Comparison of Major Groove Hydration in Isomorphous A-DNA Octamers and Dependence on Base Sequence and Local Helix Geometry†

D. B. Tippin and M. Sundaralingam*

Laboratory of Biological Macromolecular Structure, Newman & Wolfram Laboratory of Chemistry, Department of Biochemistry, The Ohio State University, 1060 Carmack Road, Columbus, Ohio 43210-1002

Received June 24, 1996; Revised Manuscript Received November 7, 1996⊗

ABSTRACT: The family of ten isomorphous tetragonal A-DNA octamers provides a unique opportunity to examine major groove hydration in terms of base sequence and local parameter effects. The presence of a severely underwound central py·pu base step (average = 24.1°), which lies on a crystallographic 2-fold in the unit cell, provides a sharp change in the local environment in which to study and separate the effects of base sequence and local helix geometry on major groove hydration. For this reason, and to avoid bias secondary to end effects, hydration analysis was restricted to the central four dyad-related base pairs. This study finds that d(CG) base pairs are better hydrated than d(TA) base pairs, 2.5 H_2O vs 1.3 H_2O; steps with high twist angles are better hydrated than steps with low twist angles, 6.9 H_2O vs 0 H_2O; negative roll angles are better hydrated than positive roll angles, 2.8 H_2O vs 1.8 H_2O; and flanking base pairs are better hydrated than central base pairs, 2.6 H_2O vs 2.0 H_2O, a phenomenon which is sequence independent, occurring for both d(CG) and d(TA) base pairs. The twist angle and base roll combine to significantly affect the pattern and degree of major groove hydration in this family of octamers. A previous study of A-DNA octamers and their helix parameters established a strong dependency on crystal packing forces with little or no dependence on the base sequence [Ramakrishnan & Sundaralingam (1993) *J. Biomol. Struct. Dyn. 11*, 11−26]. We find that the degree and pattern of major groove hydration are strongly influenced by the local helix parameters, implying an indirect, but significant, relationship between major groove hydration and environmental forces, i.e., crystal packing, drug binding, and protein−DNA interactions.

Most crystallographic hydration analyses have focused on minor groove hydration in B and Z forms of DNA. Few studies have examined major groove hydration, and fewer still have examined the hydration of A-DNA. Chuprina *et al*. (1991) have shown that the positions of solvent molecules in the minor groove of B-DNA depend principally on groove width. Wide minor grooves promote hydration of individual bases by solvent molecules lying in the plane of the base to which they are bound, while narrow grooves promote a single spine of hydration such that water molecules lie midway between the planes of the base pairs to which they are bound. For the most part, the positions of solvent molecules in the minor groove of B-DNA are independent of base sequence. Only the location of the hydrogen atoms in the hydrogen-bonding network appears to be affected by base sequence. Other studies have established hydration patterns for the minor groove in Z-DNA, where water molecules connect adjacent cytosine O2 keto groups and bridge guanine N2 nitrogens with adjacent phosphate groups (Gessner *et al*., 1994; Ban *et al*., 1996). Discussion of minor groove hydration in A-DNA crystal structures is difficult: crystal packing, where the duplex termini butt into the minor grooves of neighboring duplexes, displaces most of the water molecules which had occupied the groove when the DNA was in solution.

DNA binding studies have provided ample evidence for the importance of primary shell hydration in DNA recognition and behavior. Shakked and co-workers find major groove hydration in DNA to be crucial for recognition and binding by the *trp* repressor, and they identify 10 conserved hydration sites in the major groove of the *trp* operator (Shakked *et al*., 1994). Netropsin binding to the minor groove in B-DNA releases 22−53 water molecules from the hydration shell of AT tracts, and the number of water molecules released depends on the particular sequence involved (Chalikian *et al*., 1994). On the basis of crystallographic analysis and computer simulations, Boehncke *et al*. suggest that the tight hydration shell covering the phosphate backbone is instrumental in preventing distamycin from binding in the major groove of duplex DNA (Boehncke *et al*., 1991). Rau *et al*. hypothesize, on the basis of direct measurement of DNA helices, that multivalent counterion adsorption, i.e., polyamines and divalent cations, reorganizes the water at distinct sites on the surface of DNA in a pattern complementary to unadsorbed sites on the opposing surface of a neighboring duplex and, thus, creates attractive long-range hydration forces (Rau *et al*., 1992). This novel hypothesis is in contrast to the more active role polyamines and divalent cations are thought to play in DNA crystallization and condensation through direct bridging and/or modification of electrostatic interactions between neighboring duplexes. There can be little doubt that knowledge of solvent−DNA interaction is crucial for studying the dynamics of protein−DNA recognition and binding.

† This work was supported by National Institutes of Health Grant GM-17378 and the Ohio Eminent Scholar Endowment.

* Address correspondence to this author. Phone: (614) 292-2925. Fax: (614) 292-2524. E-mail: sunda@biot.mps.ohio-state.edu.

⊗ Abstract published in *Advance ACS Abstracts,* January 1, 1997.

S0006-2960(96)01519-X CCC: $14.00

The role of hydration in A-DNA has been addressed in two recent papers (Eisenstein & Shakked, 1995; Tippin & Sundaralingam, 1996) which discuss the importance of water-mediated interactions in stabilizing interduplex interactions. Tippin and Sundaralingam group these interactions into four distinct classes that are solely dependent on the base sequence. Eisenstein and Shakked point out similarities in these direct and water-mediated interactions with the recognition motifs of protein−DNA complexes like the TATA binding protein and the TATA box. Through the comparative analysis of five A-DNA octamers, they also demonstrate that the major groove in A-DNA is extensively hydrated and can form an ordered network of fused polygons with the hydration shell of the sugar−phosphate backbone. This is consistent with the crystal structure of another A-DNA octamer, GGTATACC, and its 5-bromouracil analogue which contain fused pentagons of water molecules in the major groove and resemble clathrate-type structures (Kennard *et al*., 1986). In a review of A-DNA crystal structures, an average of 10 water molecules are reported per base pair (Wahl & Sundaralingam, 1996). The vast majority of these belong to the first hydration shell with 30% residing in the major groove and 20% in the minor groove, while the remainder interact with the sugar−phosphate backbone. Water molecules in the major and minor grooves are significantly more conserved than those associated with the sugar−phosphate backbone.

