The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium † , ‡

ARTICLE *in* BIOCHEMISTRY · JULY 1998

Impact Factor: 3.02 · DOI: 10.1021/bi973073c · Source: PubMed

4 AUTHORS, INCLUDING:

Lori McFail-Isom
University of Central Arkansas
15 PUBLICATIONS **1,404** CITATIONS
SEE PROFILE

Loren Dean Williams
Georgia Institute of Technology
110 PUBLICATIONS **5,022** CITATIONS
SEE PROFILE

The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium[†,‡]

Xiuqi Shui, Lori McFail-Isom, Gary G. Hu, and Loren Dean Williams*

School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400

Received December 15, 1997; Revised Manuscript Received March 11, 1998

ABSTRACT: We describe a very accurate addition (called structure X here) to the B-DNA dodecamer family of X-ray structures. Our results confirm the observation of Drew and Dickerson [(1981) *J. Mol. Biol. 151*, 535−556] that the spine of hydration in AT tract DNA is two layers deep. However, our results suggest that the primary spine is partially occupied by sodium ions. We suggest that many sequence-dependent features of DNA conformation are mediated by site specific binding of cations. For example, preferential localization of cations, as described here within the minor groove of structure X, is probably the structural origin of AT tract bending and groove narrowing. The secondary spine, which does not interact directly with the DNA, is as geometrically regular as the primary spine, providing a model for transmission of sequence information into solvent regions. A fully hydrated magnesium ion located in the major groove of structure X appears to pull cytosine bases partially out from the helical stack, exposing π-systems to partial positive charges of the magnesium ion and its outer sphere. A partially ordered spermine molecule is located within the major groove of structure X. Dodecamer structures are derived from crystals of $[d(CGCGAATTCGCG)]_2$ in space group $P2_12_12_1$ (a = 25 Å, b = 40 Å, and c = 66 Å). On average, those crystals diffracted to around 2.5 Å resolution with 2500 unique reflections. Structure X, with the same space group, DNA sequence, and crystal form as the "Dickerson dodecamer", is refined against a complete, low-temperature, 1.4 Å resolution data set, with over 11000 reflections. Structure X appears to be conformationally more ordered than previous structures, suggesting that at least a portion of the conformational heterogeneity previously attributed to DNA sequence in fact arises from experimental error.

Our current image of DNA as a conformationally polymorphic molecule is derived from seminal work of Dickerson and co-workers, who used single-crystal X-ray diffraction to "observe" molecular structures of DNA dodecamers (*1*−*8*). Their observations led to the current consensus that sequence and environment influence DNA conformation, which in turn modulates interactions of ions, small molecules, and proteins. An appreciation of DNA conformational heterogeneity is a prerequisite for rational design of therapeutically and technologically useful DNA-binding agents. Sixty-eight isomorphous members of the $[d(CGCXAATTYGCG)]_2$ (X = G or A, Y = C or T) dodecamer family (Figure 1) are presently found in the Brookhaven Protein Data Bank (PDB) (*9*) and Nucleic Acid Data Bank (NDB) (*10*). Unfortunately, all previous dodecamer crystals diffracted rather poorly, to 2.0−3.0 Å, averaging around 2500 unique reflections.

Many generally accepted conclusions about DNA conformation are derived from the original "Dickerson dodecamer" [structure 1BNA; DNA duplex $[d(CGCGAATTCGCG)]_2$]. Here we re-evaluate, modify, and extend some of those conclusions. Our re-analysis is made possible by a new determination (structure X) of the dodecamer structure.

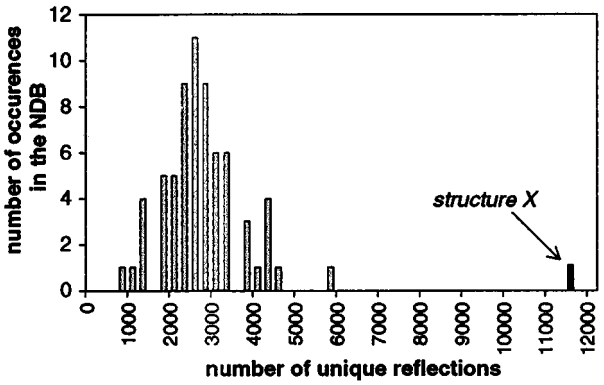

FIGURE 1: Previous dodecamer structures were derived from diffraction data averaging around 2.5 Å resolution. Shown here is a histogram of the number of reflections reported vs the number of NDB entries for 68 isomorphous structures of $[d(CGCXAATTYGCG)]_2$ (X = G or A, Y = C or T).

Structure X is of much higher resolution and greater accuracy than previous dodecamer structures. Structure X is refined against 1.4 Å resolution data, with 11 438 unique reflections (Figure 1). This high-resolution data set has allowed us to determine the three-dimensional structure of the dodecamer and its associated solvent molecules and ions much more confidently and in greater detail than was possible previously.

The hydration of structure X reveals several surprises. The primary spine of hydration appears to be composed partially of sodium ions. Our crystallographic data are consistent with partial occupancy of the primary spine solvent sites by both

[†] This work was funded by the National Science Foundation (Grant MCB-9056300), the American Cancer Society (Grant RPG-95-116-03-GMC), and the Georgia Tech/Medical College of Georgia Program, Georgia Tech/Emory Program.

[‡] Atomic coordinates and structure factors of structure X have been deposited in the NDB (entry code BDL084) and PDB.

* To whom correspondence should be addressed.

S0006-2960(97)03073-0 CCC: $15.00
Published on Web 05/22/1998

water molecules and sodium ions. In addition, a regular and highly ordered secondary spine rides piggyback on the primary spine. The pattern of interaction follows a pyramid scheme, with the primary sodium level forming a template for the secondary water level, which does not interact directly with the DNA.

In addition, structure X provides the first high-resolution view of spermine bound to B-DNA. A partially ordered spermine molecule is located within the major groove. A fully hydrated magnesium ion is also located in the major groove.

Some of the conformational heterogeneity previously attributed to DNA sequence arises in part from experimental error. Our comparison with previous structures suggests that variation of many helical parameters of previous dodecamers is poorly determined. Structure X appears to be conformationally more ordered than previous structures. In addition, contributions of minor chemical modifications to the variation of the conformation of those low-resolution structures do not seem to be significant relative to contributions from experimental error.

EXPERIMENTAL PROCEDURES

Crystallization. The ammonium salt of reverse-phase HPLC-purified d(CGCGAATTCGCG) was purchased from the Midland Certified Reagent Company (Midland, TX). Crystals were grown at 25 °C in a constant-temperature incubator from sitting drops that initially contained 1.8 mM d(CGCGAATTCGCG), 31 mM sodium cacodylate (pH 6.5), 9.6 mM $MgCl_2$, 7% 2-methyl-2,4-pentanediol (MPD), 38 mM spermine tetrahydrochloride, and a 1.0 mM putative minor groove binder [*N*,*N*′-bis(4-aminobutyl)-2,7-anthraquinonedisulfonanamide dihydrochloride; compound **6** in ref *11*]. The drops were equilibrated by vapor diffusion against a reservoir of 30% MPD. Orthorhombic ($P2_12_12_1$) crystals appeared within 20 days and grew to 0.6 × 0.4 × 0.4 mm over several months. The minor groove binder is not observed in the electron density maps (there are no disordered regions), indicating it did not enter the crystal with high occupancy. Sixty minutes before mounting, the concentration of MPD in the mother liquor was increased to approximately 35%, after which a crystal was plucked up in a loop and transferred directly into a −137 °C N_2 stream bathing the goniostat.

Data Collection and Reduction. X-ray intensity data were collected at a low temperature (−137 °C) with multiwire detectors (ADSC). Copper Kα radiation ($\lambda = 1.54$ Å) was generated with a fine-focus Rigaku RU200 rotating anode and monochromated with a graphite grating. Data were merged and reduced with ADSC software.

Refinement. Except for solvent molecules, coordinates of duplex [d(CGCGAATTCGCG)]$_2$], (PDB entry 227D; *12*) were used for the starting model. The structure was annealed and refined with XPLOR (version 3.851, *13*) which was also used to calculate sum ($2F_o - F_c$) and difference ($F_o - F_c$) Fourier maps. Structure X was refined against all data between 10 and 1.4 Å resolution (11 438 unique reflections). Data collection and refinement statistics are given in Table 1. $|F_o|$ values were not scaled to $|F_c|$ by any means other than a single isotropic scale factor. No attempt was made to artificially lower the *R*-factor.

Table 1: Crystallographic and Refinement Statistics for Structure X

unit cell dimensions	$\alpha = 90°$, $a = 25.186$ Å $\beta = 90°$, $b = 40.208$ Å $\gamma = 90°$, $c = 65.656$ Å
DNA (asymmetric unit)	[d(CGCGAATTCGCG)]$_2$
space group	$P2_12_12_1$
temperature of data collection (°C)	−137
number of reflections	58 307
number of unique reflections	11 594
number of reflections used in refinement	11 438[a]
R-merge[b] by shell	
resolution range (Å)	
2.35−10.0	3.5
1.87−2.35	4.9
1.63−1.87	6.6
1.48−1.63	8.0
1.38−1.48	11.7
1.38−10.0 (overall)	4.4
maximum resolution of observed reflections (Å)	1.38
maximum resolution of refinement (Å)	1.40
number of DNA atoms	486
number of full occupancy water molecules	142
number of half-occupancy water molecules	18
number of spermine atoms	7
number of magnesium ions plus coordinating water molecules	7
rms deviation of DNA bonds from ideality (Å)	0.007
rms deviation of DNA angles from ideality (deg)	1.42
R-factor (%)	19.7
R-free (%)	21.6
estimated coordinate error (Å)	0.14[c]/0.17[d]

Number of Reflections, *R*-factor, and Completeness by Resolution

resolution range (Å)	number of reflections	*R*-factor (%)	completeness[c] (%)
2.78−10.00	1815	17.66	100
2.22−2.78	1713	18.95	98
1.94−2.22	1637	19.26	96
1.76−1.94	1568	20.28	93
1.64−1.76	1361	21.35	80
1.54−1.64	1164	23.92	70
1.46−1.54	1128	25.88	67
1.40−1.46	1052	28.55	64

[a] Systematic absences and reflections with a resolution of >10 Å were deleted. No σ cutoff was applied. [b] *R*-merge = ∑[abs(ave − obs)]/∑(ave). [c] Estimated by the method of Read (*69*). A σ graph is shown in panel A of Figure 1S in the Supporting Information. [d] Estimated by the method of Luzzati (*24*). A Luzzati graph is shown in panel B of Figure 1S in the Supporting Information.

