Hydration of the DNA Bases Is Local

Bohdan Schneider* and Helen M. Berman‡
*J. Heyrovsky Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, 18223 Prague, Czech Republic, and ‡Rutgers University, Department of Chemistry, New Brunswick, New Jersey 08903 USA

ABSTRACT Ordered hydration sites were determined for the nucleotide bases in B-type conformations using the crystal structure data on 14 B-DNA decamer structures. A method of density representation was extended so that positions, occupancies, and distributions of the hydration sites were predicted around a B-DNA double helix by a method analogous to crystallographic refinement. The predicted hydration sites correctly reproduce the main features of hydration around the B-DNA dodecamer. In contrast to the previous observations, the newly available crystal data show the same extent of hydration of guanine and adenine, and of cytosine and thymine.

INTRODUCTION

It is well established that DNA hydration is directly related to the conformation and properties of DNA molecules (Wang, 1955; Falk et al., 1970; Milton and Galley, 1986; Umehara et al., 1990; Rentzeperis et al., 1993; Schreiner et al., 1991; Chen and Prohofsky, 1993; Chalikian et al., 1994; Langan et al., 1992). In addition, the hydration of DNA bears directly on its recognition properties. To describe the hydration patterns around the bases of DNA, a method of density representation of spatial distributions has been developed (Schneider et al., 1993) using the results of known crystal structures.

Other methods have also been developed to visualize preferred directions of hydrogen bonding to important hydrophilic groups. Rosenfield et al. (1984) and Murray-Rust and Glusker (1984) have studied preferred hydrogen bonding directions of -C=O, -OH, and -NH groups. Each atom (point) in the scatter plot of hydrogen-bonded groups around the central one has been given a Gaussian distribution; the density of points in the scatter plot has been calculated and contoured. A technique of contouring the densities of points (water oxygens) has also been used for modeling solvent sites around hydrophilic groups of proteins. Pitt et al. (1993) have determined likely hydration sites by calculating "3-D probability maps," and the algorithm by Roe and Teeter (1993) is based on seeded cluster analysis of the scattergrams.

A previous study (Schneider et al., 1993) showed different hydration patterns around the bases in A-, B-, and Z-DNA conformations. The observed hydration around all bases of a particular type was then used to create a "hydrated building block." The B-DNA building blocks were used to successfully predict the "spine of hydration" around a B-DNA dodecamer using structural information derived from known crystal structures. This information consisted of the observed hydration around all bases of a particular type that was then integrated into a hydrated building block. A drawback of the procedure was that the building blocks used to make the predicted models were themselves derived from a sample set containing many dodecamer structures. Because the crystal structures of B-DNA dodecamers are isomorphous and their sequences show variability only in the central regions, some features of the hydration sites could have been a result of the crystalline arrangement or of the similarity in the sequences.

At the time of the previous study, only a small number of non-dodecamer B-DNA structures were available. There are now more than a dozen decamer structures that crystallize in several different space groups and have more variable sequences (Table 1). It was therefore decided to create new building blocks with the decamer structures and to see how well these building blocks could be used to predict hydration patterns in the dodecamer structures. The better quality of the new building blocks allowed us to reevaluate the extent of hydration around each of the bases. In addition, the method was substantially extended so that it is possible to calculate not only positions but also occupancies and distributions of the hydration sites.

METHODS

Construction of the hydrated building blocks

The Nucleic Acid Database (Berman et al., 1992) was used to search for B-DNA decamer structures. The structures used for this analysis are shown in Table 1. For each base in the sample of structures, all water molecules within 3.40 Å from any base atom were determined with the program BANG (Carrell, 1979). Most of these water molecules were between 2.6 and 2.8 Å from base atoms, indicating strong hydrogen bonds. To include all potentially hydrogen-bonded water molecules in the analysis the distance for retrieving water molecules was increased from 3.20 Å (Schneider et al., 1992, 1993) to 3.40 Å in the present analysis.

The bases at the ends of nucleotide strands were excluded from the analysis to eliminate potential bias caused by interactions sterically possible only at the strand ends. By eliminating the end bases, the numbers of bases of each type became similar (Table 2), making the statistical samples more balanced. Modified bases as well as bases in mismatched pairs were excluded from the analysis.