Table 1 lists 10 sequences which make up the tetragonal A-DNA family of octamers. The conformation of these octamers was shown previously to be predominantly influenced by crystal lattice interactions, with only a minor dependence on base sequence (Ramakrishnan & Sundaralingam, 1993). Among the helical parameters of the examined structures, only the propeller twist and base pair buckle display any sequence dependency; the twist, rise, slide, inclination, tilt, and roll can be readily explained by similarities in the crystal packing forces. For truly isomorphous DNA structures whose helical parameters are mostly dependent on lattice interactions, it stands to reason that any variation in major groove hydration should be the result of differences in base sequence. Starting with this basic but incomplete premise, we attempted to elucidate the relationship between major groove hydration and base sequence in the tetragonal A-DNA octamers but instead found a strong correlation with local helix geometry as well.

EXPERIMENTAL PROCEDURES

The tetragonal family of A-DNA octamers was chosen for hydration analysis because it contains the most isoforms of any DNA family and the greatest number of structures with reported solvent coordinates. Solvent coordinates for eight tetragonal A-DNA octamers were retrieved from the Nucleic Acid Database (Berman *et al*., 1992) while coordinates for one other, d(GTGCGCAC), was retrieved from the Brookhaven Protein Data Bank (Bernstein *et al*., 1977). Of these, one analysis reports only three independent water molecules, d(GTCTAGAC), and a second analysis reports an ordered spermine interacting with the floor of the major groove of d(GTGTACAC) and interfering with base hydration. These structures were, therefore, excluded from the analysis. In all, eight independent sequences were chosen for solvent analysis: d(CTCTAGAG), d(CCCTAGGG), d(CCCCGGGG), d(GCCCGGGC), d(GGCCGGCC), d(GGGCGCCC), d(GTGCGCAC), and d(GTACGTAC) (Table 1). Coordinates for d(CCCTAGGG) were obtained from previous work (Tippin & Sundaralingam, 1996). The central four base pairs in each structure, which lie on a crystallographic dyad, were chosen for the initial hydration study because of the unique helix parameters in this region (i.e., very low average twist angle of 24.1°) and to avoid possible bias from end effects. In the analysis of d(CCCTAGGG) peak positions were identified as solvent molecules if they were greater than 2 σ in an $F_o - F_c$ density map and they met reasonable criteria for proper angles and bond distances, i.e., less than 3.4 Å. The omit map was calculated by omitting all the solvent molecules in our model. For the other seven independently refined sequences, published coordinates were used and solvent molecules within 3.4 Å of other polar atoms were considered to be hydrogen bonded. Duplex structures were analyzed on a Silicon Graphics Indigo 2 Extreme using the graphics program CHAIN (Sack & Spurlino, 1991). Solvent accessibilities were calculated using the program X-PLOR (Brunger *et al*., 1987).

Table 1: Tetragonal A-DNA Octamers, $P4_32_12$

sequence	$a = b$ (Å)	c (Å)	resoln (Å)	no. of H_2O[a]	reference
CCCTAGGG	42.2	24.9	1.9	76	Tippin *et al.* (1996)
CTCTAGAG	42.5	24.3	2.2	34	Hunter *et al.* (1987)
CCCCGGGG	43.4	24.8	2.2	78	Haran *et al.* (1987)
GCCCGGGC	43.3	24.6	1.8	34	Heinemann *et al.* (1987)
GGCCGGCC	42.1	25.2	2.2	80	Wang *et al.* (1982)
GGGCGCCC	43.3	24.7	1.7	87	Eisenstein *et al.* (1988)
GTGCGCAC	42.2	25.1	1.6	84	Bingman *et al.* (1992)
GTACGTAC	42.5	24.8	2.0	51	Courseille *et al.* (1990)
GTGTACAC	42.4	24.8	2.0	86	Jain *et al.* (1989)
GTCTAGAC	42.6	24.6	2.5	6	Cervi *et al.* (1992)

[a] The number of water molecules listed is the number per duplex. The last two sequences were excluded from this analysis.

RESULTS

Twist Parameter. The most striking feature of DNA hydration within this family of octamers appears to be correlated with the helical twist parameter. The tetragonal A-DNA octamers possess very low twist angles at their central base steps and are flanked by base steps with higher than average twist angles (Table 2). With only two exceptions, all octamers in this family lack solvent molecules between the base planes of nucleotides 4 and 5, and in both these exceptions, d(CCCCGGGG) and d(GTGCGCAC), the modeled water molecule lies on the crystallographic 2-fold. As such, map densities may arise from the concentration of noise on the dyad axis or a monovalent cation and not from the actual presence of an ordered water molecule. Furthermore, the position of the special site water molecule in d(GTGCGCAC) is a second sphere water, making no contacts with any nearby base (Figure 3). In d(CCCCGGGG) the same water molecule lies between 2.6 and 3.0 Å from potential hydrophilic sites on the central base pairs and approximately 2.8 Å from two other primary sphere water molecules (Figure 2). Except for the nonoctahedral geometry, this special site water behaves very much like a metal cation.