During the course of the refinement against parameters of the dictionary of param8.dna, incorrect ribose bond angles within the dictionary pulled the model systematically away from the best fit to the data. Similarly, the data pulled the model systematically away from the dictionary. The final model produced by that refinement (*structure W*) is characterized by relatively high rms deviations of angles from those of param8.dna (3.18°) but lower deviations from those of dna-rna-multi-endo.param (2.08°). The parameter file dna-rna-multi-endo.param includes improved distances and angles recently described by Berman and co-workers (*14*). Thus, *structure W* fits the improved dictionary best, even though it was not refined against that dictionary. When the refinement was conducted with the improved dictionary of dna-rna-multi-endo.param to produce structure X, the systematic discrepancies between model and dictionary disappeared entirely. The ability of the high-resolution data to detect such dictionary errors gives us considerable confidence

Table 2: Structure X Backbone Torsion Angles α−ζ, Glycosyl Angles χ, Pseudorotation Phase Angles and Amplitudes, and Sugar Puckers[a]

residue	χ (deg)	γ (deg)	δ (deg)	ϵ (deg)	ζ (deg)	α (deg)	β (deg)	phase (deg)	amplitude	pucker
C(1)	−106	−70	145	−172	−98	−	−	164	35	C2′-endo
G(2)	−85	43	148	−151	−157	−70	−172	160	38	C2′-endo
C(3)	−132	50	93	−165	−81	−39	131	65	34	C4′-exo
G(4)	−94	50	145	−167	−145	−65	174	157	38	C2′-endo
A(5)	−119	54	133	−178	−91	−47	156	158	32	C2′-endo
A(6)	−109	46	128	177	−96	−64	−173	153	31	C2′-endo
T(7)	−120	49	113	−177	−96	−49	173	126	35	C1′-exo
T(8)	−120	53	114	171	−95	−54	168	128	37	C1′-exo
C(9)	−112	51	138	−156	−96	−53	−173	158	33	C2′-endo
G(10)	−84	40	143	−100	146	−60	163	147	45	C2′-endo
C(11)	−113	51	144	−164	−126	−73	144	164	37	C2′-endo
G(12)	−79	−66	148	−	−	53	144	207	26	C3′-exo
G(24)	−141	53	88	−	−	−62	170	13	35	C3′-endo
C(23)	−150	51	84	−161	−78	−57	130	17	37	C3′-endo
G(22)	−86	45	144	−149	−173	−67	177	152	41	C2′-endo
C(21)	−124	51	96	−170	−81	−58	160	94	41	O1′-endo
T(20)	−110	43	135	−164	−107	−45	172	152	38	C2′-endo
T(19)	−116	52	128	−173	−107	−49	178	143	35	C1′-exo
A(18)	−112	45	113	166	−91	−61	176	128	33	C1′-exo
A(17)	−108	54	146	−178	−92	−56	−160	173	32	C2′-endo
G(16)	−102	62	141	151	−90	−62	175	169	31	C2′-endo
C(15)	−137	57	82	−174	−85	−57	160	42	38	C4′-exo
G(14)	−110	49	115	178	−94	−58	167	130	34	C1′-exo
C(13)	−107	59	138	−169	−105	−	−	164	31	C2′-endo

[a] Backbone torsion angles are O3′−P-α-O5′-*β*-C5′-*γ*-C4′-*δ*-C3′-*ϵ*-O3′-ζ-P−O5′.

in the quality of our data, in our final structure, and in the improved dictionary. A comparison of the models (*structures W* and *X*) produced by the different dictionaries demonstrates that incorrect sugar angles within a dictionary are propagated to incorrect sugar puckers, and to incorrect propeller twists and buckles. The deviations from ideality in Table 1 are from the parameters of dna-rna-multi-endo.param.

The refinement was completed by addition of solvent molecules to corresponding sum and difference peaks. Ten percent of the reflections were not used during refinement (were sequestered for *R*-free) but were used to refine the final model, calculate the final maps, and determine the final statistics. Overall, 85% of predicted unique reflections were observed. Water molecules show good hydrogen bonding geometry to neighboring molecules. Fewer than 10 of the 160 water molecules are over 3.6 Å from any hydrogen bonding partner. The most intense electron density peak in the solvent region is centered within an octahedron of six other peaks. The intense central peak was assigned as a magnesium ion, and the six surrounding peaks were assigned as the first magnesium hydration sphere. A tube of electron density was apparent in the initial electron density maps and throughout the refinement. This region of the map was left undisturbed until the final round of refinement, at which time it was fitted with seven atoms of a spermine molecule. The invariance of final sum and refined omit maps confirms the placement of the spermine atoms. rms deviations of atomic positions, Luzzati, and σ plots (Figure 1S and 2S of the Supporting Information) were calculated with XPLOR. Helical parameters were calculated with CURVES (*15*).

RESULTS

The new B-DNA structure, termed structure X here, diffracts to high angle and is refined against 1.4 Å data, yielding electron density maps of extremely high quality (Figures 2 and 3). Phosphodiester backbone torsion angles, glycosyl angles, and sugar puckers for structure X are given in Table 2. Helical parameters are given in Table 3.

Conformation and Hydration of Structure X. As summarized in a series of recent reviews (*16*−*19*), many aspects of the conformation and interactions of the original dodecamer (structure 1BNA; *1*−*3*, *6*, *8*) are presently thought to be universal features of B-DNA. Our comparison of structure X to structure 1BNA supports many previous conclusions but indicates extension or re-evaluation of some features is in order.

Is the minor groove of the AT tract narrow and the minor groove of the GC tract wide? Structure X is consistent with the previous appraisal of minor groove widths; the effect of base composition on groove width is clear. On average, the minor groove of structure X is narrower than that of structure 1BNA (Table 4). In comparison to structure 1BNA, the widest GC tract minor groove of structure X is expanded while the AT tract is compacted. Is the AT tract straight? Structure X is fully consistent with this previous appraisal. Is one of the GC−AT junctions bent and the other straight? Again, structure X is fully consistent with the previous appraisal. Structure X manifests the previously observed bend, localized near one AT−GC junction.

Is the spine of hydration composed exclusively of water molecules? Does sodium penetrate the spine? Structure X confirms many previously described features of the spine of hydration (Figure 4 and Table 5) but indicates that the previous model for minor groove hydration may need revision. In structure 1BNA, each spine molecule is described as four-coordinate, interacting with two or three sites on the DNA and one or two water molecules (*2*, *6*). These spine molecules invariably interact with edges of two bases, on the floor of the minor groove. By our geometric criteria, some of these spine molecules engage in additional interactions, with the O4′ of deoxyriboses. In comparison with structure 1BNA, the primary spine molecules of structure X interact more intimately with the DNA, with

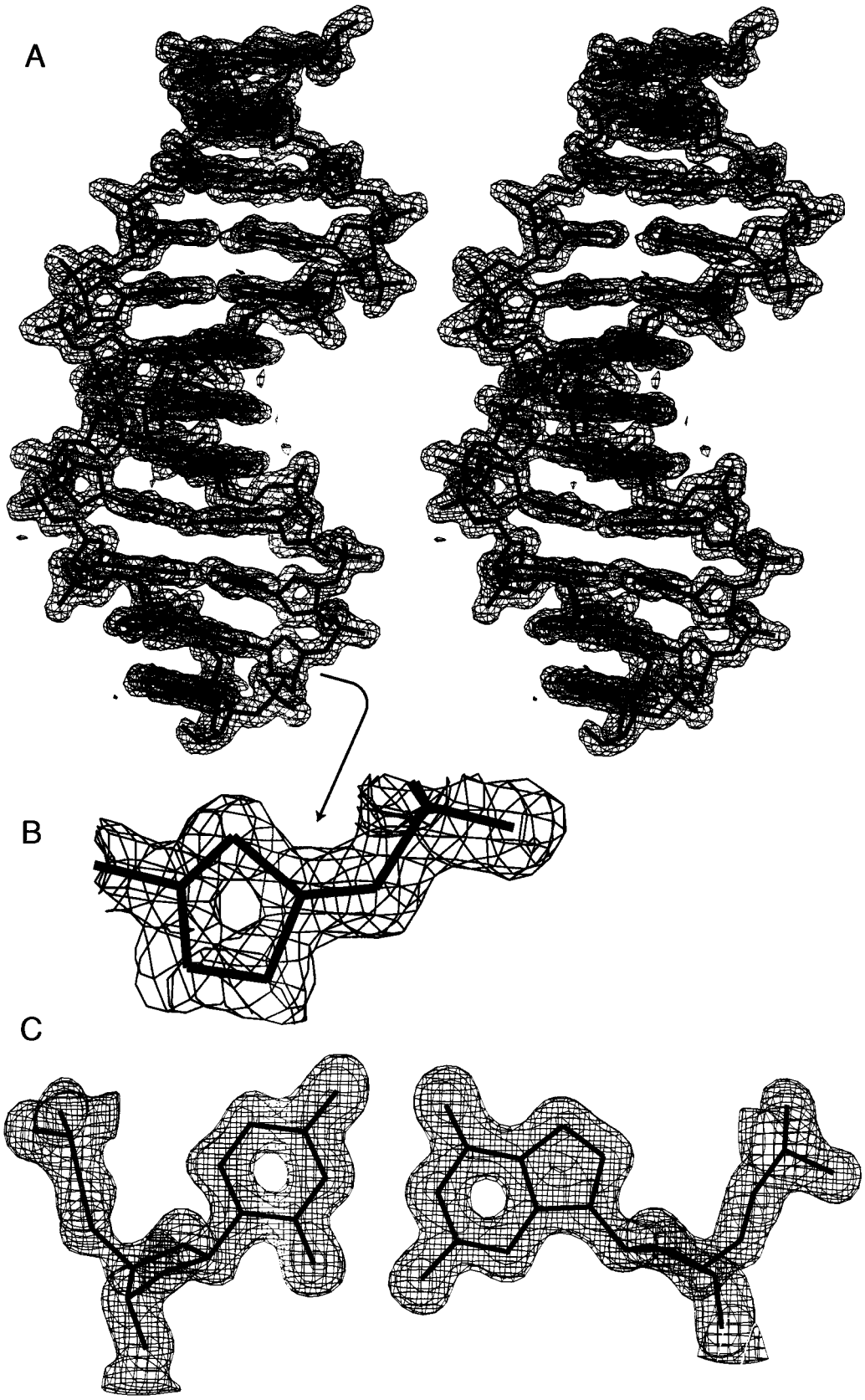

FIGURE 2: Stereoview sum 1.4 Å electron density map ($2F_o - F_c$) contoured at 1.0σ surrounding (A) the DNA duplex, (B) the deoxyribose of G(2), and (C) the C(11)-G(14) base pair. The fine detail of the maps, demonstrated by holes in the centers of the deoxyribose and pyrimidine ring systems, confirms the high quality of the data and the close fit of the model to the data.

higher DNA coordination numbers. Primary spine molecules interact with three or four sites on the DNA. The three invariant interactions are with edges of two bases on the floor of the groove and with the O4′ of a deoxyribose. In addition, where the groove is narrowest, a fourth interaction emerges, linking the primary spine to the O4′ on the opposing strand. The distances between the spine molecules and these O4′ atoms fall well within established criteria for

