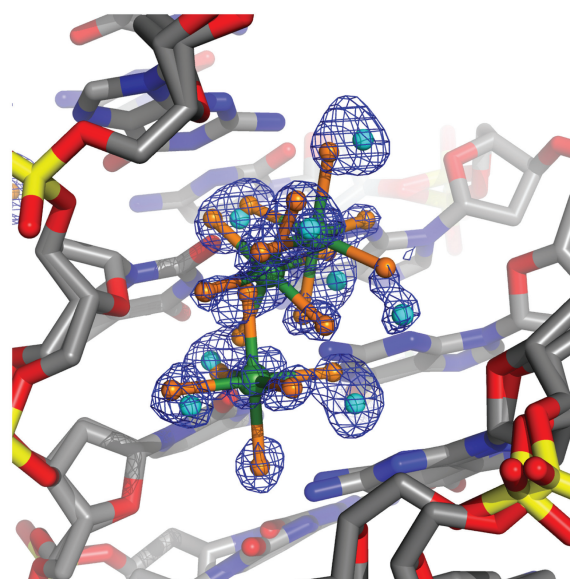
Figure 4. The major groove. Three hexahydrated magnesium complexes complete for overlapping sites within the major groove of the DNA. The data here support a model in which a magnesium ion in solution would shift between sites, with occupancy of only one site at a time. In the X-ray structure, these complexes are partially occupied, indicating occupancy on an either/or basis in the crystalline ensemble.

A total of 12 848 unique reflections in the resolution range 16.1–0.96 Å were used in the refinement, with no sigma filter. The final *R* factor is 10.3% for all data and the *R*-free is 12.5%. The asymmetric unit contains a single-strand of d(CCAGGCCTGG), which is paired to another strand by a crystallographic 2-fold axis. The final electron density maps are clean and unambiguous (Figure 1B).

Multiple positions of DNA atoms

A single conformation of the DNA (state A) was initially assigned and refined. For a few of the phosphate groups and 8 of 10 bp, the data fit well to a single base pair position (Figure 2A). A second state (state B) was clearly apparent based on electron density adjacent to the atoms of state A. State B was built and the combined models were refined. Two of the base pairs

Table 1. Crystallographic and refinement statistics

Wavelength (Å)	1.00
Space group	C2
Unit cell	$a = 32.136$ Å, $b = 25.172$ Å, $c = 34.094$ Å
	$\alpha = \gamma = 90°$, $\beta = 116.25°$
Resolution range (Å)	16.0–0.96
Number of reflections (all)	92 542
Number of unique reflections	12 848
Completeness (%)[a]	92.44 (50.27)
Average I/σ (I)	53.4
R_{merge} (%)[b]	7.9
Refinement statistics	
DNA (asymmetric unit)	d(CCAGGCCTGG)
Number of DNA atoms	287
Number of water molecules[c]	82 (excluding Mg^{2+} first shell)
Number of Mg^{2+} $(H_2O)_6$[c]	5 (all partially occupied)
R_{work} (%)[d]	10.3
R_{free} (%)[e]	12.5 (1087 reflections)
RMS deviation of bonds from ideal (Å)	0.012
RMS deviation of angles from ideal (°)	1.98
Average isotropic B value	8.61
PDB ID code	3GGB

[a]The values in parentheses refers to the highest resolution shells.
[b]$R_{\mathrm{merge}} = \Sigma|I - <I>|/\Sigma I$, where $I =$ observed intensity and $<I> =$ mean intensity obtained from multiple observations, including of symmetry-related reflections.
[c]Includes partially occupied species as well as atoms on special position.
[d]$R_{\mathrm{work}} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. Reflections flagged for the Free R test (7.8%) are excluded from the calculation. The final R factor for all reflections is 11.97%.
[e]R_{free} as defined by Brünger (14).

(Figure 2B and C) and a substantial portion of the backbone were thus fit to multiple positions. Eleven atoms in both states were refined anisotropically with SHELXL (21), which was utilized for preliminary assignments of occupancies, followed by manual adjustment based on inspection of $2F_o - F_c$ and $F_o - F_c$ peaks. We must include multiple states for the G(5/15) in the refined model.