Received for publication 16 June 1995 and in final form 8 September 1995.
Address reprint requests to Dr. Helen M. Berman, Department of Chemistry, Rutgers University, P.O. Box 939, Piscataway, NJ 08855-0939. Tel.: 908-932-4667; Fax: 908-445-5958; E-mail: berman@dnarna.rutgers.edu.

0006-3495/95/12/2661/09 $2.00

TABLE 1 Structures used in this study

NDB code*	Sequence‡	Space group	Base water§	Reference
BDJ008	CCAA***GA***TTGG	C2	2.10	Privé et al., 1987
BDJ017	CCAGGCCTGG	C2	1.60	Heinemann and Alings, 1989
BDJ019	CCAACGTTGG	C2	3.00	Privé et al., 1991
BDJ025	CGATCGATCG	$P2_12_12_1$	3.05	Grzeskowiak et al., 1991
BDJ031	CGATTAATCG	$P2_12_12_1$	2.45	Quintana et al., 1992
BDJ036	CGATATATCG/Ca	$P2_12_12_1$	2.35	Yuan et al., 1992
BDJ037	CGATATATCG/Mg	$P2_12_12_1$	1.15	Yuan et al., 1992
BDJ039	CCGGCGCCGG	R3	1.85	Heinemann et al., 1992
BDJ043	CCAACITTGG	$P3_221$	1.00	Lipanov et al., 1993
BDJ044	CCAACITTGG	C2	2.00	Lipanov et al., 1993
—	CCAA***IA***TTGG	C2	2.20	Dickerson,¶ unpublished, 1993
BDJ051	CATGGCCATG	$P2_12_12_1$	1.60	Goodsell et al., 1993
BDJB27	CCAGGC(^{Me}C)TGG	P6	1.65	Heinemann and Hahn, 1992
BDJB48	CGATCG(^{Me}A)TCG	$P3_221$	1.80	Baikalov et al., 1993

*Structure identification used in NDB.
‡Sequence of a oligonucleotide. Base modification is indicated by a superscript; bases involved in a mismatch are in ***bold italics***.
§Number of crystallographically ordered water molecules within 3.40 Å of any base atom divided by the number of nucleotides in the asymmetric unit.
¶Dickerson, R. E. 1993. Unpublished structure of a B-DNA decamer CCAAIATTGG.

The statistical *zI* test was used to quantitatively compare numbers of water molecules attached to the bases. The *zI* test can be used for comparing two random samples of different but comparable sizes (for details see Langley, 1971). For example, a contingency table was built to test the null hypothesis that the number of water molecules attached to the cytosine major groove is not different from that of thymine (Table 3). The *z* value was calculated and compared with the tabulated values at a desired significance level. The *z* values have the same probability distribution as the probabilities of the standard *t*-test. In the example shown, the distributions are not significantly different at the 10% confidence level.

Bases of a particular type with the attached water molecules were superimposed to construct hydrated building blocks. In purines the atoms N9, N7, N3, and in pyrimidines the atoms N1, N3, C5 were overlapped using the program BMF (Sussman, 1984). Coordinates of the base templates over which all bases were superimposed were taken from the *ab initio* study by Sponer and Hobza (1994). These newly created hydrated building blocks that contain tens to hundreds of partially occupied water molecules were used to calculate water densities around the bases in the following manner: i) The positions of individual water molecules were assigned occupancies of $1/N$, where N is the number of bases used for construction of the hydrated building block; ii) the program X-Plor (Brünger, 1992) was used to convert these positions into structure factors $F_{MOD}(hkl)$ and subsequently into iii) electron densities of water molecules; iv) the densities were displayed and peaks manually fitted using the program Frodo (Jones, 1978). The centers of peaks are called hydration sites. A detailed description of steps i) to iv) is in Schneider et al. (1993). v) The positions of the hydration sites were refined. The pictorial summary of the process is shown in Fig. 1.