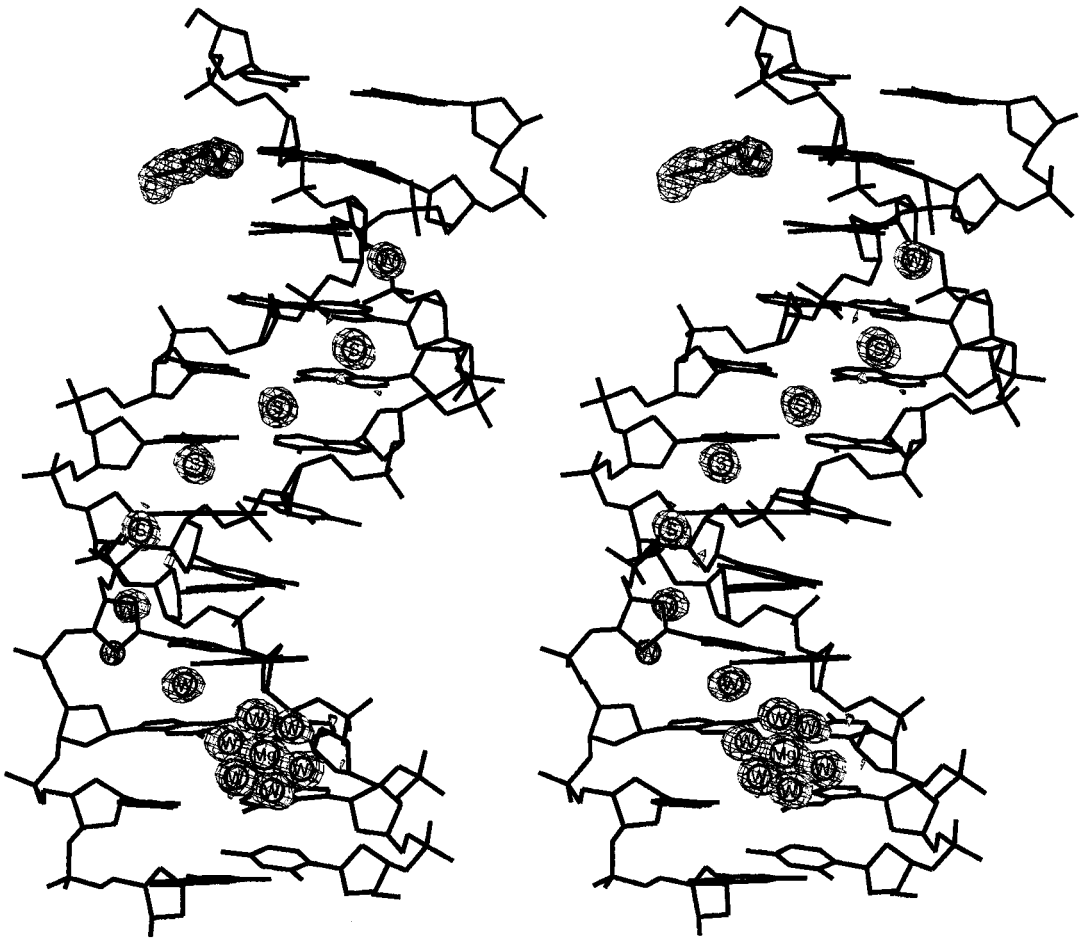

FIGURE 3: Stereoview sum ($2F_o - F_c$) electron density of structure X contoured at 1.0σ surrounding the hydrated magnesium ion, the partial spermine molecule, and the minor groove sodium and water molecules that interact with DNA bases.

Table 3: Structure X Helical Parameters (Global)

	shear S_x (Å)	stretch S_y (Å)	stagger S_z (Å)	buckle κ (deg)	propeller twist ω (deg)	opening σ (deg)
C(1)-G(24)	0.2	0.0	0.0	7.2	−23.0	−1.2
G(2)-C(23)	0.0	0.0	0.5	10.8	−19.0	−0.2
C(3)-G(22)	0.1	0.0	0.2	−6.0	−8.7	−0.3
G(4)-C(21)	−0.1	0.1	0.0	12.8	−11.4	2.6
A(5)-T(20)	0.0	0.0	0.0	5.9	−19.9	1.5
A(6)-T(19)	0.0	0.0	0.0	0.9	−18.9	5.0
T(7)-A(18)	0.0	0.0	0.1	−1.0	−20.2	4.3
T(8)-A(17)	−0.1	0.1	0.1	−1.6	−23.1	3.5
C(9)-G(16)	0.0	0.1	0.1	−14.5	−13.1	−0.6
G(10)-C(15)	0.2	0.1	0.1	5.7	−10.2	5.5
C(11)-G(14)	0.1	0.0	0.1	−2.1	−22.9	−3.8
G(12)-C(13)	0.0	0.1	0.2	5.6	−7.3	−1.7

(Local)	shift D_x (Å)	slide D_y (Å)	rise Dz (Å)	tilt τ (deg)	roll ρ (deg)	twist π (deg)
C(1)-G(24)−G(2)-C(23)	0.2	−0.4	3.3	−3.3	8.9	34.6
G(2)-C(23)−C(3)-G(22)	0.5	0.1	3.7	3.6	−12.9	42.6
C(3)-G(22)−G(4)-C(21)	−0.1	0.2	3.0	1.1	16.6	26.3
G(4)-C(21)−A(5)-T(20)	−0.6	−0.7	3.4	−3.3	−0.7	36.9
A(5)-T(20)−A(6)-T(19)	0.1	−0.9	3.3	−0.5	−0.5	37.2
A(6)-T(19)−T(7)-A(18)	0.0	−1.0	3.2	−0.4	−4.0	33.2
T(7)-A(18)−T(8)-A(17)	−0.2	−0.9	3.2	1.2	0.3	33.8
T(8)-A(17)−C(9)-G(16)	−0.2	−0.8	3.6	1.8	−6.2	41.1
C(9)-G(16)−G(10)-C(15)	0.8	0.4	3.0	−3.2	9.2	28.9
G(10)-C(15)−C(11)-G(14)	−1.5	−0.2	3.4	−3.5	−13.9	40.4
C(11)-G(14)−G(12)-C(13)	−0.2	−0.3	3.1	0.6	10.0	35.0

interactions such as hydrogen bonds. Thus, where the minor groove is narrowest, the primary spine of structure X stitches the two backbones together, linking two O4′ atoms offset by three residues in the 3′ direction.

We suggest that the primary spine contains sites (indicated by S in Table 5 and Figure 5) that are occupied partially by sodium ions and partially by water molecules. We have conducted valence calculations (*20*, *21*) that provide strong

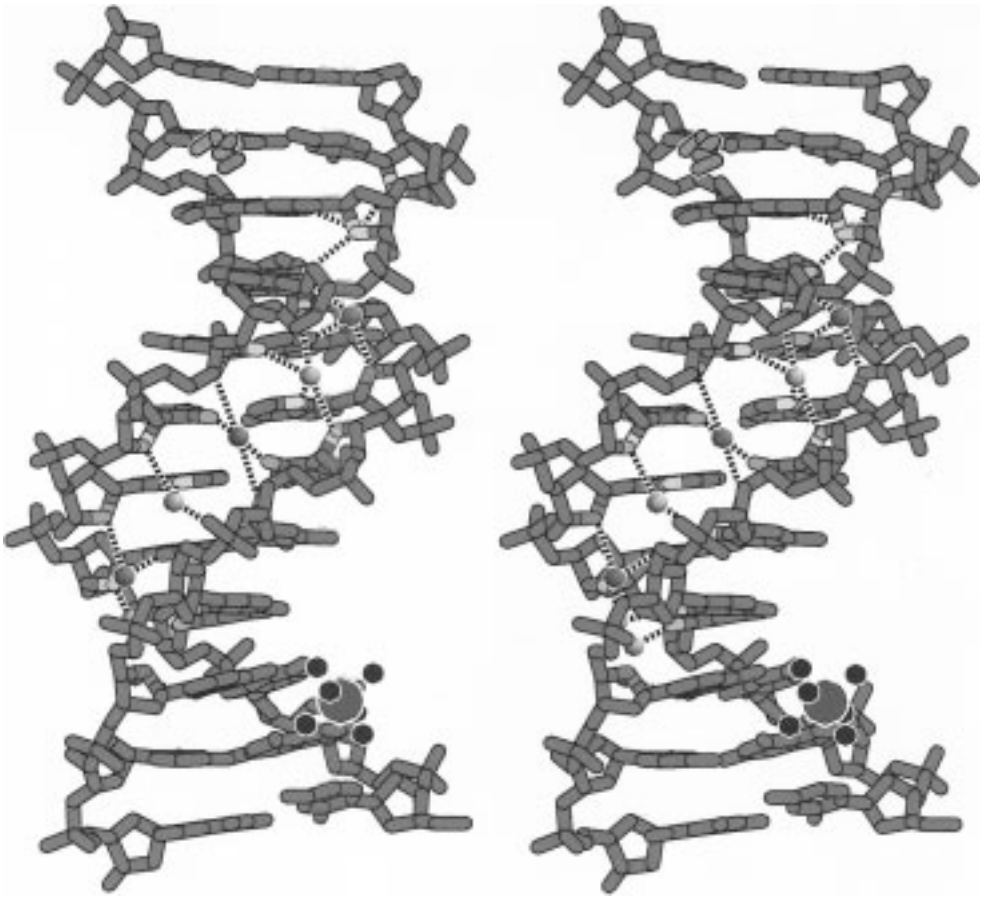

FIGURE 4: Stereoview of structure X and the primary spine. Primary spine molecules are colored alternately red and yellow. The DNA is blue except for atoms that interact with the spine. The colors of the DNA atoms match the molecules and ions with which they interact. Red spine molecules interact with red DNA atoms and so on. The spermine fragment is purple, and the magnesium ion is green. The first hydration shell of the magnesium ion is dark blue.

Table 4: Minor Groove Width[a]

phosphate pair	structure X distance (Å)	structure 1BNA distance (Å)	Δ distance (Å)
A(5)−G(24)	9.3	8.7	0.6
A(6)−G(22)	6.7	7.1	−0.4
T(7)−G(22)	4.7	5.3	−0.6
T(8)−T(21)	3.1	4.2	−1.1
C(9)−T(20)	3.5	4.0	−0.5
G(10)−T(19)	3.1	3.1	0.0
C(11)−A(18)	4.6	5.2	−0.6
G(12)−A(18)	3.9	4.8	−0.9

[a] P−P separation minus 5.8 Å.

support for this partial sodium/partial water model of the primary spine. The bond valence model assumes that the valence of metal is equal to the sum of its bond valences (*20*). Bond valences are determined by empirical relationships between bond lengths and strengths. Probable metal sites in inorganic crystals have been located by calculating maps of valence for specific metals (*22*). In a search for sodium ions bound to protein molecules, Nayal and Di Cera (*21*) have calculated the sodium specific valence for more than three hundred thousand water molecules in the PDB.