Several steps were taken to assure that the multiple-state model was fully justified by the data. (i) A model with only one state for each guanine (which shows the most subtle displacement between the two states), has worse refinement statistics than a two-state guanine model. The two-state guanine model gives an R/R_f of 10.29/12.49 (RMSD bonds/angles: 0.012/1.984), while the one-state guanine model gives an R/R_f of 10.47/12.62 (RMSD bonds/angles: 0.012/2.006). These one-state G models were obtained in several ways, including independent refinements starting with each member alone of the two-state model. Each of these refinements converges to the same structure, indicating that the one state model is unbiased. (ii) If one restricts the structure to one state of guanine, and two states for cytosine, then guanine forms optimal hydrogen bonds with cytosine A (2.9, 3.0, 3.0 Å), and sub-optimal hydrogen bonds with cytosine B (2.74, 3.11, 3.42 Å). (iii) Both states of guanine are evident in unbiased electron density maps.

The data here clearly indicate disorder of seven out of nine phosphate groups. The electron density of the two-state phosphate group with the least pronounced displacement is shown side by side with the electron density of a one-state phosphate group in Supplementary Data. Multiple states of phosphate groups, or lack thereof, are ambiguously indicated in the electron density.

Identification of magnesium ions

Hexahydrated magnesium complexes were added to the final model where the $2F_o - F_c$ and $F_o - F_c$ difference Fourier electron density maps satisfied the required octahedral coordination geometry of magnesium (22). The distance between any first shell water and the proximal magnesium ion is ~2.08 Å (Supplementary Table 3S). SHELXL (21) was utilized to obtain the initial estimates of occupancies of multiple overlapping magnesium complexes. The occupancy of each magnesium complex was adjusted by inspection of $2F_o - F_c$ and $F_o - F_c$ maps. Although the process of building models with multiple overlapping magnesium complexes was arduous, the final maps are clean and unambiguous (Figure 4).

RESULTS

Backbone heterogeneity

The high-resolution data here indicate that bases and backbone of the CCAGGCCTGG[0.96], along with water molecules and ions, occupy multiple discrete positions in a single crystal. A preliminary single-state model fit the diffraction data poorly. Negative difference Fourier density ($F_c - F_o$, Figure 2C) was observed on top of some atoms in the single-state model indicating excess electrons in those locations. Positive difference Fourier density ($F_o - F_c$) was observed adjacent to atoms (Figure 2C) indicating missing electrons. These difference peaks are not observed for the final multi-positional model, indicating good fit of model to data. In total, the correctness of the multi-positional model is supported by clean and continuous electron density, well-behaved thermal ellipsoids, low thermal factors (the average isotropic thermal factor; <10 for all DNA atoms) which are similar between corresponding atoms of the alternative states, real-space R-factors, and good statistics of fit including R-factor/R-free (Table 1).

CCAGGCCTGG[0.96] is seen predominantly in two states, called states A and B. The states are not fully exclusive in that for a given molecule of CCAGGCCTGG[0.96] one segment could be in State A and another segment in State B. State A is the dominant state, with 60–75% occupancy, although at the G(9)–G(10) step the populations of the two states are roughly equivalent. Covalent and non-covalent interactions allow some inference of correlations between positions of phosphates, deoxyriboses, bases, water and ions. When atoms of one nucleotide change position, the atoms of adjacent nucleotides also change, propagating deformation along the polymer backbone. For example, for nucleotide C(7), the phosphorous atom shifts from state A to state B

($\Delta\text{Å} = 2.08$ Å), along with the adjacent C5′ atom ($\Delta\text{Å} = 0.84$ Å). When these atoms of residue C(7) change position, the C3′ atom of C(6) also shifts ($\Delta\text{Å} = 2.03$), as well as all of the C(6) ribose atoms, including C5′ ($\Delta\text{Å} = 0.91$). These changes are likewise correlated with a shift in the position of the phosphorous atom of C(6) ($\Delta\text{Å} = 0.6$ Å), which, in turn, are coupled to changes in the atomic positions of G(5).