The average twist angle for the central base step in this family is 24.1°, which should be compared to the average twist angle for the entire helix, 32°. The average twist angle

Table 2: Relationship among Twist Angle, Roll Angle, Solvent-Accessible Area, and Base Hydration[a]

	center BP			flanking BP			waters/step		
sequence	twist	roll	SAA	twist	roll	SAA	central	flanking	total
CTCTAGAG	20.8	2.54	35.8	40.8	−1.51	43.4	0	4	4
CCCTAGGG	28.4	1.70	31.4	36.7	−4.06	42.3	4	6	10
CCCCGGGG	25.4	−5.95	51.4	35.6	0.06	47.0	6	6	12
GCCCGGGC	23.2	−3.33	55.7	33.6	−0.81	41.8	6	6	12
GGCCGGCC	24.2	2.75	41.4	37.5	−2.09	39.9	6	6	12
GGGCGCCC	23.9	0.00	50.7	34.8	0.18	43.2	4	4	8
GTGCGCAC	25.7	4.19	44.0	35.7	−1.99	33.0	4	6	10
GTACGTAC	23.0	4.28	57.7	36.0	−1.72	25.2	2	4	6
av	24.1	0.77[b]	46.0	36.3	−1.58	39.4	4	5.3	
av/BP							2	2.6	
av/BP (−roll)[c]							3	2.7	2.8
av/BP (+roll)							1.6	2.0	1.8

[a] Twist and roll are in degrees. BP = base pairs. SAA = solvent-accessible area in Å^2. The solvent-accessible area was calculated using the program X-PLOR for central and flanking base pairs by adding the solvent-accessible areas of the major groove nitrogen and oxygen atoms. Waters/step are calculated as the number residing between adjacent base planes within hydrogen-bonding distance of a polar base atom. [b] Average value for the central base roll increases to 2.58° if d(CCCCGGGG) and d(GGCCGGCC) are excluded from the calculation. [c] The last two rows list the number of solvent molecules for both negative and positive base rolls per pase pair.

Table 3: Dependence of Hydration on Base Type and Local Helix Parameters[a]

	no. of H_2O/bp		no. of H_2O/bp		no. of H_2O/bp	ΔH_2O/bp	% Δ
av TA	1.33	central TA	1.00	flanking TA	2.00	1.0	100
av CG	2.54	central CG	2.33	flanking CG	2.71	0.38	16
ΔH_2O/bp	1.21		1.33		0.71		
% Δ	91		133		36		

[a] bp = base pair; av = average value. This table lists the average hydration for TA and CG base pairs. The no. of H_2O/bp is calculated as the number of DNA-bound solvent molecules within hydrogen-bonding distance of a base pair polar atom. Rows show the differences in hydration between the same type of base pair at two different helix locations (central and flanking). Columns show the differences in hydration between the two types of base pairs (TA vs CG) at the same helix location. Both the absolute difference (ΔH_2O/bp) and relative percent difference (% Δ) in hydration are listed.

of the flanking base steps is 36.3°, 4° higher than the overall average and 12° higher than that of the central base steps. Solvent molecules are absent in the central base step for both AT and CG base steps, and this absence appears to be independent of base sequence. The flanking base steps, however, average 6.9 solvent molecules per base step, including first and second shell waters, and are similarly independent of base sequence. Incidentally, intercalating agents such as actinomycin underwind duplex DNA by 10−20° at their point of intercalation, the same range of underwinding exhibited by the central base step in the tetragonal A-DNA octamers.

Roll Parameter. There appears to be a strong correlation between the base roll and the number of hydrogen bonds formed with solvent molecules. In general, positive rolls, i.e., bases open toward the minor groove, establish fewer solvent interactions, averaging 1.8 H_2O per base pair, while negative base rolls allow for more solvent interactions, averaging 2.8 H_2O per base pair. In this family of octamers the roll tends to be positive for the central base step and negative for the flanking base steps. The average base roll for the central region is 0.77°, which increases to 2.58° if d(CCCCGGGG) and d(GCCCGGGC) are excluded, while the average base roll for the flanking base pairs is −1.58° (Table 2). In other words, the flanking base pairs open toward the major groove, while the central base pairs open towards the minor groove. Dickerson and co-workers (Yanugi *et al*., 1991) have identified two similar step profiles in B-DNA. One profile exhibits a high twist, low rise, and negative base roll (high twist profile). The other profile exhibits a low twist, high rise, and positive base roll (low twist profile).

Sequence Effects. Any attempt to elucidate the sequence dependency of base hydration requires separating the effects due to helical parameters, i.e. DNA conformation, from those due to the particular base pairs involved. This is particularly difficult to quantify statistically because of the limited number of sequences, which is large in terms of available isomorphous structures but small from a statistical perspective. In this particular family of DNA, the average TA base pair binds 1.33 solvent molecules and the average CG base pair binds 2.5 solvent molecules, which correlates well with solvent accessibility predictions (Table 4). Central TA base pairs bind an average of 1.0 solvent molecule, while central CG base pairs bind an average of 2.3 solvent molecules. For the flanking base pairs, the average TA pair binds 2.0 solvent molecules and the average CG pair binds 2.7 solvent molecules. From Table 3, it is seen that CG base pairs are 91% more hydrated than TA base pairs. However, central CG base pairs are 133% more hydrated than central TA base pairs, while flanking CG base pairs are only 36% more hydrated than flanking TA base pairs. It would appear that either the low twist angle and/or the increased positive roll (LTP) of the central base pair increases the differences between the solvent binding potentials of CG and TA base pairs, while the high twist angle and/or the increased negative roll (HTP) of the flanking base pairs affect(s) smaller differences. After comparing the solvent binding differences between CG and TA base pairs for the central and flanking steps, it is necessary to compare the solvent binding

Table 4: Dependence of Solvent-Accessible Area on Base Type and Local Helix Parameters[a]

	av SAA (Å^2)		av SAA (Å^2)		av SAA (Å^2)	ΔSAA	% Δ
av TA	15.3	central TA	16.8	flanking TA	12.6	4.2	33
av CG	23.1	central CG	25.1	flanking CG	20.7	4.4	21
ΔSAA/bp	7.8		8.3		8.1		
% ΔSAA	51		49		64		