The sodium specific valence (v_{Na^+}) of an oxygen-liganded solvent molecule in a crystal structure is determined by the relationship

$$v_{Na^+} = \sum (R_j/R_o)^{-N}$$

where R_o is 1.622, R_j is the distance from the solvent molecule to oxygen ligand j, N is 4.29, and the sum is over the total number of oxygen ligands. For all water molecules in the PDB, v_{Na^+} averages 0.18 valence unit (*21*) with a 95% confidence limit of 0.0005 (calculated here from sample size and $\sigma = 0.13$). The v_{Na^+} values for the spine of structure X are given in Table 5. The average v_{Na^+} for the primary spine is 0.48 valence unit with a 95% confidence limit of 0.01. Thus, the average v_{Na^+} for the spine molecules of structure X is greater than would be expected if the spine were composed exclusively of water molecules. The difference between the bulk water average and the spine average (Δv_{Na^+}) is statistically significant. The Δv_{Na^+} equals 0.30 valence unit, which exceeds the sum of the 95% confidence limits for the bulk and spine averages. In fact, Δv_{Na^+} exceeds the sum of the 99% confidence limits (not shown). Our results support a model in which the primary spine sites are not fully occupied by either water molecules or sodium ions. In no case does v_{Na^+} approach 1.0 valence unit, which would indicate a fully occupied sodium ion. The most reasonable explanation for the intermediate value of v_{Na^+} is that primary spine sites are partially occupied by both sodium ions and water molecules. An alternative model that cannot be ruled out by our data is one in which the spine sites are partially occupied by magnesium ions.

The thermal factors of the primary spine of structure X are anomalous in comparison with those of the rest of the complex. All minor groove-bound solvent molecules, in-

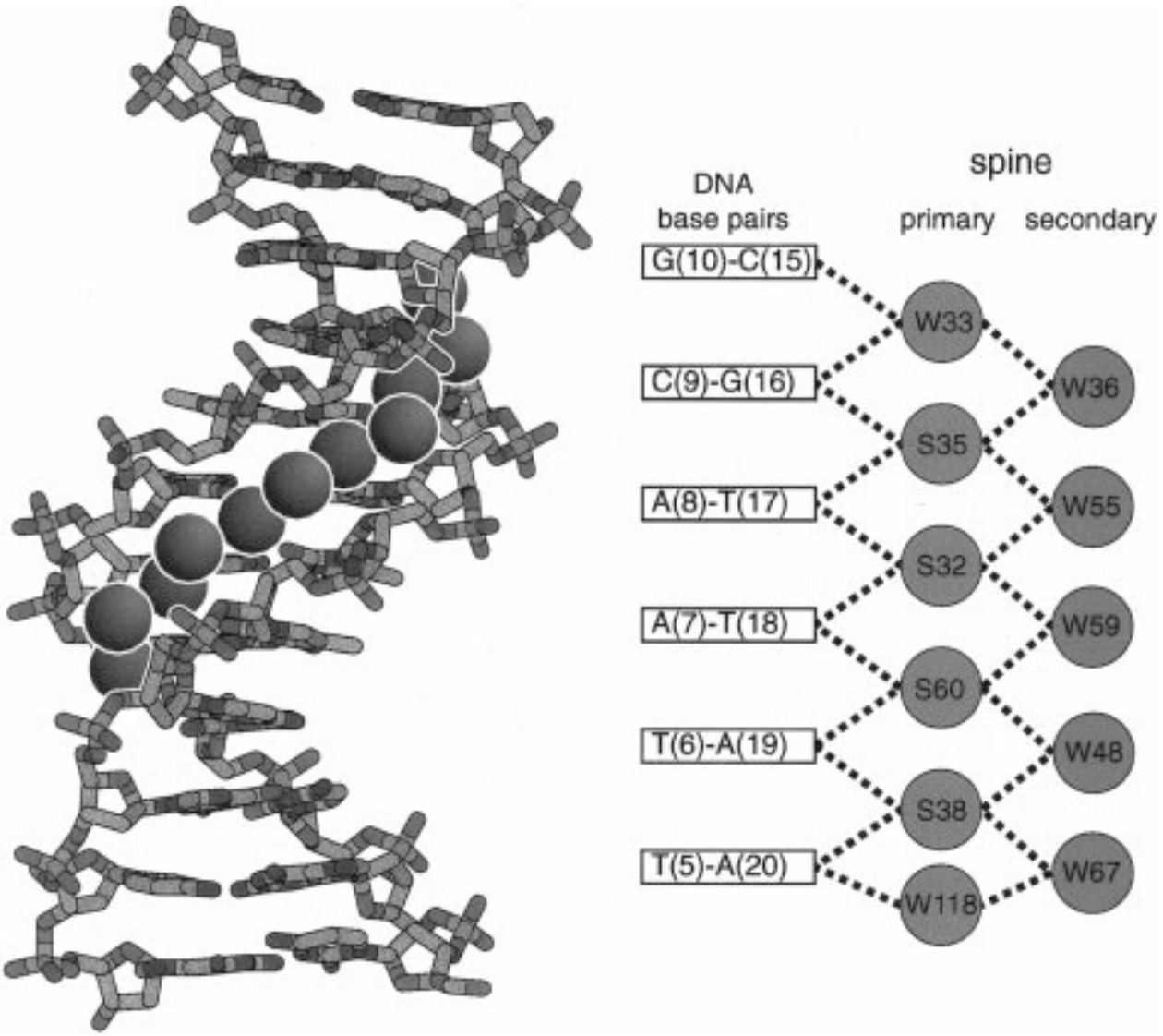

FIGURE 5: Primary and secondary layers of the spine of hydration of structure X. The primary spine is cyan, and the secondary spine is magenta. DNA atoms are colored by CPK. On the right is a schematic diagram of the primary and secondary spines of hydration with the same coloring scheme.

cluding the primary spine, were treated as water molecules during structure refinement. The average thermal factor for all water molecules in the completed model is 25.4 Å^2 (σ = 8.8). The average thermal factor of the primary spine (11.4 Å^2, σ = 3.4) is much smaller than the average for all solvent molecules. Excluding the magnesium and its primary hydration sphere, the lowest non-DNA thermal factors in the completed model are those of the primary spine (S35, 5.89 Å^2; S32, 9.00 Å^2; and W33, 9.02 Å^2). In fact, the average thermal factor of the primary spine is similar to the average for the DNA (9.7 Å^2, σ = 3.5).

A second striking feature of the minor groove hydration of structure X is the degree of order within the secondary spine. As described by Drew and Dickerson for structure 1BNA (*2*), the secondary spine rides piggyback on the primary spine (Figure 5 and Table 5). The minor groove hydration system is built like a pyramid, with each internal primary spine ion providing half-support for two secondary spine water molecules. In structure X, primary spine W33 and S35 support secondary spine molecule W36, S35 and S32 support W55, S32 and S60 support W59, and S60 and S38 support W48. The secondary spine terminates where S38 and W118 support W67. W118 is not formally part of the primary spine as it forms a single interaction with a DNA base.

Do dodecamers form intrinsically disordered crystals due to weak lattice interactions and/or DNA conformational polymorphism? This presumption is proven false by this report. The origins of improved crystal quality, compared with previous efforts, are unresolved. The increase in crystalline order may derive from (i) higher levels of spermine and lower levels of MPD than those used previously [crystallization conditions for B-form dodecamers have been reviewed (*23*)], (ii) specific conditions and the duration of cryoprotection, (iii) initial and final temperatures of shock cooling, (iv) the rate of cooling, or (v) the low occupancy of a minor groove binder even though it is not observed in the electron density maps. We have obtained high-resolution data sets for [d(CGCAAATTTGCG)]$_2$ and for several covalently modified dodecamers (X. Shui and L. D. Williams, unpublished). Others have recently observed that, upon proper cooling, dodecamer crystals of drug complexes and of chemically modified DNA routinely diffract to 1.5−

Table 5: Minor Groove Hydration of Structure X[a]

water or ion	ligand type	atom	distance (Å)	atom	distance (Å)
W45	base	G(22) N3	2.7		
	sugar	G(22) O4′	3.1		
	water	W167	3.1	W58	3.3
W167	base	C(21) O2	2.9		
	sugar	G(22) O4′	3.0		
	water	W118	2.4	W45	3.1
W47	base	G(4) N2	3.3	G(4) N3	2.8
	sugar	A(5) O4′	3.2	G(12) O3′ [b]	2.8
	water	W172	2.9	W66	2.7
W118	base	A(5) N3	2.9		
($\upsilon_{Na^+} = 0.43$)	sugar	A(6) O4′	3.2		
	water	W167	2.4	W67	3.1
S38	base	A(6) N3	2.9	T(20) O2	2.8
($\upsilon_{Na^+} = 0.48$)	sugar	T(7) O4′	3.2		
	water	W48	2.7	W67	2.8
S60	base	T(7) O2	2.7	T(19) O2	2.8
($\upsilon_{Na^+} = 0.49$)	sugar	T(8) O4′	3.3	T(20) O4′	3.1
	water	W48	3.0	W59	2.8
S32	base	T(8) O2	2.9	A(18) N3	3.0
($\upsilon_{Na^+} = 0.48$)	sugar	C(9) O4′	3.1	T(19) O4′	3.0
	water	W55	2.8	W59	2.8
S35	base	C(9) O2	2.7	A(17) N3	2.9
($\upsilon_{Na^+} = 0.46$)	sugar			A(18) O4′	3.0
	water	W55	2.8	W36	2.7
W33	base	G(10) N3	2.8	G(16) N2	3.1
($\upsilon_{Na^+} = 0.41$)	sugar	G(10) O4′	3.2	C(11) O4′	2.7
	water	W56	3.7	W36	2.7

[a] Solvent molecules that interact with DNA bases within the minor groove. [b] Related to the previous duplex by (1.5 − *x*, 1.0 − *y*, *z* + 0.5).

1.8 Å (S. Neidle, personal communication; M. Egli, personal communication).

Is spermine required for crystallization but not observable in electron density maps because it is disordered? Does spermine enter dodecamer crystals? Throughout the refinement of structure X, electron density maps show clear evidence of a spermine molecule bound in the major groove (Figure 3). Since no spermine molecules are contained within the 68 dodecamer entries in the PDB and NDB, a great deal of care was taken to minimize bias during refinement of structure X. In the final analysis, evidence for the correctness of the spermine molecule is overwhelming. (i) All other aspects of the refinement were completed before the spermine molecule was added to the model. The pre- and post-spermine maps are equally unambiguous. (ii) After completion of the refinement, the location of the spermine molecule was reconfirmed by refined omit maps. (iii) A spermine molecule has now been observed in a nearly identical position in a very high-resolution structure of $[d(CGCAAATTTGCG)]_2$ (X. Shui and L. D. Williams, unpublished).