Base pair heterogeneity

All the bases of CCAGGCCTGG$^{0.96}$ form Watson–Crick base pairs, with good hydrogen bonding geometry. The two C•G pairs in the center of the duplex [base pairs G(5)•C(16) and C(6)•G(15)] each occupy two discrete positions. To the best of our knowledge, this is the first observation of full base pair positional heterogeneity observed by X-ray diffraction of DNA. Rees and coworkers have previously described positional heterogeneity of half of a base pair (i.e. of one base) (1). The two polymorphic base pairs here are related by a crystallographic 2-fold axis. Only base pair G(5)•C(16) will be described, except in relating inter-base pair parameters. To switch from G(5)A•C(16)A to G(5)B•C(16)B, (i.e. from the base pair in state A to the base pair in state B) the base pair flattens and rotates towards the major groove. The rotation axis passes approximately through the C1′ of G(5) and is nearly parallel to the helical axis. The base pair rotates as a unit, maintaining reasonable hydrogen bonding geometry. Differences in G(5)A–C(16)A and G(5)B–C(16)B are manifest by changes in essentially all interbase parameters (Supplementary Data), especially shear, stagger and buckle. Upon conversion from G(5)A–C(16)A to G(5)B–C(16)B, the shear increases by ~0.15 Å, the stagger increases by >0.2 Å and the buckle drops by ~10°. The conversion of both base pairs from state A to state B [G(5)A–C(16)A to G(5)B–C(16)B and C(6)A–G(15)A to C(6)B–G(15)B] causes changes in all inter-base pair parameters (Supplementary Data), especially shift, tilt and roll. Shift changes by 0.7 Å, tilt by 6° and roll by 10°.

Average versus discrete

The previously described CCAGGCCTGG$^{1.6}$ is fully consistent with the 1.6 Å data from which it was determined, as indicated by electron density maps and various statistics of fit. The CCAGGCCTGG$^{1.6}$ structure was determined with lower resolution data collected at higher temperature than CCAGGCCTGG$^{0.96}$. The data collected for CCAGGCCTGG$^{1.6}$ were fit to a model with a single conformational state of the DNA, plus 27 fully occupied and 15 partially occupied water molecules. CCAGGCCTGG$^{1.6}$ can be seen here to be an average structure, weighted by population, of the discrete states A and B of CCAGGCCTGG$^{0.96}$.

The averaging in CCAGGCCTGG$^{1.6}$ can be seen by a comparison of atomic positions, bond angles and helical parameters with those of the two states of CCAGGCCTGG$^{0.96}$. For example, atomic positions of CCAGGCCTGG$^{1.6}$ are population-weighted averages of those of the

discrete states of CCAGGCCTGG$^{0.96}$. Specifically, the N1 of C(6)$^{1.6}$ falls on the line between the N1 of C(6)A and the N1 of C(6)B. The N1 of C(6)$^{1.6}$ is closer to the N1 of C(6)A (0.60 Å), which is 75% populated, than to the N1 of C(6)B (1.06 Å), which is 25% populated.

Sugar puckers also reveal the averaging phenomenon. The ribose phase angles (23) of CCAGGCCTGG$^{1.6}$ are equal to the population-weighted averages of those of states A and B of CCAGGCCTGG$^{0.96}$. For example, for C(7)$^{1.6}$ the ribose phase angle is 109°, which is equivalent to the 107° population-weighted average of the phase angles of C(7)A and C(7)B of CCAGGCCTGG$^{0.96}$ ($95° \times 0.75 + 142° \times 0.25 = 107°$, where 95° and 142° are the phase angles of the two states C(7) of CCAGGCCTGG$^{0.96}$ with populations of 0.75 and 0.25, respectively).

Magnesium ions

A total of 16 fully hydrated magnesium complexes are observed per CCAGGCCTGG$^{0.96}$ duplex (i.e. there are 8 magnesium ions per asymmetric unit). As with the water molecules, closely spaced magnesium ions are interpreted as single ions in multiple competing states. None of the magnesium ions are fully occupied. Several appear by their overlap and interactions to represent alternative states for a given ion. The magnesium complexes are numbered 21–25, accompanied with a letter (a–c) to specify alternative states. Magnesium 24 occupies three overlapping sites (a, b and c) with occupancies of 35% (a), 15% (b) and 35% (c) (Figure 4). Magnesium 22 occupies two states, with occupancies of 60% (a) and 15% (b). Magnesium 21 is the most highly occupied at 90%. Magnesium 23 has 40% occupancy. Magnesium 25 occupies a crystallographic special position, where a 2-fold symmetry axis directly runs through the metal center; hence its occupancy of 40% should be doubled when considering the true molecular occupancy per DNA duplex. A summary of the magnesium complexes, including occupancies, thermal factors and coordination geometries, is given in Supplementary Table 2S.