[a] SAA = solvent-accessible area; bp = base pair; av = average value. This table lists the average solvent-accessible areas for TA and CG base pairs. Rows show the differences between the same type of base pair at two different helix locations (central and flanking). Columns show the differences between the two types of base pairs (TA vs CG) at the same helix location. Both the absolute (ΔSAA) and relative percent difference (% Δ) in solvent-accessible area are listed.

differences between the central and flanking base steps for CG and TA base pairs. From Table 3, it is clear that flanking TA pairs are 100% more hydrated than central TA pairs, while flanking CG pairs are only 16% more hydrated than central CG pairs, suggesting that the solvation of TA base pairs is more susceptible to changes in helical parameters than is the solvation of CG base pairs. To determine the differential contribution of the base sequence and helical parameters to base hydration, a simple comparative ratio may be used (Table 3):

$$\text{sequence-dependent } \Delta = [(\text{diff between central TA and CG } H_2O) + (\text{diff between flanking TA and CG } H_2O)]/2 = (1.33 + 0.71)/2 = 1.02\ H_2O \text{ per base pair} \quad (1)$$

$$\text{parameter-dependent } \Delta = [(\text{diff between central and flanking TA } H_2O) + (\text{diff between central and flanking CG } H_2O)]/2 = (1.00 + 0.38)/2 = 0.69\ H_2O \text{ per base pair} \quad (2)$$

where the values in parentheses correspond to either the sum of the differences between the two types of base pairs (TA and CG) at the same helix location as in eq 1 or the sum of the differences between the same type of base pair at two different helix locations (central and flanking) as in eq 2. The dissimilar results tell us that sequence effects have a slightly greater influence on step hydration than does the actual change in helix parameters. Accordingly, the base sequence is responsible for 68% of the difference seen in base pair hydration (Δ 1.02 H_2O/base pair), and the helical parameters are responsible for the other 32% (Δ 0.69 H_2O/base pair). This contribution of base sequence and helical geometry applies only to the helical parameters exhibited by the central four bases of the tetragonal A-DNA octamers. The ratio will undoubtedly change when base steps of other sequences with helical parameters different from those analyzed in this family are compared.

Solvent Accessibility. Solvent accessibility of the four central base pairs was calculated for a sphere with a diameter of 2.8 Å using the Lee and Richards algorithm in the program X-PLOR. Table 2 shows the accessible surface area of donors (N atoms) and acceptors (O atoms) in the major groove of the studied sequences. Values were calculated for O6 and N7 atoms in guanines, N6 and N7 atoms in adenines, the N4 atom in cytosines, and the O4 atom in thymines. Analysis established an inverse relationship between the twist angle and the solvent-accessible area of the major groove, for the central four base pairs. As the twist angle decreases, the solvent accessibility of a given base pair increases. This becomes more apparent if sequences possessing a central TA step are analyzed separately from those possessing a central CG step since, intrinsically, CG base pairs display larger solvent-accessible areas in duplex DNA and are, therefore, expected to interact with more solvent molecules. For AT base pairs the C5 methyl of thymine protrudes into the major groove discouraging the approach of polar solvent molecules. The average solvent-accessible area of the central base pairs is somewhat higher than that of the flanking base pairs, 46.0 versus 39.4 Å^2. Despite their smaller solvent-accessible area, the flanking base pairs form more hydrogen bonds with solvent molecules than do the central base pairs, binding an average of 2.6 water molecules per base pair compared to an average of 2.0 water molecules per base pair for the central bases. By some unidentified mechanism, the low central twist angle, which results in greater solvent accessibility, disrupts hydrogen-bonding contacts between nucleotide bases and solvent molecules. This is quite apparent with d(CTCTAGAG), whose central base step has the lowest reported twist angle (20.8°) of the tetragonal A-DNA octamers, displays a higher solvent-accessible surface area than the central TA base pairs in d(CCCTAGGG), and, yet, forms no hydrogen bonds with neighboring solvent molecules. Its flanking base pairs display the highest twist angle in the family (40.8°) and, consequently, are much better hydrated (4 H_2O) than the central base pairs. d(CCCTAGGG), on the other hand, exhibits the highest central twist angle of any reported A-DNA octamer (28.4°), the lowest-solvent-accessible area found for a central base step (31.4 Å^2), and is quite well hydrated on both sides of the central base step (4H_2O) (Table 2).

A similar analysis to the one used in studying the differential contribution of the base sequence and helix parameters on major groove hydration can be applied to deciphering the paradoxical relationship between the solvent-accessible area and base hydration. Table 4 shows that the average solvent-accessible area is 15.3 Å^2 for TA base pairs and 23.1 Å^2 for CG base pairs, which is consistent with the relative degrees of hydration for TA base pairs (1.3 H_2O/base pair) and CG base pairs (2.5 H_2O/base pair). At the central base step the average solvent-accessible area is 16.8 Å^2 for TA base pairs and 25.1 Å^2 for CG base pairs. Flanking TA base pairs display an average solvent-accessible area of 12.6 Å^2, while flanking CG base pairs display an average solvent-accessible area of 20.7 Å^2. Statistically speaking, central CG base pairs possess 49% more solvent-accessible area than do central TA base pairs, while flanking CG base pairs possess 64% more solvent-accessible area than do flanking TA base pairs. Overall, the average CG base pair displays 51% more solvent-accessible area than does