The hydration sites were determined and refined for the four DNA bases and the Dickerson-Drew dodecamer (Drew et al., 1981). To determine hydration sites around the dodecamer, the current building blocks were incorporated into the conformation of DNA found in the crystal by using the previously developed program. The water electron density was calculated in the same way as for isolated bases. The fully hydrated dodecamer structure contains 2228 partially occupied water positions, which result in 55 hydration sites.

Refinement of the hydration sites

The densities calculated from $F_{MOD}(hkl)$ are not true crystallographic electron densities but are the Fourier transform of the hydrated building blocks. Nevertheless, the $F_{MOD}(hkl)$ and the derived densities can be treated the same as F_{obs} and electron densities. Thus, the positions of the hydration sites can be derived by manually fitting the densities and then subjected to crystallographic refinement. As a result of such a refinement, new positions can be calculated. The temperature displacement factors are a measure of the distribution and the occupancies are a measure of the likelihood that a water molecule is at that site.

The electron density is not the same at all hydration sites, so that the sites do not have the same occupancies. Initial occupancies were estimated from the electron densities ρ_i of the sites as observed in FRODO. Occupancies of all of the water molecules in the hydrated building block were summed to get the expected stoichiometric number of water molecules in the block N_w: N_w = (number of water molecules in the hydrated building block)/(number of bases in the hydrated building block). Occupancies Q_i of the individual hydration sites were then calculated: $Q_i = N_w\ (\rho_i/\Sigma_k\ \rho_k)$, where $\Sigma_k\ \rho_k$ is the sum of water densities of all the hydration sites in a studied base.

The program SHELXL-93 (Sheldrick, 1993) was used for the refinement. Required values of estimated standard deviations (esd) of structure factors $F_{MOD}(hkl)$, which substitute for observed structure factors F_{obs} in the refinement, were set at a value of 0.01 for every reflection. It represents about 1% of an average $F_{MOD}(hkl)$.

Bases were refined in a cell of dimensions $a = b = 15$, $c = 10$ Å, and $\alpha = \beta = \gamma = 90°$, the space group P1, and crystallographic resolution was set to 1.5 Å (1388 reflections). The final refinement parameters and their definitions are shown in Table 4. The weighting scheme used in the refinement was extensively tested to obtain flat distributions of both the $F_{MOD}(hkl)$ and the residual factor R as a function of resolution, and to keep the value of the goodness of fit close to 1. Even though these distributions depend on the weighting scheme and on the values of weighting parameters a and b, the positions, temperature displacement factors, and R factor do not show this dependence.

The refinement started from the manually fitted positions of hydration sites and their estimated occupancies. The isotropic refinement converged after 8 to 12 cycles. Attempts to refine occupancies resulted in such small changes that these values were kept at their original values. The anisotropic refinement converged rapidly and the R factor was reduced by 5%. The anisotropy of the parameters is shown in Fig. 1 *c* (Johnson, 1976). In a normal crystallographic refinement, these ellipsoids represent the thermal distribution. In this case they represent the spatial distribution of the water molecules illustrated in Fig. 1 *a*.

The 55 hydration sites of the dodecamer (Drew et al., 1981) were refined in a unit cell with dimensions of $a = 24$, $b = 32$, $c = 48$ Å, and $\alpha = \beta = \gamma = 90°$, in space group P1, and with crystallographic resolution of 1.5 Å (22,869 reflections). The refinement parameters are in Table 4. After an extensive testing, the same weighting scheme used for the refinement of the isolated bases was used. The initial hydration site occupancies were estimated by a procedure different from that of the case of the bases. The highest electron density observed in the map (about 24 σ) was scaled to an occupancy of 1.00; all other peaks were proportionally decreased down to the lowest fitted peak of 10 σ with an occupancy of 0.40. The stoichiometric number of water molecules around the dodecamer was then 30.2. The isotropic temperature displacement factors and the occupancy factors were alternatively refined in eight steps, each step taking about eight least-square cycles to converge. The stoichiometric number of water oxygens dropped from 30.20 to 29.10 during this process. The anisotropic refinement proceeded in three steps, each consisting of 12 least-square

TABLE 2 Statistics of hydrated building blocks

Bases*	Guanine (42)			Cytosine (43)			Adenine (43)			Thymine (45)		
Groove‡	m	M	All	m	M	All	m	M	All	m	M	All
Waters¶	40	61	101	32	52	84	35	57	92	35	60	95

*The reported numbers of bases used to construct the hydrated building blocks. Twenty-two guanines attached to 52 water molecules and 23 cytosines attached to 43 water molecules at the ends of nucleotide strands were excluded from the analysis and are not reported here. No modified (including inosines) or mismatched bases were analyzed.