Most DNA phosphate groups of CCAGGCCTGG$^{0.96}$ make contact with at least one magnesium–water complex. The local electrostatic environment surrounding the phosphate groups are variable. In total the magnesium complexes enable lateral packing of helices in the crystal lattices through electrostatics and hydrogen bonding. The interactions of magnesium complexes with the DNA backbone involve hydrogen bonds between first shell water molecules and oxygen atoms in the sugar/phosphate backbones, O1P, O2P, O3′ or O4′. Interactions with O5′ atoms are not observed. In some cases, both of non-bridging phosphate oxygens (O1P, O2P) of a phosphate interact with first shell water molecules of a common magnesium ion to form a six-membered ring system

The positional heterogeneity of at least some of the magnesium complexes appears to be related directly to heterogeneity of the nearby DNA (24). The polymorphism of the central GGCC region of the DNA appears to be linked to that of magnesium complexes 23(a/b) and 24(a/b). Magnesium 24a forms hydrogen bonds within

the minor groove predominantly with the A state of the DNA, with G(15)A and the O4′ of the C(16)A, N3 and O4′ of G(15)A/B and the N2 atoms of G(14). Magnesium 24a also forms a hydrogen bond with the O6 of G(5)A of a symmetry related duplex. The alternate state of this magnesium ion (magnesium 24b) forms hydrogen bonds with the N2 of G(14), O2 of C7 and O4′ of T(8). Magnesium 21 acts as an anchor for both states, forming hydrogen bonds (through first shell water molecules) with the phosphates of C(6)A/B and C(7)A/B.

Hydration

A network of 64 fully occupied and 122 partially occupied water molecules per asymmetric unit fill much of the available solvent region. The aggregate water molecule population, summing over all occupancies, is 106 water molecules per duplex, distributed over 86 sites. The majority of the water molecules (118/131) associated with CCAGGCCTGG$^{0.96}$ are partially occupied. Pairs of closely spaced, partially occupied water molecules are interpreted here as single water molecules in multiple competing states, reducing the number of unique water molecules from 118 to 88.

The terminal base pairs of CCAGGCCTGG$^{0.96}$ interact with localized water molecules, which form the 'ribbon of hydration' in the minor groove noted previously by Dickerson (25,26). The hydrogen bonding interactions of these water molecules, which are primarily located within the planes of the base pairs, involve the N3 and O2 and an O4′ of the terminal C(1)–G(20) base pair, the O2 of C(2), and N3 of both A(3) and G(4). In the vicinity of the third base pair from the terminus, this pattern of the minor groove hydration is interrupted by magnesium 25.

Three-center (bifurcated) hydrogen bonds

Bifurcated H-bonds, more correctly called three-centered hydrogen bonds (27,28), were previously proposed to link adjacent base pairs in A-tract DNA, within the major groove (29,30). However, more recently it was suggested from analysis of a high-resolution structures that these intra-base pair hydrogen bonds are not significant, and that the relevant interactions are in fact limited to conventional Watson–Crick hydrogen bonds (31). Here the geometry of the central CpC step of CCAGGCCTGG$^{1.6}$ was interpreted as providing evidence of weak bifurcated hydrogen bonds between adjacent base pairs. However, our interpretation of CCAGGCCTGG$^{0.96}$ is that the hydrogen bonds are limited to the conventional Watson–Crick type. Specifically, when hydrogen atoms are placed appropriately on the 4-amino groups of the cytosines of both states A and B (see Supplementary Data), the relevant N–H–O6(G) distances are at least 2.9 Å, which is beyond the limit of what is generally accepted as hydrogen bonding (32).

DISCUSSION

The structure of CCAGGCCTGG$^{0.96}$ described here contains B-DNA, water molecules and counterions, all in multiple states (Figures 1, 2 and 4). The two central

C•G base pairs of CCAGGCCTGG$^{0.96}$, at a $^{5'}$GpC$^{3'}$ step, occupy multiple positions (Figure 2C). The positions are separated by atomic translations of up to 1.8 Å (Figure 3). Each position is fractionally occupied. To the best of our knowledge, this structure represents the most extreme base pair positional polymorphism observed thus far by X-ray diffraction of DNA. Additional states with low populations cannot be excluded.