One or both ends of the spermine molecule are disordered; only seven of the fourteen non-hydrogen atoms of the spermine molecule are visible in the electron density maps. Thermal disorder can be ruled out as the source of the disorder as the data were collected at low temperatures, indicating that a portion of the spermine is statically disordered. The ordered section of the spermine molecule is fully hydrated but makes a single contact with the floor of the major groove. The disorder of the spermine molecule makes the registry (i.e., the atomic assignments as carbon and nitrogen) difficult to unambiguously assign. However, an apparent hydrogen bond to the N7 of G(10) (3.4 Å) allows at least tentative assignment of one of the internal nitrogen atoms of the spermine molecule.

Table 6: Interactions of the First Magnesium Hydration Shell with DNA

DNA residue	DNA atom	water molecule	distance (Å)
G(2)	O6	W2	2.6
G(2)	N7	W5	2.8
G(22)	O6	W1	2.7
A(6)[a]	O1P	W6	2.9
A(6)[a]	O2P	W6	3.3
A(6)[a]	O2P	W4	2.7
T(7)[a]	O1P	W3	2.7

[a] Related to the previous DNA duplex by a symmetry operator ($-x$, $y + 1/2$, $-z + 1/2$).

Role of Magnesium. A hydrated magnesium ion is observed in the electron density maps of structure X (Figure 3). The first hydration shell of the magnesium ion forms hydrogen bonds to the O6 and N7 positions of guanines of the G(2)-C(23) and C(3)-G(22) base pairs (Figure 4 and Table 6). The magnesium hydration shell does not form hydrogen bonds with any of the other bases. However, our detailed structural analysis suggests that the first hydration sphere of the magnesium engages in cation−π interactions with DNA bases (L. McFail-Isom, X. Shui, and L. D, Williams, unpublished). The magnesium hydration shell also forms hydrogen bonds to phosphate oxygen atoms of a symmetry-related duplex.

The total negative charge of the DNA (22 negative charges) is not compensated by the observed cations (one spermine molecule and one magnesium ion and four partial sodium spines gives less than ten positive charges). This discrepancy suggests that some electron density peaks assigned to water molecules are in fact partially occupied ions or partially ordered ions.

Variation of DNA Helical Parameters; Effects of Sequence, Chemical Modification, and Coordinate Error. To determine the effects of resolution on coordinate error, and to validate various methods of estimating errors, we have examined the effects of resolution on deviations from a "true" structure.

Fourier truncation causes errors in simulated structures. A refined DNA dodecamer, with solvent removed, was used as a true structure. This true structure was used to calculate $|F_{obs}(hkl)|^s$, a complete asymmetric unit of synthetic data. These data are free of random error in structure factor amplitudes. The same starting model was annealed and refined to convergence against $|F_{obs}(hkl)|^s$ truncated at six resolutions: 1.4, 1.7, 2.0, 2.5, 3.0, and 3.5 Å. Thus, six simulated structures, a−f (Table 7), were generated. Deviations of structures a−f from the originating structure, which was used to generate the initial error-free data, result from a combination of Fourier truncation and force field errors. This set of structures was used to determine validities of various methods of estimating coordinate error.

Three independent methods were used to estimate or calculate coordinate errors of the simulated series. In method 1 (calculated errors), errors of simulated structures were calculated directly by comparison with the true structure, which was used to generate the data. In method 2, errors of test structures were estimated by deviations from the most accurate (reference) structure. Errors of structures b−f were estimated by their deviations from structure a, which is

Table 7: Refinement Statistics and Coordinate Error for Simulated Structures with Varying Extents of Fourier Truncation

structure	resolution (Å)	number of reflections	reflections per atom	*R*-factor (%)	*R*-free (%)	coordinate error[a]	estimated coordinate error[b]	estimated coordinate error[c]
a	1.4	13542	27.9	5.8	5.8	0.04	NA[d]	0.05
b	1.7	7677	15.8	5.1	9.2	0.05	0.02	0.05
c	2.0	4776	9.8	6.6	7.0	0.09	0.07	0.07
d	2.5	2473	5.1	7.5	8.4	0.14	0.12	0.11
e	3.0	1451	3.0	6.0	11.0	0.11	0.09	0.10
f	3.5	915	1.9	7.3	8.9	0.22	0.20	0.10

[a] rms deviations of atomic positions of test structures a−f from the "true" structure, which was used to generate the data. [b] Shui method: rms deviations of atomic positions of test structures b−f from the most accurate structure (structure a). [c] Luzzati method (*24*): Luzzati plots for this series are shown in Figure 2S in the Supporting Information. [d] NA, not available.

Table 8: Refinement Statistics for Duplicate Dodecamers[a]

structure[b]	reported resolution (Å)	number of reflections	reflections per atom	$F > n\sigma(F)$	*R*-factor	temperature (°C)	reference
X[c]	1.4	11438	17.3	$n = 0$	0.197	−137	this report
BDL084							
1BNA[c]	1.9	2725	4.2	$n = 2$	0.178	room temperature	*1*
BDL001							
2BNA[c]	2.7	1836	3.2	$n = 2$	0.210	−257	*4*
BDL002							
7BNA[c]	1.9	5599	NA[k]	$n = 0$	0.206	room temperature	*7*
BDL005							
9BNA[c]	1.9	3979	7.2	not reported	0.188	room temperature	*70*
BDL020							
3BNA[d]	3.0	1515	2.8	$n = 2$	0.173	20	*5*
BDLB03							
4BNA[d]	2.3	2919	4.8	$n = 2$	0.216	7	*5*
BDLB04							
4DNB[e]	2.0	2379	4.1	$n = 2$	0.169	room temperature	*71*
BDLB13							
265D[f]	2.01	4233	6.9	$n = 2$	0.172	−173	*j*
BDLB72							
266D[g]	2.03	4049	7.3	$n = 2$	0.166	5	*j*
BDLB73							
270D[g]	2.01	4028	6.6	$n = 0$	0.176	−173	*j*
BDLB74							
290D[h]	2.5	2072	3.6	$n = 2$	0.170	room temperature	*72*
BDLS79							
291D[i]	2.1	3605	6.4	$n = 2$	0.207	room temperature	*72*
BDLS80							

[a] For each structure, the unit cell is as follows: $a = 25$ Å, $b = 40$ Å, $c = 66$ Å, and $\alpha = \beta = \gamma = 90°$; the space group is $P2_12_12_1$. The first four entries (structures 1BNA, 2BNA, 7BNA, and 9BNA) are chemically identical to structure X. [b] (Top) Code for the Brookhaven Protein Data Bank. (Bottom) Code for the Nucleic Acid Data Bank. [c] Unmodified $[d(CGCGAATTCGCG)]_2$. [d] $[d(CGCGAATT^{5Br}CGCG)]_2$. [e] $[d(CGCGA^{6Me}ATTCGCG)]_2$. [f] $[d(CG^{5Me}CGAATT^{5Me}CGCG)]_2$. [g] $[d(CGCGAATT^{5Me}CGCG)]_2$. [h] $[d(CGCGAA^{(M)C}T^{(M)C}TCGCG)]_2$; residues $^{(M)C}$T(7), $^{(M)C}$T(8), $^{(M)C}$T(19), and $^{(M)C}$T(20) are 6′-α-methyl carbocyclic thymidines. [i] $[d(CGCGAA^{(OH)C}T^{(OH)C}TCGCG)]_2$; residues $^{(OH)C}$T(7), $^{(OH)C}$T(8), $^{(OH)C}$T(19), and $^{(OH)C}$T(20) are 6′-α-hydroxy carbocyclic thymidines. [k] NA, not available. [j] B. Partridge and S. Salisbury, unpublished observations.

refined against the highest-resolution data and is therefore the most accurate. This method assumes that errors of test structures are significantly greater than those of the reference structure. In method 3 (Luzzati), errors for structures a−f were estimated by the method of Luzzati (*24*). Luzzati plots for the synthetic series are shown in Figure 2S of the Supporting Information. The Luzzati method is a generally accepted method for estimating coordinate errors of X-ray structures but requires structure factor amplitudes and so cannot be applied to many PDB and NDB entries.

For the series of synthetic structures (a−f), the three methods give generally consistent results (Table 7). In only two cases do estimated rms coordinate errors deviate by more than a factor of 2 from the true coordinate errors. The method of Luzzati appears to underestimate error at low resolution (structure f). The method of Shui underestimates error at high resolution (structure b). The reason for the failure of the Shui method at high resolution is straightforward. The reference (structure a, 1.4 Å resolution) is not significantly more accurate than the test (structure b, 1.7 Å resolution). However, for structures c−f and for duplicate dodecamer structures in the PDB, reasonable estimates of errors can be expected from deviations from a 1.4 Å resolution structure.

Conformational polymorphism is caused by a combination of coordinate error and intrinsic heterogeneity. The results above demonstrate that crystallographic errors can be reliably estimated by comparison with a significantly more accurate, isomorphous test structure. Using a comparison with structure X, which is far more accurate, we have performed an analysis of deviations within a subset of the published dodecamer family of three-dimensional structures. Figure 6 illustrates an example of significant differences that distinguish structures that are chemically and crystallographically identical. One goal of the work described here is to determine if conformational heterogeneity in DNA dodecamer structures is intrinsic, resulting from effects of sequence and/or chemical modification. An alternative source of