‡M, major groove; m, minor groove.

¶The number of water molecules within 3.4 Å of the atoms in the major groove, the minor groove, and both grooves for the designated base.

cycles. The first and last steps refined anisotropic temperature displacement factors, and the middle step refined occupancies of the hydration sites. The final number of water oxygens was 28.9 and the residual R factor decreased by 5% from its converged isotropic value to the final value of 25%.

The previously determined hydration sites were also refined using the same cell parameters, resolution, and weighting schemes. The final positions of the hydration sites were compared to the positions of the current sites by calculating their distances, angles, and torsion angles to selected base atoms, and more quantitatively, by calculating their root mean square deviations after the superposition by the program BMF (Table 5).

RESULTS

Hydration of bases

Fig. 2 shows the electron densities of the hydrated building block and the manually fitted positions. The hydrogen bonding of the refined hydration sites is given in Table 5. These hydration sites are generally very close to their positions before the refinement, with the largest difference of 0.33 Å occurring between the W1 sites in adenine. The increase of the limiting distance for selecting water molecules as potentially hydrogen bonded from 3.20 to 3.40 Å did not change the length of hydrogen bonds between hydration sites and base atoms. The only visible effect is the emergence of two distant and weak hydration sites attached to the guanine atom N2.

The occupancies of the currently reported hydration sites are systematically higher than for the previously determined sites (Table 6) because the current hydrated building blocks

TABLE 3 Example of the *zI* test to determine if the numbers of water molecules attached to the cytosine and thymine in the major groove are the same

	Frequency* (F) Number of bases	Observation‡ (O) Number of waters	Expectation§ (E) Number of waters	Probability¶ (P) Fraction of x or y in sample
Cytosine (x)	43	52	54.7	0.4886
Thymine (y)	45	60	57.3	0.5114

*Number of bases in the sample. Observations with smaller frequency (F) have index x.

‡Number of water molecules (O) closer than 3.40 Å to the major groove atoms of cytosine (O_x) or thymine (O_y).

§Expected number of water molecules E_x and E_y. $E_x = F_x\,(O_x + O_y)/(F_x + F_y)$, analogously for E_y. F_x (F_y) is frequency of observation for cytosine (thymine).

¶Probability P_y is calculated as $P_y = F_y/(F_x + F_y)$. In this case this represents the fraction of x or y in the sample. The value of z is calculated as $z = (|E_x - O_x| - c)/(P_y E_x)^{1/2}$, where $c = 0.2$ if $E_x > O_x$ and $c = 0.5$ when $E_x < O_x$. In this case, z equals 0.4727. The limiting value for z is 1.64 at the 10% level and 1.96 at the 5% level. Therefore, it can be concluded that the number of waters around x (cytosine) and y (thymine) are the same.