We propose that positional polymorphism, including that of DNA bases, is a general property of B-form DNA structures that will be observed with high frequency in high-resolution structures. We have examined the small number of high-resolution structures of B-DNA currently available in the database. Most contain bases in multiple positions. For example, Rees and coworkers (1) described positional polymorphism of individual B-DNA bases. Their 0.74 Å resolution structure showed heterogeneity of 6 out of 9 phosphate groups, 1 out of 10 bases (but not an entire base pair), 2 out of 4 calcium ions and many water molecules. Similarly, Dervan (4) reported base pair positional polymorphism, albeit with relatively subtle atomic translations of <0.6 Å. We have determined that at least one of the high-resolution B-DNA structures that does not contain bases in multiple positions requires re-refinement and revision to incorporate the base pair positional polymorphism indicated by the data (see below). Combined, these results suggest that bases and intact base pairs of B-DNA are very frequently polymorphic in crystal structures, even though the polymorphism is not evident in lower resolution structures.

Observed positional polymorphism in X-ray structures implies polymorphism in solution, which is consistent with results from vibrational spectroscopy studies that have revealed sugar pucker polymorphism within double stranded of poly(dA–dT) and poly(dA)•poly(dT) (33). B-DNA appears to be more polymorphic in crystals and in solution than other helical conformations. Evidence that A-helices and especially Z-helices (34,35), are less polymorphic in solution than B-DNA, is given by their propensity to readily form well-ordered crystals, with low conformational heterogeneity. We are unaware of recent efforts to address the nature and degree of these differences in solution.

Thus far, only C•G base pairs have been observed in multiple positions within B-form DNA crystal structures. The polymorphism reported here is at a $^{5'}$GpC$^{3'}$ step, which is not expected to be the most dynamic in solution, based on the predictions of Olson and Zhurkin (36). It will be interesting to see, as the number of high-resolution structures grows, if the observed examples of base pair positional polymorphism grow to include all base pair steps, and if the frequencies of polymorphisms prove to be consistent with the predictions of Olson and Zhurkin. It may be that A•T base pairs, especially those in A-tracts, are less apt to adopt multiple positions compared to C•G base pairs. This prediction is based on a putative link observed in the present study between the polymorphism of DNA conformation and polymorphism of groove hydration. The hydration of C•G base pairs is intrinsically more polymorphic and dynamic than that of A•T base pairs (37,38). The

potential energy wells for water molecules adjacent C–G base pairs can be broad and shallow. In contrast, water molecules in the minor groove associated with contiguous A•T base pairs tend to be localized in well-defined potential energy wells, giving highly ordered hydration arrays.

The original CCAGGCCTGG[1.6] structure suggested homogeneity of both the DNA and the surrounding hydration environment. In fact, the heterogeneity is seen to be extensive, as indicated by high-resolution data here, including multiple positions of 14 of 18 phosphate groups (per duplex), 2 of 10 bp, all of the observed counterions and most of the water molecules. Visualization and accurate modeling of such a polymorphic structure was enabled by high resolution and high quality data collected at low temperature, along with iterative manual model building.

CCAGGCCTGG[0.96] was fit to data collected at low temperature, where interconversion between conformers is slow on the timescale of data collection. The electron density of the backbone shows discrete peaks of electron density (Supplementary Data). The CCAGGCCTGG[0.96] maps and conformation suggest discrete conformers, separated by an energetic barrier, rather than isoenergetic continuum of intermediate states. CCAGGCCTGG[1.6], by contrast, was fit to data collected at room temperature, where interconversion between conformers is fast on the timescale of data collection. Therefore atomic positions there are averaged by both crystallographic disorder and by thermal fluctuations.