Table 5: Summary of Quantitative Hydration Results in the Tetragonal Family of A-DNA Octamers

parameter	condition	hydration results (no. of waters/base pair)
twist	high twist	more solvent molecules found in base step (flank = 6.9)
	low twist	fewer solvent molecules found in base step (center = 0)
roll	positive roll	fewer hydrogen bonds with the solvent (1.8/base pair)
	negative roll	more hydrogen bonds with the solvent (2.8/base pair)
base pair	CG	more hydrogen bonds with the solvent (2.5/base pair)
	TA	fewer hydrogen bonds with the solvent (1.3/base pair)
location	central base pair	fewer hydrogen bonds with the solvent (2.0/base pair)
	flanking base pair	more hydrogen bonds with the solvent (2.6/base pair)

the average TA base pair. Solvent-accessible differences are considerably less when flanking base pairs are compared to central base pairs. Central TA base pairs possess 33% more solvent-accessible area than do flanking TA base pairs, 16.8 versus 12.6 Å^2, while central CG base pairs possess 21% more solvent-accessible area than do flanking CG base pairs, 25.1 versus 20.7 Å^2. Using the previously mentioned formula and the values listed in Table 4, two equations can be constructed to describe the dependence of solvent accessability on helix parameters and base sequence:

$$\Delta \text{ due to helical parameters:} \quad \Delta(4.2 + 4.4)/2 = \Delta 4.3 \text{ Å}^2 \tag{3}$$

$$\Delta \text{ due to sequence:} \quad \Delta(8.3 + 8.1)/2 = \Delta 8.2 \text{ Å}^2 \tag{4}$$

where the value in parentheses refers to the sum of the differences in surface area between the same type of base pair at different base steps (central versus flanking) in eq 3 or the sum of the differences in surface area between different types of base pairs (CG versus TA) at the same helix location in eq 4. Helical parameters account for 4.3 Å^2 of the difference between the average central and flanking base pair solvent-accessible area. The change in base type accounts for 8.2 Å^2, translating into 91% more solvent-accessible area than accounted for by the helical parameters. At least for the central four base pairs, the solvent-accessible area is more dependent on base sequence than on helix parameters. The degree of hydration, on the other hand, is more dependent on the helix parameters than is the surface area, offering an explanation as to why the degree of hydration can be inversely proportional to the solvent-accessible area for the central four base pairs. With similar helical parameters, the degree of hydration should be proportional to the solvent-accessible area of the base pairs and, hence, to the types of bases involved. However, given the appropriate helical geometry, as is apparently the case with the central four bases of the tetragonal A-DNA octamers, the strong correlation between solvent-accessible area and base solvation can be weakened.

DISCUSSION

A summary of the quantitative results is provided in Table 5. Factors that promote local major groove hydration are high twist angles, negative base rolls, and CG base pairs. In duplex DNA, low twist angles increase the interstrand phosphate−phosphate distances, thereby opening the major groove and exposing the floor to more solvent molecules. Bound water molecules can then be more readily displaced by free solvent molecules from the solution. The increased solvent exchange associated with lower twist angles can smear out the electron density of major groove hydration so that it appears as background noise in the electron density maps. Such an explanation is also consistent with a less stable region of local DNA structure.

Whereas the twist quantitatively affects the solvent-accessible area of a particular base step, the base roll appears to affect the distribution of the solvent-accessible area in the major groove and, hence, the directions of hydrogen bonding with solvent molecules. Examination of the flanking base pairs reveals that higher twist angles expose less solvent-accessible area from the polar atoms in the major groove (Table 2). The flanking base pairs, however, exhibit a negative roll (HTP) and, thus, open toward the major groove. Stacked base pairs that open toward the major groove would expose more surface area in the direction of their midline, encouraging the formation of hydrogen bonds with solvent molecules that lie within the base step. The less hydrated central base step exhibits a positive base roll (LTP), opening toward the minor groove and, thus, exposing less solvent area for formation of hydrogen bonds in the major groove (Figures 1−4).

Both quantitatively and qualitatively, major groove hydration is very much dependent on local helix geometry. This is exquisitely demonstrated by the complete absence of water molecules in the central base step of the tetragonal A-DNA octamers, which cannot be attributed to the presence of a crystallographic 2-fold nor to sequence effects—there are four different base combinations represented in this study. Inspection of the major grooves in d(CTCTAGAG) and d(CCCTAGGG) reveals markedly different degrees and patterns of hydration for the central four bases even though they are identical in sequence. Examination of Table 2 shows that d(CTCTAGAG) and d(CCCTAGGG) possess different helix parameters. d(CTCTAGAG) has the lowest central and the highest flanking twist angles in the tetragonal A-DNA family of octamers. d(CCCTAGGG) lies at the other extreme, possessing the highest central twist angle in the family. Further corroboration for the predominance of local helix geometry in major groove hydration is provided by comparison of d(CCCTAGGG) and d(GGCCGGCC). Despite differing base sequences, they display remarkably similar patterns of hydration in the major groove of the central four base pairs, and this similarity extends to secondary shell water molecules. Under similar conditions, d(GGCCGGCC) would be expected to exhibit a higher degree of hydration due to its higher CG content. d(CCCTAGGG), however, has the highest central twist angle and the largest flanking negative rolls in this family (Table 2). This study has shown that higher twist angles and negative base rolls foster major groove hydration and, in this particular case, allow d(CTAG) to achieve the same degree and pattern of major groove hydration as that found for d(CCGG) (Figures 1 and 2).