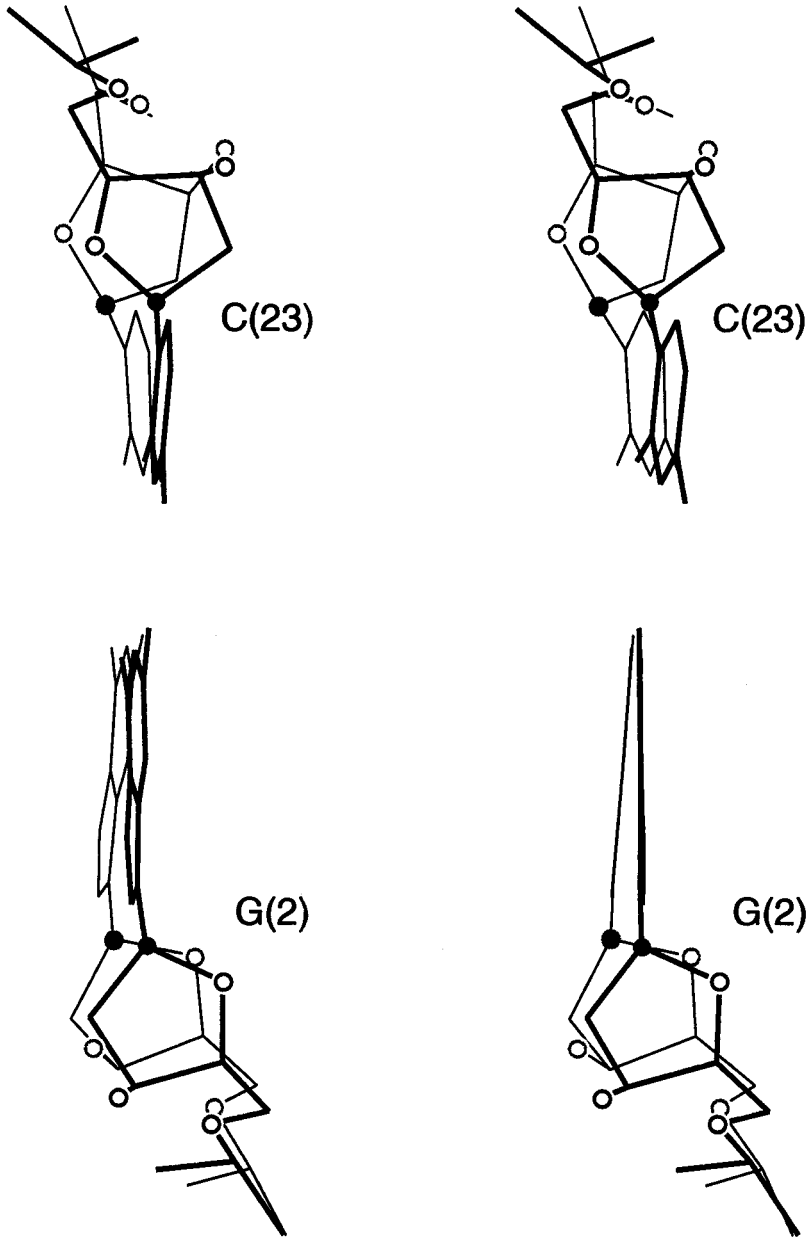

FIGURE 6: Stereoview of the G(2)-C(23) base pairs of structure X (thick lines) and the DNA dodecamer duplex 1BNA (*1*) (thin lines). The superimposition was performed using all DNA atoms of the dodecamer duplexes. The C1′ atoms are highlighted by dots, and the oxygen atoms of the deoxyriboses are indicated. The view is into the minor groove to illustrate differences in the buckles.

apparent conformational heterogeneity is experimental error. Inaccuracies of atomic positions are propagated to helical parameters.

We establish above that coordinate errors of test structures are correctly estimated by deviations from a more accurate reference structure. Here errors of helical parameters and torsion angles have been estimated in the same way. Table 8 gives resolution, *R*-factors, and the number of data/number of atoms for the twelve duplicate structures. Structures 1BNA, 2BNA, 7BNA, and 9BNA are chemically identical to structure X, while the other duplicate structures differ from structure X by minor chemical modifications (Table 10). None of the chemical modifications of the twelve duplicate structures affect base pairing interactions. For example, structure 265D has 5-methylcytosines at four positions of the dodecamer duplex.

For the twelve duplicate structures, a hypotheses of identical torsion angles and parameters was formulated and tested. The R^2 statistic was used to compare torsion angles and helical parameters of the twelve duplicates with those of structure X. R^2 gives the proportion of variation in a dependent variable that is explained by a regression. All else being equal, parameters with little intrinsic variation are expected show lower R^2 statistics than parameters with large intrinsic variation. Constant error with a decreasing extent of intrinsic variation would increase R^2.

The regression here is $P[n,\mathrm{D}(m)]$ to $P(n,\mathrm{X})$, where P represents a helical parameter or torsion angle, n represents the residue, base pair, or base pair step (1−24 for most torsion angles, 1−12 for intra-base pair parameters, or 1−11 for inter-base pair parameters), X represents structure X, and D(m) represents a duplicate structure ($m = 1-12$). Structure X is considered independent, and the duplicate structures are considered dependent. For example, if the 12 propeller twists

Table 9: Comparison of Base−Base Parameters of Duplicate Structures with Structure X [R^2 (%)][a]

structure	shear	stretch	stagger	buckle	propeller twist	opening
1BNA	5	2	15	68	56	78
2BNA	0	24	45	82	57	66
7BNA	5	0	24	68	56	75
9BNA	18	0	1	70	42	45
3BNA	6	9	13	33	42	63
4BNA	23	49	44	74	48	57
4DNB	2	7	46	37	59	58
265D	1	6	12	49	57	45
266D	15	5	29	68	46	17
270D	21	0	10	67	63	62
290D	0	4	11	38	14	20
291D	40	3	48	53	76	24
average	11	9	25	59	51	51

structure	shift D_x	slide D_y	rise D_z	tilt τ	roll ρ	twist ω
1BNA	62	85	54	71	76	42
2BNA	58	85	73	44	55	19
7BNA	66	86	50	71	73	59
9BNA	70	96	31	17	73	82
3BNA	66	81	8	18	80	30
4BNA	47	79	55	36	31	64
4DNB	42	75	35	32	68	81
265D	64	87	51	70	77	62
266D	62	77	17	60	51	50
270D	60	86	57	24	74	94
290D	20	58	9	18	50	40
291D	38	86	17	79	58	87
average	55	82	38	45	64	59

[a] The R^2 gives the proportion of variation in a dependent variable that is explained by a regression line. The regression here is $P[n,\mathrm{D}(m)]$ to $P(n,\mathrm{X})$, where P represents a helical parameter or torsion angle, n represents the base pair or base pair step (1−12 for each intra-base pair parameter or torsion angle or 1−11 for each inter-base pair parameter), X represents structure X, and D(m) represents a duplicate structure (m 1−12). Structure X is considered independent, and the duplicate structures are considered dependent.

Table 10: Deviations of Helical Parameters within Duplicate DNA Dodecamers[a]

structure	shear (Å)	stretch (Å)	stagger (Å)	buckle (deg)	propeller twist (deg)	opening (deg)
X	0.10[b]	0.06[b]	0.16[b]	7.6	5.9	3.0[b]
1BNA	0.38	0.09	0.21	6.4	8.9	5.3
2BNA	0.37	0.08	0.23	9.2	11.9	7.2
7BNA	0.29	0.10	0.20	5.4	8.6	4.6
9BNA	0.25	0.13	0.18	6.8	10.2	4.3
3BNA	0.47	0.14	0.32	8.7	10.7	9.5
4BNA	0.46	0.12	0.43	11.4	15.9	6.7
4DNB	0.61	0.18	0.40	5.0	8.0	5.4
265D	0.40	0.10	0.30	8.1	5.6	6.4
266D	0.37	0.24	0.31	6.3	6.3	4.3
270D	0.31	0.16	0.24	4.4	5.6	5.1
290D	0.52	0.15	0.34	6.9	6.7	5.4
291D	0.42	0.16	0.19	4.9	9.8	2.8

	shift D_x (Å)	slide D_y (Å)	rise D_z (Å)	tilt τ (deg)	roll ρ (deg)	twist ω (deg)
X	0.59	0.50	0.23	2.5[b]	9.7	5.0
1BNA	0.67	0.48	0.21	3.4	7.2	4.4
2BNA	0.73	0.41	0.26	3.8	7.7	4.2
7BNA	0.66	0.50	0.18	3.3	6.9	4.7
9BNA	0.61	0.53	0.21	4.0	7.9	3.2
3BNA	0.72	0.42	0.33	3.4	7.3	3.8
4BNA	0.88	0.39	0.31	4.7	8.2	4.2
4DNB	0.60	0.38	0.15	3.8	7.3	6.9
265D	0.56	0.41	0.25	4.0	7.5	5.0
266D	0.61	0.59	0.15	4.8	6.7	5.1
270D	0.69	0.50	0.17	3.6	7.0	5.8
290D	0.56	0.39	0.24	3.8	8.9	5.0
291D	0.60	0.51	0.14	3.3	8.2	6.1

[a] Standard deviation from the mean within that structure. [b] The dispersions of these parameters are smaller in structure X than in any of the duplicate structures. The single exception is the opening of structure 291D.

of a duplicate structure are identical to the corresponding propeller twists of structure X, the R^2 for that parameter of that structure would equal 100%.

The results indicate that parameter deviations are poorly determined for most helical parameters (Table 9). R^2 statistics are worst for shear, stretch, and stagger, which generally vary by less than 0.5 Å (σ) within a given structure (Table 10). Buckle, propeller twist, opening, shift, rise, tilt, roll, and twist give R^2 statistics of 40−60%, indicating a significant probable contribution from coordinate error. Deviations of slide appear to be the most reliable, with R^2 statistics around 80%. In sum, the correlation between the deviations of helical parameters in structure X and those in previous, lower-resolution structures is not strong

The relative contributions of chemical modifications and structural error to deviations of helical parameters can be estimated. For most helical parameters, the R^2 statistics of chemically identical duplicates are not greater than those of duplicate structures. Therefore, contributions of the chemical modifications to variation of these parameters are not greater than contributions from other factors. For several parameters such as buckle, propeller twist, shift, and slide, the R^2 statistics of chemically identical duplicates are greater than average for all duplicate structures. For these parameters, chemical modifications contribute measurably to variation.

Structure X appears to be conformationally more ordered than previous structures, consistent with the implication above that some of the conformational heterogeneity that was previously attributed to sequence in fact arises from coordinate error. The relative contributions of DNA sequence and coordinate error to conformational heterogeneity were estimated here by comparing deviations within structures (Table 10). The parameters shear, stretch, stagger, and tilt show less variation within structure X than within any duplicate structure. Similarly, the parameters propeller twist, opening, and shift show less deviation than nearly all duplicate structures. Therefore, the variation of these seven parameters in previous dodecamer structures appears to be inflated. Only for the parameters buckle, slide, rise, roll, and twist does the variation within structure X appear to be representative of the duplicate set. Therefore, for these five parameters, the observed variation may result from intrinsic differences arising from sequence and chemical modification.

Variation of DNA torsion angles is caused by sequence, chemical modification, and coordinate error. For each of the torsion angles γ, ϵ, α, and β, the R^2 statistic (Table 11) averages around 25%, indicating consistently poor correlation of variation within those structures to variation within structure X. For χ, ϵ, and ζ the R^2 statistic for each of the four chemically identical duplicates is greater than the average for all duplicate structures. Therefore, the contribution of the chemical modifications to these torsion angles appears to rise above random coordinate error.

Table 11: Comparison of Torsion Angles of Duplicate Structures with Structure X [R^2 (%)]

	χ (deg)	γ (deg)	δ (deg)	ϵ (deg)	ζ (deg)	α (deg)	β (deg)
1BNA	58	41	63	21	85	13	13
2BNA	62	58	70	32	82	21	9
7BNA	52	17	46	32	83	2	9
9BNA	51	29	45	21	88	2	8
3BNA	33	22	38	14	79	46	16
4BNA	40	24	56	26	70	22	8
4DNB	40	12	46	4	46	7	43
265D	59	29	43	20	74	4	18
266D	47	34	18	16	55	18	13
270D	66	17	54	21	78	5	57
290D	45	27	27	6	36	2	12
291D	55	7	42	12	39	7	54
average	51	27	46	19	68	13	22

DISCUSSION

X-ray crystallography continues to provide highly detailed structural information on nucleic acid structure. Effects of DNA sequence (*25*−*27*), base modification (*28*, *29*), mismatches (*30*, *31*), minor groove binders (*32*−*34*), intercalators (*35*−*38*), lattice forces (*39*, *16*), and hydration (*40*) on DNA conformation have been explained by similarities within and differences between an increasing number of X-ray structures (*18*, *19*). As summarized in a series of recent reviews (*16*−*19*), many features of the original structure of [d(CGCGAATTCGCG)]$_2$ (structure 1BNA) are thought to apply to DNA in general. Here we suggest that some previous conclusions need extension or re-evaluation.