Currently, six B-form DNA structures are available from data that extends beyond 1.0 Å. None of these show full base pair positional heterogeneity to the extent observed here, although single-base positional heterogeneity is common. We considered the possibility that the heterogeneity observed in CCAGGCCTGG[0.96] is anomalous since it is more extensive than that of previous structures. However, given that for at least one published structure the actual structural heterogeneity is greater than originally described, it is possible that other crystal structures represent averaged-out base pair positions. To test this possibility, we extracted the coordinates and data for the high-resolution structure CCAGCGCTGG[0.99] (PDB entry 1EN9, resolution 0.99 Å). The published structure of CCAGCGCTGG[0.99] contains bases and base pairs each at single positions. By re-refinement, we determined that the data gives a better fit to a model with multiple base pair positions, at the central $^{5'}CpG^{3'}$ step, than to the published single base pair positional model. The electron density maps of the single state model clearly show the excess $F_o - F_c$ density surrounding the central C–G pairs, which can be fit to multiple positions. This conclusion is supported by statistical indicators such as the real-space *R*-factor showing sudden spike of the value at the central G(6) of the published structure. Therefore, the correct model of CCAGCGCTGG[0.99] contains multiple positions of the central base pairs, similar to the multiple states of the central base pairs of CCAGGCCTGG[0.96]. Crystallization of DNA can select a small subset of solution conformers, none of which are necessarily among the most highly populated in

solution. Since both CCAGCGCTGG[0.99] and CCAGGCC TGG[0.96] were crystallized under similar conditions, yielding nearly isomorphous structures, it is assumed that lattice effects (39) are the same.

The observation of multiple positions of atoms in a X-ray structure indicates heterogeneity from site to site in the crystal or possibly thermal transitions between states during data acquisition. The averaging of discrete positions is a well-known phenomenon of the crystallographic method. At modest resolutions, only averages of discrete states are 'observed' in the electron density. When atoms are in multiple states, higher resolution (higher diffraction angle) and lower temperature structures can show discrete positions. For heterogeneous structures, the average position is often incorrect at some level. It can fall directly on an energy barrier, between the minima.

The CCAGGCCTGG[1.6] structure was determined with technology that has been vastly improved over the last 20 years. CCAGGCCTGG[1.6] contains only a single conformational state of DNA and single positions for most water molecules. CCAGGCCTGG[1.6] is an average over the states of CCAGGCCTGG[0.96], which contains multiple discrete states of DNA and its saline environment. In particular, the positions of the base pairs at the center of CCAGGCCTGG[1.6] are now seen to be population-weighted averages of two discrete positions of CCAGGCCTGG[0.96]. The combined results suggest that multiple positions of base pairs might be a common but undetected property of DNA.

High resolution also reveals a more complex and subtle milieu of magnesium ions than anticipated from lower resolution analysis of DNA. None of the magnesium ions associated with CCAGGCCTGG[0.96] are fully occupied. There are 10 partially occupied magnesium ions per duplex, in some cases with significant positional overlap (Figure 4). The estimated occupancies range from 15% to 90%. For a given region, the sum of the partial occupancies is always <100%. At lower resolution dodecamer structures (8,10), the magnesium ions appeared to be more highly localized. There are fewer of them, but at higher occupancies. Again, the accumulation of additional high-resolution structures will help establish which models best describe the interactions of magnesium ions with DNA.

In the current structure, there are a total of 10 partially occupied magnesium ions, contributing a total charge of +13.2 C per two unit cells (i.e. the biological unit of one DNA duplex). There are −18 C arising from 18 phosphate groups of the DNA duplex. The current model thus lacks cations to account for the remaining 4.8 C required for charge neutrality. Sodium and magnesium were the only cations in the crystallization solution. It is likely that much of the unobserved cationic charge arises from sodium ions that are not included in the model. Magnesium ions are readily identifiable by coordination geometry (40). In contrast, a partially occupied sodium ion is nearly indistinguishable from water in electron density maps due to similarities in number of electrons and ionic/molecular radii between Na^+ and water (1.4 Å for oxygen versus 0.95 Å for Na^+); mixed occupancy between Na^+ and water further obscures the difference

in radii (41). In subsequent work, the locations of mono-valent cations with distinctive scattering characteristics, namely Thallium (I) and Rubidium (I) ions, will be described in ultrahigh resolution structures, with the aim of fully characterizing the electrostatic environment surrounding DNA.

ACCESSION NUMBER

PDB ID 3GGB.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–6, Supplementary Discussion of the BI–BII conformation, Supplementary Tables 1–3 and Supplementary References [1,2,23,42–47].