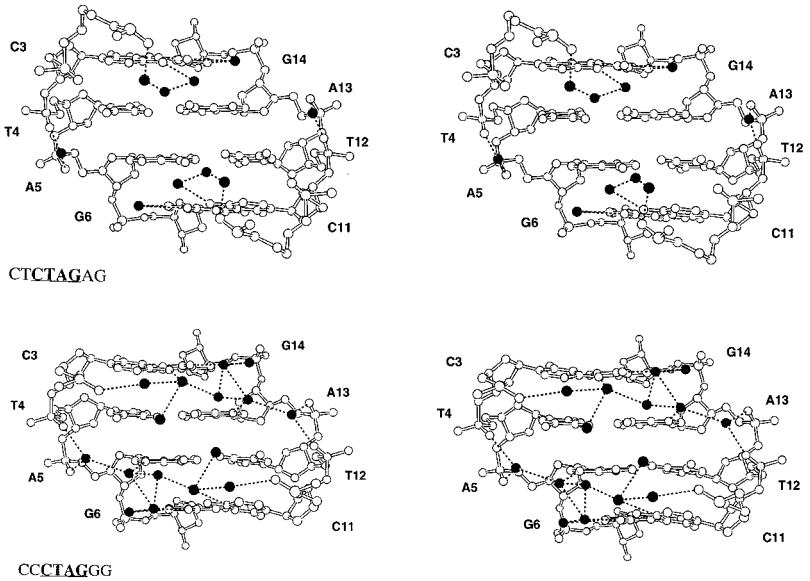

FIGURE 1: Stereoviews of major groove hydration for the central four base pairs in dCTCTAGAG) and d(CCCTAGGG). All figures have been aligned with respect to their backbone phosphorus atoms. Both dyad-related flanking base steps are shown to emphasize the absence of solvent molecules in the central base step. d(CTAG). d(CTCTAGAG) and d(CCCTAGGG) clearly demonstrate the importance of local helix geometry in major groove hydration. Based solely on sequence identity, these two octamers should exhibit identical patterns of hydration in the region of the major groove defined by their central four bases, but the higher twist and negative roll angles in d(CCCTAGGG) lead to higher degrees of major groove hydration (Table 2).

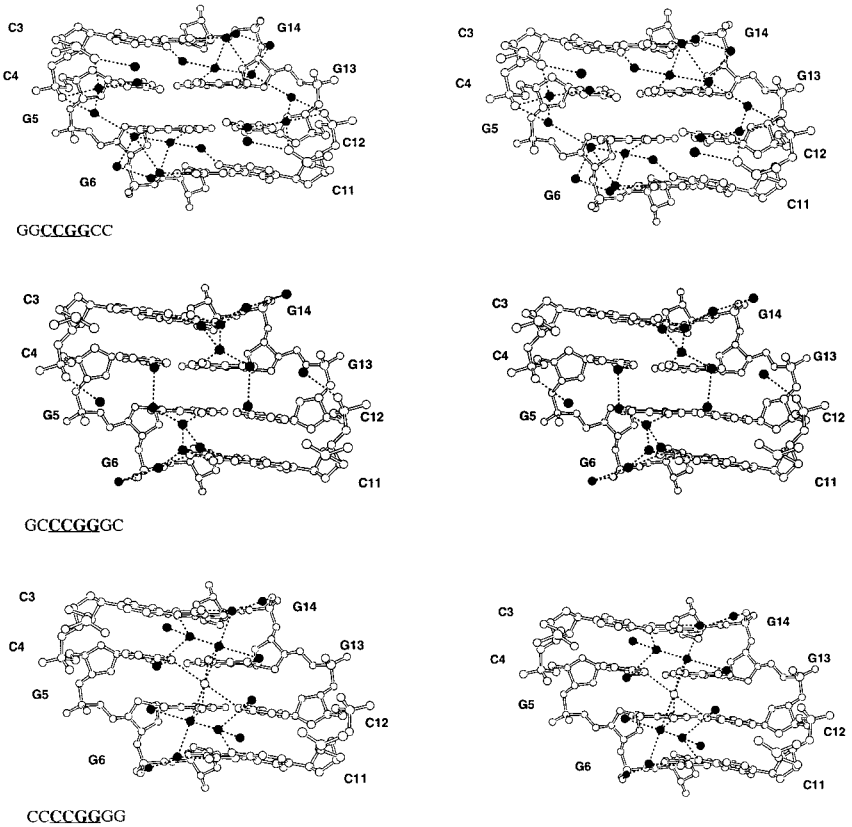

FIGURE 2: d(CCGG). Three sequences comprise this subgroup of tetragonal A-DNA octamers, d(GGCCGGCC), d(GCCCGGGC), and d(CCCCGGGG). The empty balls are solvent molecules reported to lie on crystallographic 2-fold in d(CCCCGGGG). They possess similar helix parameters, similar degrees of hydration, and similar patterns of hydration in their major grooves (Table 2). d(CCGG) is the only subgroup with negative rolls for the central base step, partially accounting for the well-hydrated base pairs and the large number of major groove waters. The CG-rich nature of these sequences also encourages solvent interactions. d(CTAG) vs d(CCGG). d(CCCTAGGG) and d(GGCCGGCC) display remarkably similar patterns of hydration in the stretch of major groove defined by their central four base pairs (Figures 1 and 2). They share 16 common solvent positions out of a possible 16. CG base pairs tend to be better hydrated than TA base pairs. d(CCCTAGGG), however, has the highest central twist angle and the largest negative flanking roll in this family, allowing the central four bases to form an extensive pattern of hydration very similar to that found in the major groove of d(GGCCGGCC).

Ramakrishnan and Sundaralingam showed that the helical parameters for the tetragonal A-DNA family of octamers are sequence independent and can by explained by invoking similarities in the lattice and/or crystal packing forces.