Water on Sodium in the Spine of Hydration. Patterns of hydration of structure X reveal several surprises. The primary spine is composed of sites occupied partially by sodium ions and partially by water molecules. The minor groove of AT tract DNA appears to form a linear array of natural sodium binding sites. The fractional occupancy of sodium and water at each of these sites would depend on environmental factors such as the concentration of sodium, and other ions might compete with sodium for spine occupancy.

Sodium ions are difficult to distinguish from water molecules by macromolecular crystallographic techniques because a sodium ion and a water molecule contain the same number of electrons. Sodium ions, unlike magnesium ions, cannot be identified by strict geometeric criteria. For example, sodium ions bound to proteins are five-, six, and seven-coordinate (*21*). The octahedra of six-coordinate sodium ions are commonly distorted.

Valence calculations (*20, 21*) as used here provide clear support of a model of partial sodium occupancy of the primary spine. The bond valence model assumes that the valence of metal is equal to the sum of its bond valences (*20*). Bond valences are determined by empirical relationships between bond lengths and strengths. Probable metal sites in inorganic crystals have been located by calculating maps of valence for specific metals (*22*). Valence calculations use emperical metal−ligand bond distances to calculate the valence (v_{M^+}) of a metal (M). For an ideal solvent entity surrounded by six oxygen ligands at distances of 2.46 Å, $v_{Na^+} = 1.0$ valence unit. Conversely, a v_{Na^+} value of 1.0 should be considered strong evidence that a solvent site is fully occupied by a sodium ion. For some sites in the primary spine, the v_{Na^+} values are uniformly less than 1.0 valence unit (Table 5), suggesting that the primary spine is not composed exclusively of sodium ions. However, the v_{Na^+} values for the primary spine molecules of structure X are greater than expected for water molecules. This difference between bulk water and primary spine sites is statistically significant beyond 99% certainty limits. The simplest model that explains the observed v_{Na^+} values is one in which sodium ions penetrate the primary spine and partially occupy each site. Thus, sodium and water molecules share some sites of the primary spine of hydration.

In principle, sodium ions can also be distinguished from water molecules by differences in electron density. The electrons of a sodium ion are more tightly held and occupy less volume than those of a water molecule. Therefore, a sodium ion treated as a water molecule during crystallographic refinement should show an anomalously low thermal factor. Thermal factors of low-resolution structures are notoriously inaccurate and are more error-prone than atomic positions. In fact, Nayal and DiCera analyzed the entire PDB and found that thermal factors do not correlate with sodium specific valence (*21*). Therefore, an appropriate threshold for accuracy is required for observation of the expected relationship between the thermal factor and sodium specific valence. This relationship appears to rise over experimental error in structure X. The thermal factors of the primary spine of structure X are anomalous in comparison with those of the rest of the complex. When the magnesium and its primary hydration sphere are excluded, the lowest non-DNA thermal factors in the completed model are those of the primary spine. The average thermal factor of the primary spine is lower than the average for the DNA.

In sum, our crystallographic results support those of molecular dynamics calculations (*41*) and solution NMR experiments (*42*), demonstrating that cations penetrate the primary spine. The most reasonable scenario is one in which each primary spine entity is comprised partially of sodium and partially of water.

As described by Drew and Dickerson (*2*), the secondary water spine rides piggyback on the primary spine (Figure 5). The minor groove hydration resembles an orderly pyramid, with the sodium layer forming a template for the next layer, which is composed of water molecules.

Although the basic network of interactions is similar in structures X and 1BNA, the geometric regularity of the secondary spine is much greater in structure X. The secondary spine lacks direct interaction with the DNA. Thus, the high degree of order of the secondary spine is derived entirely from the template provided by the primary spine. The exciting aspect of this geometric regularity is that it shows how one solvent layer can form a template for a second solvent layer that is so geometrically regular it can in turn act as a template for a third layer, and so on. This templating of the primary spine upon the DNA bases, and the secondary spine upon the primary spine, provides a model for transmission of DNA sequence information into the DNA hydration region. This observation of a highly ordered spine with a primary sodium layer and a secondary water layer is consistent with results of solution experiments. The number of ordered solvent molecules displaced from AT tracts upon binding of minor groove binders (*43*, *44*) is much larger than the number of ordered molecules in the primary spine.

Electrostatic Origins of Narrow Minor Groove and Axial Bends. We suggest that many sequence-dependent features of DNA conformation are mediated by cations. We propose that AT tract induction of DNA bending (*45*−*47*) and of minor groove narrowing both originate from preferential localization of cations. Both these structural features would arise naturally from localization of sodium ions as observed in structure X. It is well-established that cation localization causes conformation changes in DNA. Mirzabekov and Rich proposed that a driving force for DNA bending would arise from phosphate neutralization by cations (*48*). Maher and Strauss-Soukup provided elegant support for this model by chemically removing anionic charges and chemically linking cationic charges to DNA. These modifications spontaneously bend DNA toward the neutralized or cationic volume (*49*−*51*). The self-repulsive ionic system of DNA collapses toward a neutral or cationic volume. On a local level, it is clear that localization of cations within the minor groove of AT DNA would decrease phosphate repulsion across the groove, closing down the groove width. Similarly, bending of DNA by AT tracts is explained by selective localization of cations. In fact, DNA should spontaneously bend around occupied cation binding sites within either groove. Our model for cation-induced DNA deformation is similar to a bending model recently proposed by Hud and Feigon (*42*) but differs from that model by the inclusion of a rationale for minor groove narrowing.

Magnesium in the Major Groove. A fully hydrated magnesium ion is located in the major groove of structure X (Figures 3 and 4). The magnesium hydration sphere interacts with hydrogen bond acceptors of the DNA but not with hydrogen bond donors. This limited range of interaction type is consistent with the high cationic charge, high hydrogen bond donating potential, and relatively high acidity of water molecules coordinating the dication. Elsewhere, we report that magnesium and other divalent cations interact not only with electron lone pairs but also with π-systems of nucleic acid bases (L. McFail-Isom, X. Shui, and L. D. Williams, unpublished) in a mode termed cation−π interaction (*52*). The magnesium ion located in the major groove of structure X pulls cytosine bases partially out from the helical stack, exposing π-systems to partial positive charges of the magnesium ion and its outer sphere.

Spermine in the Major Groove. Structure X provides the first high-resolution view of spermine bound to B-DNA. Drew and Dickerson attempted with tentative success to fit a spermine to a string of water molecules located at the other end of the structure 1BNA (*2*).

Spermine, the largest and most complex polyamine found in eukaryotes, forms polymorphic and dynamic complexes with nucleic acids. It interacts by a variety of competing and complementary interactions (electrostatic, hydrogen bonding, and hydrophobic). Spermine has been observed by X-ray diffraction bound to yeast-tRNA^{phe} (*53*), in the deep groove (*54*), or bound to the phosphate oxygen atoms (*55*) of A-DNA, bound in the major groove (*56*) or to the drug alone (*57*) of DNA−drug complexes, and bound to phosphate groups, the convex surface, and the deep groove of Z-DNA (*58*). Egli and co-workers found that a spermine molecule coordinates two phosphate groups adjacent to a catalytically critical A-bulge in the crystal structure of a chimeric, cleavage-inhibited DNA−RNA duplex (*59*).

In structure X, a portion of a spermine molecule is well-ordered and clearly can be observed in the electron density maps. Another portion is statistically disordered. The ordered portion of the spermine molecule is located in the major groove region but is well hydrated, with a single direct contact with the DNA. It has been suggested that spermine is required for crystallization but is either disordered or does not enter the crystal. This presumption is shown to be false by this report, as is a previous suggestion that dodecamers form intrinsically disordered crystals due to weak lattice interactions and/or DNA conformational polymorphism.

Effects of Sequence, Chemical Modification, and Coordinate Error on DNA Structure. Successful interpretation of X-ray crystallographic results requires concessions to experimental uncertainty. Like all experimental techniques, X-ray crystallography is affected by random and systematic errors. Differences are expected, and observed, in duplicate X-ray structures, independent determinations of the same molecule in the same crystal form. Comparisons of duplicate protein structures, or redundant protein molecules within an asymmetric unit, show coordinate errors of around 0.1 Å (*60*−*63*). The effects and magnitudes of errors in crystallographic structure determinations are often difficult to assess. We demonstrate here that valid estimates of structural error in a test structure can be obtained by comparison with a more accurate reference structure.

DNA dodecamers were initially crystallized and solved by Dickerson and co-workers (*1*−*8*). The dodecamer family in the PDB and NDB has expanded to 68 isomorphous B-conformation members of $[d(CGCXAATTYGCG)]_2$ (X = G or A, Y = C or T). The family contains mismatches, chemical modifications, and complexes with minor groove binders. On average, dodecamer crystals diffract to around 2.5 Å resolution with 2500 unique reflections (Figure 1). Structure X was refined against X-ray intensity data taken from a highly ordered specimen of the usual dodecamer crystal form. The unit cell is the same as that for previous dodecamers, and the structure are isomorphous; however, the resolution and quality of the diffraction pattern are radically different. The crystals diffract strongly to 1.4 Å resolution, giving around 3-fold more unique data than obtained previously from dodecamer crystals. The high quality of the data can be inferred from well-formed holes of electron density in the centers of five- and six-membered rings (Figure 2).

Structure X is consistent with the previous appraisal of minor groove widths. The minor groove of the AT tract is narrow, while the minor groove of the GC tracts is wide. In addition, it was concluded that the dodecamer AT tract is inherently straight and that the GC−AT junction is a flexible hinge. Dickerson and co-workers hypothesized that this hinge can be induced to bend under local forces, consistent with bending of one of the GC−AT junctions (the well-known dodecamer bend) by lattice forces but not the other. Structure X manifests the previously observed bend, localized near one AT−GC junction. The DNA is linear throughout the AT tract and the second GC tract.

We have performed an analysis of the degree of confidence with which structural polymorphism can be determined within a subset of the published dodecamer family of three-dimensional structures. The subset consists of structures 1BNA, 2BNA, 7BNA, and 9BNA, which are chemically

identical to structure X, and other duplicate structures that differ from structure X only by minor chemical modifications (Table 8).

It appears that some of the conformational heterogeneity that was previously attributed to DNA sequence in fact arises from experimental error. Structure X is conformationally more ordered and regular than previous dodecamer structures. Many helical parameters (shear, stretch, stagger, tilt, propeller twist, opening, and shift) show less deviation in the high-resolution structure than in nearly all duplicate, lower-resolution structures. Only for five parameters (buckle, slide, rise, roll, and twist) does the variation within structure X appear to be representative of the lower-resolution duplicate set. Therefore, for these five parameters, the observed variation probably results from intrinsic differences arising from sequence, environment, and chemical modification.