REFERENCES

1. Kielkopf,C.L., Ding,S., Kuhn,P. and Reese,D.C. (2000) Conformational Flexibility of B-DNA at 0.74 Angstrom Resolution: d(CCAGTACTGG)$_2$. *J. Mol. Biol.*, **296**, 787–801.
2. Chiu,T.K. and Dickerson,R.E. (2000) 1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J. Mol. Biol.*, **301**, 915–945.
3. Soler-Lopez,M., Malinina,L., Liu,J., Huynh-Dinh,T. and Subirana,J.A. (1999) Water and ions in a high resolution structure of B-DNA. *J. Biol. Chem.*, **274**, 23683–23686.
4. Chenoweth,D.M. and Dervan,P.B. (2009) Allosteric modulation of DNA by small molecules. *Proc. Natl Acad. Sci. USA*, **106**, 13175–13179.
5. Heinemann,U. and Alings,C. (1989) Crystallographic study of one turn of G/C-rich B-DNA. *J. Mol. Biol.*, **210**, 369–381.
6. Drew,H.R., Wing,R.M., Takano,T., Broka,C., Itakura,K. and Dickerson,R.E. (1981) Structure of a B-DNA dodecamer. Conformation and dynamics. *Proc. Natl Acad. Sci. USA*, **78**, 2179–2183.
7. Sussman,J.L., Holbrook,S.R., Warrant,R.W., Church,G.M. and Kim,S.-H. (1978) Crystal structure of yeast phenylalanine tRNA. I. Crystallographic refinement. *J. Mol. Biol.*, **123**, 607.
8. Sines,C.C., McFail-Isom,L., Howerton,S.B., VanDerveer,D. and Williams,L.D. (2000) Cations mediate B-DNA conformational heterogeneity. *J. Am. Chem. Soc.*, **122**, 11048–11056.
9. Shui,X., Sines,C., McFail-Isom,L., VanDerveer,D. and Williams,L.D. (1998) Structure of the potassium form of CGCGA ATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry*, **37**, 16877–16887.
10. Shui,X., McFail-Isom,L., Hu,G.G. and Williams,L.D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, **37**, 8341–8355.
11. McFail-Isom,L., Sines,C.C. and Williams,L.D. (1999) DNA structure: cations in charge? *Curr. Opin. Struct. Biol.*, **9**, 298–304.
12. Hud,N.V. and Plavec,J. (2003) A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*, **69**, 144–158.
13. Otwinowski,Z. and Minor,W. (1997) In: Carter,J.C.W. and Sweet,R.M. (eds), *Methods in Enzymol., Macromolecular Crystallography*, Vol. 276. Part A. Academic Press, New York, pp. 307–326.
14. Brunger,A.T. (1992) Free R-value - a novel statistical quantity for assessing the accuracy of crystal-structures. *Nature*, **355**, 472–475.
15. Brunger,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) Crystallography & Nmr system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D-Biol. Crystallogr.*, **54**, 905–921.
16. Gelbin,A., Schneider,B., Clowney,L., Hsieh,S.-H., Olson,W.K. and Berman,H.M. (1996) Geometric parameters in nucleic acids: sugar and phosphate constituents. *J. Am. Chem. Soc.*, **118**, 519–529.
17. Clowney,L., Jain,S.C., Srinivasan,A.R., Westbrook,J., Olson,W.K. and Berman,H.M. (1996) Geometric parameters in nucleic acids: nitrogenous bases. *J. Am. Chem. Soc.*, **118**, 509–518.
18. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect D: Biol. Crystallogr.*, **53**, 240–255.
19. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect D: Biol. Crystallogr.*, **60**, 2126–2132.
20. Farrugia,L.J. (1997) Ortep-3 for Windows. *J. Appl. Cryst.*, **30**, 565.
21. Sheldrick,G.M. (2008) A short history of SHELX. *Acta Crystallogr., Sect. A: Found. Crystallogr.*, **64**, 112–122.
22. Brown,I.D. (1988) What factors determine cation coordination numbers. *Acta Crystallogr. Sect. B.*, **44**, 545–553.
23. Altona,C. and Sundaralingam,M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.*, **94**, 8205–8212.
24. Hud,N.V. and Polak,M. (2001) DNA-cation interactions: the major and minor grooves are flexible ionophores. *Curr. Opin. Struct. Biol.*, **11**, 293–301.
25. Baikalov,I., Grzeskowiak,K., Yanagi,K., Quintana,J. and Dickerson,R.E. (1993) The crystal structure of the trigonal decamer C-G-A-T-C-G-6mea-T-C-G: A B-DNA helix with 10.6 base-pairs per turn. *J. Mol. Biol.*, **231**, 768–784.
26. Heinemann,U. and Hahn,M. (1992) CCAGGCm5CTGG. Helical fine structure, hydration, and comparison with CCAGGCCTGG. *J. Biol. Chem.*, **267**, 7332–7341.
27. Taylor,R., Kennard,O. and Versichel,W. (1983) Geometry of the N-H●O = C hydrogen bond. 1. Lone pair directionality. *J. Am. Chem. Soc.*, **105**, 5761–5766.
28. Jeffrey,G.A. and Mitra,J. (1984) 3-Center (Bifurcated) Hydrogen-bonding in the crystal-structures of amino-acids. *J. Am. Chem. Soc.*, **106**, 5546–5553.
29. Coll,M., Frederick,C.A., Wang,A.H.-J. and Rich,A. (1987) A bifurcated hydrogen-bonded conformation in the d(A-T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl Acad. Sci. USA*, **84**, 8385–8389.
30. Nelson,H.C.M., Finch,J.T., Luisi,B.F. and Klug,A. (1987) The structure of an Oligo(dA)●Oligo(dT) tract and its biological implications. *Nature*, **330**, 221–226.
31. Woods,K.K., Maehigashi,T., Howerton,S.B., Sines,C.C., Tannenbaum,S. and Williams,L.D. (2004) High-resolution structure of an extended A-tract: [d(CGCAAATTTGCG)]$_2$. *J. Am. Chem. Soc.*, **126**, 15330–15331.
32. Jeffrey,G.A. (1997) *An Introduction to Hydrogen Bonding*. Oxford University Press, New York.
33. Brahms,S., Fritsch,V., Brahms,J.G. and Westhof,E. (1992) Investigations on the dynamic structures of adenine- and thymine-containing DNA. *J. Mol. Biol.*, **223**, 455–476.