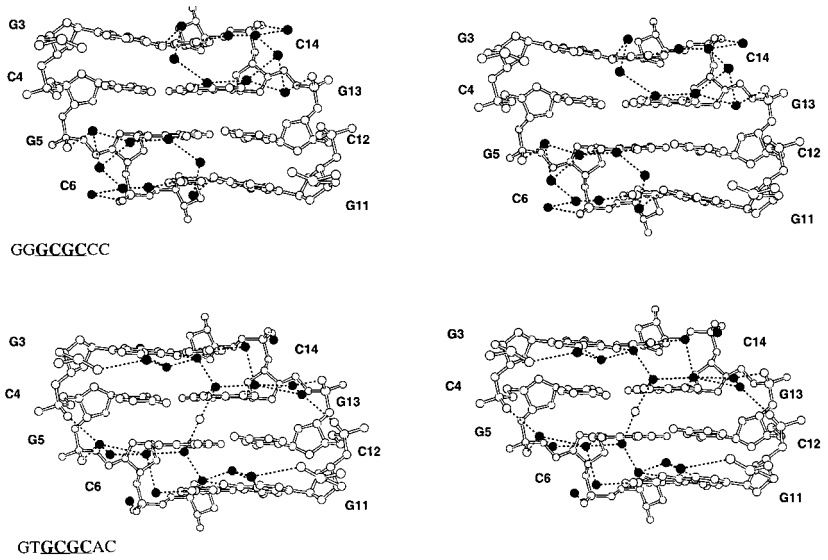

FIGURE 3: d(GCGC). d(GGGCGCCC) and d(GTGCGCAC) comprise this subgroup of isomorphous octamers. The empty balls are solvent molecules reported to lie on crystallographic 2 fold in d(GTGCGCAC). Among themselves they share 12 common solvent positions out of a possible 16 and exhibit similar degrees of hydration. The greatest difference in hydration occurs for the flanking base pairs, where d(GTGCGCAC) binds two more solvent molecules than d(GGGCGCCC) and supports an extensive hydrogen-bonding network which spans the entire width of the major groove. d(GTGCGCAC) has a flanking roll of −1.99° and is fully hydrated (3 H_2O per base pair) while d(GGGCGCCC) has a flanking roll of 0.18° and is, consequently, less hydrated (2 H_2O per base pair). Notice that d(GTGCGCAC) has an exceptionally low solvent-accessible area for its flanking base pairs (33 Å^2) but is still fully hydrated (Table 2). Also, the central base pairs exhibit solvent-accessible areas comparable to those in the d(CCGG) subgroup but possess positive rolls and are, consequently, not as well hydrated (2.0 H_2O per base pair versus 3 H_2O per base pair). d(CTAG) vs d(GCGC). Despite different types and arrangements of purines and pyrimidines, the solvent positions are amazingly similar between d(CCCTAGGG) and this subgroup, differing mainly in the direction and interaction of the hydrogen bonds. d(CCCTAGGG) and d(GTGCGCAC) both have hydrogen-bonding networks which span the width of their major grooves. In d(CCCTAGGG), the central d(TA) base pair has a higher twist angle and a less positive roll angle than does d(GTGCGCAC), allowing the central d(TA) base pair to acquire a similar degree and pattern of hydration (Table 2). d(CCGG) vs d(GCGC). d(GCCCGGGC) and d(GGGCGCCC) have 14 solvent positions in common (Figures 2 and 3). In d(GCGC) sequences, the central cytosines form no hydrogen contacts with solvent molecules, which can be explained by their positive base rolls. d(CCGG) sequences exhibit negative base rolls for their central and flanking base pairs and are, consequently, better hydrated. Note that d(GCCCGGGC) and d(GGGCGCCC) have identical water bridges between their flanking base pairs and adjacent phosphate groups, despite differences in major groove hydrophilic positions, G vs C.

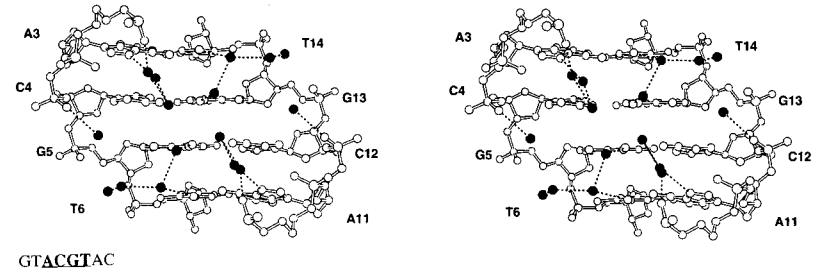

FIGURE 4: d(ACGT). Hydrationwise d(GTACGTAC) is quite unique and displays some extreme helical parameters. The large positive roll (4.28°) and low twist angle (23.0°) of the central step can account for the lower hydration state of the central base pair, where only two of the six possible sites, the O6 atoms of G(5) and G(13), make hydrogen bonds to solvent molecules. d(GTACGTAC) also exhibits the lowest solvent-accessible area for its flanking base pairs, 25.2 Å^2, which is only 40% of the solvent-accessible area in the flanking regions of other sequences.

Qualitatively, the trends in the helix parameters are similar, but quantitatively, there are variations in the magnitudes of these trends. The current study shows that similar A-DNA sequences can have different patterns of major groove hydration if they have different magnitudes for the twist and/or roll parameters [e.g., d(CTCTAGAG) and d(CCCTAGGG)]. On the other hand, different sequences can have similar patterns in major groove hydration provided the twist angles and base rolls are properly adjusted; i.e., less hydrated TA base pairs can achieve the same extent and pattern of hydration as CG base pairs if they possess higher twist angles and/or more negative base rolls [e.g., d(CCCTAGGG) is similarly hydrated to d(GCCCGGGC)]. Because Ramakrishnan and Sundaralingam found that the helix parameters are dependent on the crystal packing, and this study finds that major groove hydration is dependent on helix parameters, it stands to reason that major groove hydration is indirectly dependent on crystal packing/environmental forces. These results are consistent with studies of B-DNA hydration. Schneider and Berman (1995) demonstrated that the hydration of B-DNA is local and dependent on base geometry by replacing the central AATT of the Dickerson dodecamer with GGCC. As long as there was not a significant change in helix parameters, the positions and densities of hydration sites were similar between the two structures.