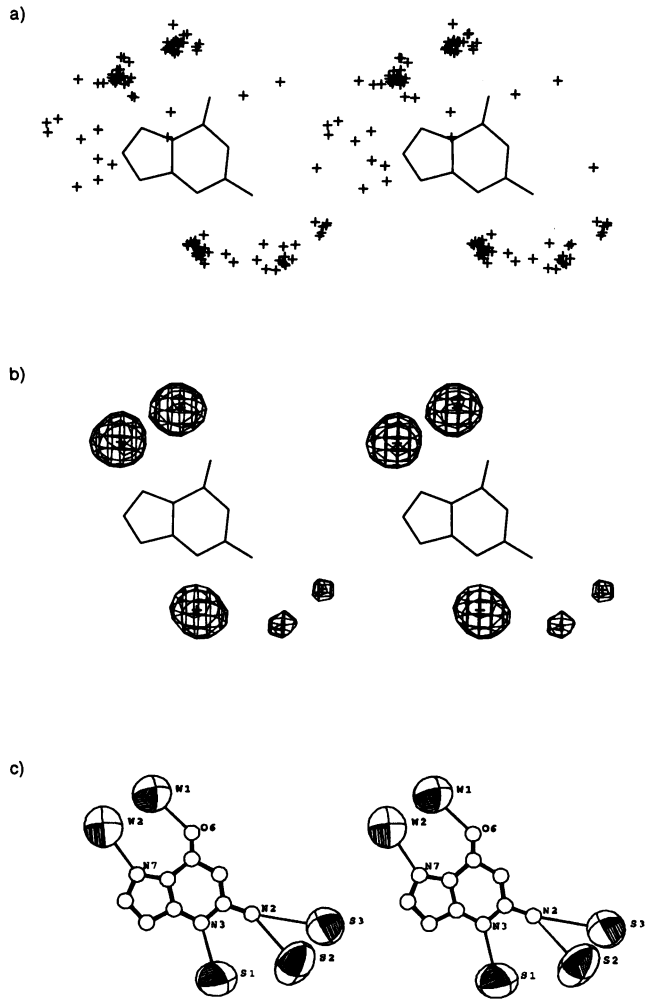

FIGURE 1 Three representations of water distributions around guanine in B-DNA. (*a*) Scattergram of 101 water molecules within 3.40 Å from any atom of 42 guanines found in 14 B-DNA decamer structures. (*b*) Electron densities of the 101 water molecules plotted at the 4 σ level. Each water is modeled as an oxygen atom with an occupancy of 1/42. (*c*) An ORTEP (Johnson, 1976) plot of the current guanine B-DNA hydration sites after the refinement. Plotted are 50% probability thermal ellipsoids. The key guanine atoms and hydration sites are labeled. All plots are in stereo.

TABLE 4 Refinement parameters of the current and previously determined hydration sites

	R*	GooF‡	Weight§		$\Delta\rho_e$¶	
Current hydration sites						
Guanine	20.0	0.806	0.35	0.30	0.07	−0.08
Adenine	20.5	1.032	0.20	0.30	0.99	−0.99
Cytosine	20.9	0.780	0.40	0.30	0.07	−0.08
Thymine	22.6	0.876	0.30	0.40	0.11	−0.06
BDL001‖	24.8	1.219	0.30	1.50	0.14	−0.08
Previously determined hydration sites						
Guanine	19.1	0.705	0.20	0.50	0.10	−0.11
Adenine	16.1	0.676	0.20	0.30	0.04	−0.05
Cytosine	17.0	0.885	0.20	0.20	0.04	−0.05
Thymine	20.0	0.834	0.30	0.30	0.08	−0.07
BDL001‖	20.9	1.043	0.36	1.46	0.08	−0.10
Guanine A-DNA**	19.6	0.677	0.10	0.30	0.04	−0.03
Guanine Z-DNA**	19.9	0.756	0.40	0.30	0.05	−0.10
Cytosine A-DNA**	20.3	0.840	0.30	0.10	0.04	−0.02
Cytosine Z-DNA**	20.0	0.722	0.40	0.30	0.05	−0.08

*Final value of the residual R factor:

$$R = \sum_{(hkl)} (F_{MOD}(hkl) - F_{calc}(hkl))/F_{MOD}(hkl);$$

where $F_{MOD}(hkl)$ are the model structure factors replacing observed structure factors of diffraction experiments and $F_{calc}(hkl)$ are the calculated structure factors.

‡Goodness of fit:

$$S = \sum_{(hkl)} (w(F_{MOD}{}^2 - F_{calc}{}^2)^2/(n-p))^{1/2};$$

where w is the weight, n the number of reflections, and p the number of refined parameters.

§ Parameters a and b of the optimal weighting scheme found:

$$w = 1/(\sigma(F_{MOD})^2 + (aP)^2 + (bP)),$$

where $P = ((F_{MOD})^2 + 2(F_{calc})^2)/3$, and $\sigma(F_{MOD})$ is the estimated standard deviation of the model structure factors F_{MOD}.