The extent of correlation between structures suggests that observed variation of shear, stretch, stagger, buckle, propeller twist, and opening is not intrinsic. Variation of slide does appear by this measure to be measurable and intrinsic at the resolution limits of previous dodecamers. For most helical parameters, contributions of the chemical modifications to variation are not greater than contributions from random coordinate error. For buckle, propeller twist, shift, and slide, the contribution of chemical modifications to variation is greater than contributions from other sources.

Significance of Uncertainty in DNA Structure. Many macromolecular crystallographers (us included) and structural databases violate a basic tenet of experimental science that "measurements require error bars" (adapted from ref *64*). Coordinates are commonly reported to 0.001 Å but in fact are much less accurate than that. The true accuracy of crystallographic structures is not generally communicated to or appreciated by other investigators. In one example, DNA dodecamer structures were used to suggest that differential propeller twists are correlated with probabilities of small molecule cleavage. Unfortunately, errors in propeller twists of published B-DNA dodecamers are sufficiently high that the reported correlation is likely to be specious.

More recently, a series of DNA decamers have been crystallized (*25*, *65*−*67*) that in some cases but not all (*68*) diffract to higher resolutions than previous dodecamer crystals. The decamer structures, like the dodecamers, contain errors that must be appreciated for proper interpretation.

ACKNOWLEDGMENT

We thank Drs. Jonathan Chaires, Martin Egli, Steven Neidle, Eaton Lattman, Horace Drew, and Laurence Wakelin for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Two figures showing the σ plot of structure X and the cross-validated Luzzati plots of structure X and simulated structures a−f (2 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Wing, R., Drew, H., Takano, T., Broka, C., Takana, S., Itakura, K., and Dickerson, R. E. (1980) *Nature 287*, 755−758.
2. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535−556.
3. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Itakura, K., and Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179−2183.
4. Drew, H. R., Samson, S., and Dickerson, R. E. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4040−4044.
5. Fratini, A. V., Kopka, M. L., Drew, H. R., and Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686−14707.
6. Kopka, M. L., Fratini, A. V., Drew, H. R., and Dickerson, R. E. (1983) *J. Mol. Biol. 163*, 129−146.
7. Holbrook, S. R., Dickerson, R. E., and Kim, S.-H. (1985) *Acta Crystallogr. B41*, 255−262.
8. Dickerson, R. E., Goodsell, D., and Kopka, M. L. (1996) *J. Mol. Biol. 256*, 108−125.
9. Bernstein, F. C., Koetzle, T. F., Williams, J. G. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
10. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R., and Schneider, B. (1992) *Biophys. J. 63*, 751−759.
11. Kan, Y., Armitage, B., and Schuster, G. B. (1997) *Biochemistry 36*, 1461−1466.
12. Laughton, C. A., Tanious, F., Nunn, C. M., Boykin, D. W., Wilson, W. D., and Neidle, S. (1996) *Biochemistry 35*, 5655−5661.
13. Brunger, A. T. (1996) *Methods Mol. Biol. 56*, 245−266.
14. Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K., and Berman, H. M. (1996) *J. Am. Chem. Soc. 118*, 509−518.
15. Lavery, R., and Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655−667.
16. Dickerson, R. E., Goodsell, D. S., and Neidle, S. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3579−3583.
17. Heinemann, U., Alings, C., and Hahn, M. (1994) *Biophys. Chem. 50*, 157−167.
18. Hartmann, B., and Lavery, R. (1996) *Q. Rev. Biophys. 29*, 309−368.
19. Berman, H. M. (1997) *Biopolymers 44*, 23−44.
20. Brown, I. (1992) *Acta Crystallogr. B48*, 553−572.
21. Nayal, M., and Di Cera, E. (1996) *J. Mol. Biol. 256*, 228−234.
22. Waltersson, K. (1978) *Acta Crystallogr. A43*, 901−905.
23. Timsit, Y., and Moras, D. (1992) in *Methods in Enzymology: DNA Structures: Part A Synthesis and Physical Analysis of DNA* (Lilley, D. M. J., and Dahlberg, J. E., Eds.) Vol. 211, pp 409−429, Academic Press, Inc., San Diego.
24. Luzzati, V. (1952) *Acta Crystallogr. 5*, 802−810.
25. Prive, G. G., Yanagi, K., and Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 201−214.
26. Bingman, C., Jain, S., Zon, G., and Sundaralingam, M. (1992) *Nucleic Acids Res. 20*, 6637−6647.
27. Young, M. A., Ravishanker, G., Beveridge, D. L., and Berman, H. M. (1995) *Biophys. J. 68*, 2454−2468.
28. McAuley-Hecht, K. E., Leonard, G. A., Gibson, N. J., Thomson, J. B., Watson, W. P., Hunter, W. N., and Brown, T. (1994) *Biochemistry 33*, 10266−10270.
29. Lipscomb, L. A., Peek, M. E., Morningstar, M. L., Verghis, S. M., Miller, E. M., Rich, A., Essigmann, J. M., and Williams, L. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 719−723.
30. Brown, T., Hunter, W. N., Kneale, G., and Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2402−2406.
31. Brown, T., Leonard, G. A., Booth, E. D., and Chambers, J. (1989) *J. Mol. Biol. 207*, 455−457.
32. Chen, X., Ramakrishnan, B., Rao, S. T., and Sundaralingam, M. (1994) *Nat. Struct. Biol. 1*, 169−175.
33. Boykin, D. W., Kumar, A., Spychala, J., Zhou, M., Lombardy, R. J., Wilson, W. D., Dykstra, C. C., Jones, S. K., Hall, J. E., and Tidwell, R. R. (1995) *J. Med. Chem. 38*, 912−916.
34. Goodsell, D. S., Kopka, M. L., and Dickerson, R. E. (1995) *Biochemistry 34*, 4983−4993.
35. Lipscomb, L. A., Peek, M. E., Zhou, F. X., Bertrand, J. A., VanDerveer, D., and Williams, L. D. (1994) *Biochemistry 33*, 3649−3659.

36. Peek, M. E., Lipscomb, L. A., Bertrand, J. A., Gao, Q., Roques, B. P., Garbay-Jaureguiberry, C., and Williams, L. D. (1994) *Biochemistry 33*, 3794−3800.
37. Lipscomb, L. A., Zhou, F. X., Presnell, S. R., Woo, R. J., Peek, M. E., Plaskon, R. R., and Williams, L. D. (1996) *Biochemistry 35*, 2818−2823.
38. Hu, G. G., Shui, X., Leng, F., Priebe, W., Chaires, J. B., and Williams, L. D. (1997) *Biochemistry 36*, 5940−5946.
39. Ramakrishnan, B., and Sundaralingam, M. (1993) *Biochemistry 32*, 11458−11468.
40. Schneider, B., and Berman, H. M. (1995) *Biophys. J. 69*, 2661−2669.
41. Young, M. A., Jayaram, B., and Beveridge, D. L. (1997) *J. Am. Chem. Soc. 119*, 59−69.
42. Hud, N. V., and Feigon, J. (1997) *J. Am. Chem. Soc. 119*, 5756−5757.
43. Marky, L. A., and Kupke, D. W. (1989) *Biochemistry 28*, 9982−9988.
44. Sidorova, N. Y., and Rau, D. C. (1995) *Biopolymers 35*, 377−384.
45. Laundon, C. H., and Griffith, J. D. (1987) *Biochemistry 26*, 3759−3762.
46. Zinkel, S. S., and Crothers, D. M. (1987) *Nature 328*, 178−181.
47. Koo, H. S., Drak, J., Rice, J. A., and Crothers, D. M. (1990) *Biochemistry 29*, 4227−4234.
48. Mirzabekov, A. D., and Rich, A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1118−1121.
49. Strauss, J. K., and Maher, L. J. (1994) *Science 266*, 1829−1834.
50. Strauss, J. K., Roberts, C., Nelson, M. G., Switzer, C., and Maher, L. J. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 9515−9520.
51. Strauss-Soukup, J. K., Vaghefi, M. M., Hogrefe, R. I., and Maher, L. J. (1997) *Biochemistry 36*, 8692−8698.
52. Ma, J., and Dougherty, D. (1997) *Chem. Rev. 97*, 1303−1324.
53. Quigley, G. J., Teeter, M. M., and Rich, A. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 64−68.
54. Jain, S., Zon, G., and Sundaralingam, M. (1989) *Biochemistry 28*, 2360−2364.
55. Ban, C., Ramakrishnan, B., and Sundaralingam, M. (1994) *Nucleic Acids Res. 22*, 5466−5476.
56. Frederick, C. A., Williams, L. D., Ughetto, G., van der Marel, G. A., van Boom, J. H., Rich, A., and Wang, A. H.-J. (1990) *Biochemistry 29*, 2538−2549.
57. Smith, C. K., Brannigan, J. A., and Moore, M. H. (1996) *J. Mol. Biol. 263*, 237−258.
58. Bancroft, D., Williams, L. D., Rich, A., and Egli, M. (1994) *Biochemistry 33*, 1073−1086.
59. Portmann, S., Grimm, S., Workman, C., Usman, N., and Egli, M. (1996) *Chem. Biol. 3*, 173−184.
60. Chambers, J. L., and Stroud, R. M. (1979) *Acta Crystallogr. B33*, 1824−1827.
61. Chothia, C., and Lesk, A. M. (1986) *EMBO J. 5*, 823−826.
62. Wlodawer, A., Borkakoti, N., Moss, D. S., and Howlin, B. (1986) *Acta Crystallogr. B42*, 379−387.
63. Stroud, R. M., and Fauman, E. B. (1995) *Protein Sci. 4*, 2392−2404.
64. Janin, J. (1990) *Biochimie 72*, 705−709.
65. Grzeskowiak, K., Yanagi, K., Prive, G. G., and Dickerson, R. E. (1991) *J. Biol. Chem. 266*, 8861−8883.
66. Heinemann, U., and Alings, C. (1991) *EMBO J. 10*, 35−43.
67. Quintana, J. R., Grzeskowiak, K., Yanagi, K., and Dickerson, R. E. (1992) *J. Mol. Biol. 225*, 379−395.
68. Heinemann, U., Alings, C., and Bansal, M. (1992) *EMBO J. 11*, 1931−1939.
69. Read, R. J. (1986) *Acta Crystrallogr. A42*, 140−149.
70. Westhof, E. (1987) *J. Biomol. Struct. Dyn. 5*, 581−600.
71. Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H., and Rich, A. (1988) *J. Biol. Chem. 263*, 17872−17879.
72. Portmann, S., Altmann, K.-H., Reynes, N., and Egli, M. (1997) *J. Am. Chem. Soc. 119*, 2396−2403.

BI973073C