Recognition of DNA by proteins and drugs first involves interaction with and the subsequent displacement of primary sphere solvent molecules. A 1994 study found that 25%−100% of the enthalpy of ligand binding is due to the reestablishment of solvent hydrogen bonds after solvent

displacement from ligand binding sites (Chervenak & Toone, 1994). Therefore, solvent molecules must be viewed as active participants in the processes which interact with and, hence, regulate DNA. It has been argued that proteins and drugs do not merely recognize DNA but rather the solvent−DNA complex and that, after binding, reorganization of solvent molecules may be of primary importance in maintaining stability of the drug or protein−DNA complex. For example, the central base step in the tetragonal A-DNA family of octamers is underwound by about 10° with respect to the average helical twist. This is the same degree of underwinding seen in base steps which harbor intercalating agents. Our study finds that this degree of underwinding does not support major groove hydration to the extent that larger twist angles do, which may have some bearing on the binding of intercalating agents and other hydrophobic interactions. Intercalating agents preferably intercalate at py·pu base steps, the same steps which exhibit severe underwinding and dehydration in response to crystal packing forces. Rich *et al*. recently reported the molecular structure of iododoxorubicin complexed with d(TGTACA) and d(CGATCG) (Berger *et al*., 1995). Iododoxorubicin is an anticancer drug in phase II clinical trials and exhibits reduced basicity and improved lipophilicity compared to related anthracycline drugs. Its geometry is not significantly different from previous anthracycline complexes, but the local solvent environment about the py·pu base step is markedly affected. Obviously, the mechanism of intercalation requires lower twist angles and altered states of hydration. But does the ability of py·pu steps to assume lower twist angles alter local major groove hydration and allow the intercalation of anthracycline compounds, or does the intercalation of anthracyclines underwind the twist angle and affect major groove hydration? Similar arguments may be applied to crystal packing forces and protein−DNA interactions, which are known to affect DNA structure.

To fully appreciate the interactions between DNA and its environment, it is necessary to understand the way in which primary sphere solvent molecules respond to changes in the local helix geometry. The results of this analysis mark the first steps in developing predictive algorithms for the hydration of A-DNA based on knowledge of the base sequence and local helix parameters. The same approach used to analyze major groove hydration in this paper could be readily extended to other sequences and families of DNA oligomers.

REFERENCES

Ban, C., Ramakrishnan, B., & Sundaralingam, M. (1996) *Biophys. J. 71*, 1215−1221.

Berger, I., Su, L., Spitzner, J. R., Kang C., Burke, T. G., & Rich, A. (1995) *Nucleic Acids Res. 23*, 4488−4494.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hseih, S., Srinivasen, A. R., & Schneider, B. (1992) *Biophys. J. 63*, 751−759.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rogers, J. B., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.

Bingman, C., Li, X., Zon, G., & Sundaralingam, M. (1992) *Biochemistry 31*, 12803−12812.

Boehncke, K., Nonella M., Schulten, K., & Wang, A. H. (1991) *Biochemistry 30*, 5465−5475.

Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458−460.

Cervi, A. R., D'Estaintot, B. L., & Hunter, W. N. (1992) *Acta Crystallogr. B48*, 714−719.

Chalikian, T. V., Plum, G. E., Sarvazyan, A. P., & Breslauer, K. J. (1994) *Biochemistry 33*, 8629−8640.

Chervenak, M. C., & Toone, E. J. (1994) *J. Am. Chem. Soc. 116*, 10533−10539.

Chuprina, V. P., Heinemann, U., Nurislamov, A. A., Zielenkiewicz, P., & Dickerson, R. E. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 593−597.

Courseille, C., Dautant, A., Hospital, M., D'Estaintot, B. L., Precigoux, G., Molko, D., & Teoule, T. (1990) *Acta Crystallogr. A46*, Fc9−Fc12.

Eisenstein, M., & Shakked, A. (1995) *J. Mol. Biol. 248*, 662−678.

Eisenstein, M., Hope, H., Haran, T. E., Frolow, F., Shakked, Z., & Rabinovich, D. (1988) *Acta Crystallogr., Sect. B 44*, 625−628.

Gessner, R. V., Quigley, G. J., & Egli, M. (1994) *J. Mol. Biol. 236*, 1154−1168.

Haran, T. E., Shakked, Z., Wang, A. H.-J., & Rich, A. (1987) *J. Biomol. Struct. Dyn. 5*, 199−217.

Heinemann, U., Lauble, H., Frank, R., & Bloecker, H. (1987) *Nucleic Acids Res. 15*, 9531−9550.

Hunter, W. N., D'Estaintot, B. L., & Kennard, O. (1989) *Biochemistry 28*, 2444−2451.

Jain, S. C., Zon, G., & Sundaralingam, M. (1989) *Biochemistry 28*, 2360−2364.

Kennard, O., Cruse, W. B. T., Nachman, J., Prange, T., Shakked, Z., & Rabinovich, D. (1986) *J. Biomol. Struct. Dyn. 3*, 623−647.

Ramakrishnan, B., & Sundaralingam, M. (1993) *J. Biomol. Struct. Dyn. 11*, 11−26.

Rau, D. C., & Parsegian, V. A. (1992) *Biophys. J. 61*, 246−259.

Sack, J. S., & Spurlino, J. C. (1991) *Crystallographic Modeling Program, version 5.2.*, Baylor College of Medicine, Houston, TX.

Schneider, B., & Berman, H. (1995) *Biophys. J. 69*, 2661−2669.

Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F., Rabinovich D., Joachimiak, A., & Sigler, P. B. (1994) *Nature 368*, 469−473.

Tippin, D. B., & Sundaralingam, M. S. (1996) *Acta Crystallogr., Sect. D 52*, 997−1003.

Wahl, M. C., & Sundaralingam, M. S. (1996) in *Nucleic Acids Reviews* (Berman, H., & Breslauer, K., Eds.) John Wiley & Sons, Inc., New York (in press).

Wang, A. H.-J., Fugii, S., Van Boom, J. H., & Rich, A. (1982) *Proc. Natl. Acad. Sci U.S.A. 79*, 3968−3972.

Yanugi, K., Prive, G., & Dickerson, R. (1991) *J. Mol. Biol. 217*, 201−214.

BI9615194