34. Brzezinski,K., Brzuszkiewicz,A., Dauter,M., Kubicki,M., Jaskolski,M. and Dauter,Z. (2011) High regularity of Z-DNA revealed by ultra high-resolution crystal structure at 0.55 A. *Nucleic Acids Res.*, **1**, 1.

35. Wang,A.H., Quigley,G.J., Kolpak,F.J., Crawford,J.L., van Boom,J.H., van der Marel,G. and Rich,A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, **282**, 680–686.

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

37. Watkins,D., Mohan,S., Koudelka,G.B. and Williams,L.D. (2010) Sequence recognition of DNA by protein-induced conformational transitions. *J. Mol. Biol.*, **396**, 1145–1164.

38. Watkins,D., Hsiao,C., Woods,K.K., Koudelka,G.B. and Williams,L.D. (2008) P22 C2 repressor-operator complex: mechanisms of direct and indirect readout. *Biochemistry*, **47**, 2325–2338.

39. Dickerson,R.E., Goodsell,D.S. and Neidle,S. (1994) "...The Tyranny of the Lattice...". *Proc. Natl Acad. Sci. USA*, **91**, 3579–3583.

40. Phillips,K., Dauter,Z., Murchie,A.I.H., Lilley,D.M.J. and Luisi,B. (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 angstrom resolution. *J. Mol. Biol.*, **273**, 171–182.

41. Williams,L.D. (2005) *DNA Binders and Related Subjects*, Vol. 253. Springer, Berlin, pp. 77–88.

42. Fratini,A.V., Kopka,M.L., Drew,H.R. and Dickerson,R.E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATT^{Br}CGCG. *J. Biol. Chem.*, **257**, 14686–14707.

43. Hartmann,B., Piazzola,D. and Lavery,R. (1993) BI-BII transitions in B-DNA. *Nucleic Acids Res.*, **21**, 561–568.

44. Stofer,E. and Lavery,R. (1994) Measuring the geometry of DNA grooves. *Biopolymers*, **34**, 337–346.

45. Soler-Lopez,M., Malinina,L. and Subirana,J.A. (2000) Solvent organization in an oligonucleotide crystal. The structure of d(GC GAATTCG)$_2$ at atomic resolution. *J. Biol. Chem.*, **275**, 23034–23044.

46. Egli,M., Tereshko,V., Teplova,M., Minasov,G., Joachimiak,A., Sanishvili,R., Weeks,C.M., Miller,R., Maier,M.A., An,H.Y. *et al.* (1998) X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers*, **48**, 234–252.

47. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.