¶Highest and lowest electron density that was not fitted by any hydration site in the refinement (i.e., residual electron density).

‖Refinement of the predicted hydration sites in the Dickerson-Drew dodecamer (Drew et al., 1981). The range of equivalent temperature factors Beq was 27–53 Å^2 for the current and 19–37 Å^2 for the previous hydration sites.

**The A-DNA and Z-DNA hydration sites (see Table 3 in Schneider et al., 1993) were also refined with these final values of occupancies and Beq: A-DNA guanine: S1 0.20/36; W1a 0.35/27; W1b 0.15/40; W2 0.40/36; cytosine; S1 0.25/37; W1 0.35/24; W2a 0.15/47; W2b 0.05/43; Z-DNA guanine: S1 0.90/25; W1a 0.50/31; W1b 0.40/49; W2a 0.40/47; W2b 0.25/35; cytosine: S1a 0.65/33; S1b 0.65/33; W1 0.90/26.

TABLE 5 Geometries of the refined hydration sites

Atom				Distance‡	Angle‡	Torsion‡	
A	B	C	HS*	(C-HS)	(B-C-HS)	(A-B-C-HS)	Fit§
Guanine							
N1	C2	N3	S1	2.79	113	160	0.04
N1	C2	N2	S2	3.11	141	145	¶
N1	C2	N2	S3	3.31	169	37	¶
C5	C6	O6	W1	2.58	135	−4	0.09
N9	C8	N7	W2	2.67	120	−163	0.14
		S2	S3	2.38			
		W1	W2	2.85			
Adenine							
N1	C2	N3	S1	2.76	110	174	0.12
C5	C6	N6	W1	2.99	136	−6	0.33
N9	C8	N7	W2	2.71	126	−171	0.06
		W1	W2	2.20			
Cytosine							
N1	C2	O2	S1	2.63	158	50	0.04
N3	C4	N4	W1	2.97	116	−178	0.01
C2	N1	C6	W2	3.02	141	−153	0.20
Thymine							
N1	C2	O2	S1	2.69	156	64	0.03
N3	C4	O4	W1	2.72	144	−175	0.06
C2	N1	C6	W2	3.09	127	−168	¶

*Hydration sites are labeled "S" in the minor groove and "W" in the major groove and are uniquely labeled.

‡Distances are in Ångstroms, angles in degrees. Distances are listed when shorter than 3.40 Å.

§Distances between the corresponding current and previously determined hydration sites after their least-square fit by the program BMF (in Å).

¶The site was not observed in Schneider *et al.* (1993).

contain more water molecules than the previously determined building blocks. The ratio of the number of stoichiometric water molecules to the number of bases in the sample is 2.14 for guanine, 2.14 for adenine, 1.95 for cytosine, and 2.11 for thymine. In the previous analysis these values were 1.0, 1.14, .64, and 1.05 for guanine, adenine, cytosine, and thymine, respectively. These differences are due to the fact that decamer structures have generally higher crystallographic resolution and consequently larger numbers of ordered water molecules than the dodecamer structures.

This quantitative difference between the previous and current hydration sites is accompanied by another qualitative shift in their properties. In the previous study, the adenine and thymine appeared to be more extensively hydrated than guanine and cytosine, especially in the minor groove. Even when these observations agreed with some experiments on DNA hydration (Tunis and Hearst, 1968; Mrevlishvili, 1981), the better balanced decamer hydrated building blocks show that CG is hydrated to an extent similar to that of TA.

As explained in the Methods section and in Table 3, the statistical zI test was used to decide whether the distributions of water molecules are significantly different between guanine and adenine, between cytosine and thymine, and between the CG and TA pairs. Each test was made separately for the minor groove, major groove, and for water molecules around the whole base (numbers of potentially hydrogen-bonded water molecules are listed in Table 2). All nine tests confirmed the null hypothesis that the compared distributions are the same at the 5% confidence level. Therefore, there are no statistically significant differences in the extent of hydration of the individual bases.

The extent of hydration was also compared between the C and G bases at the ends of the nucleotide strands versus inside the strands (Table 2), and no significant differences were found between these two distributions.

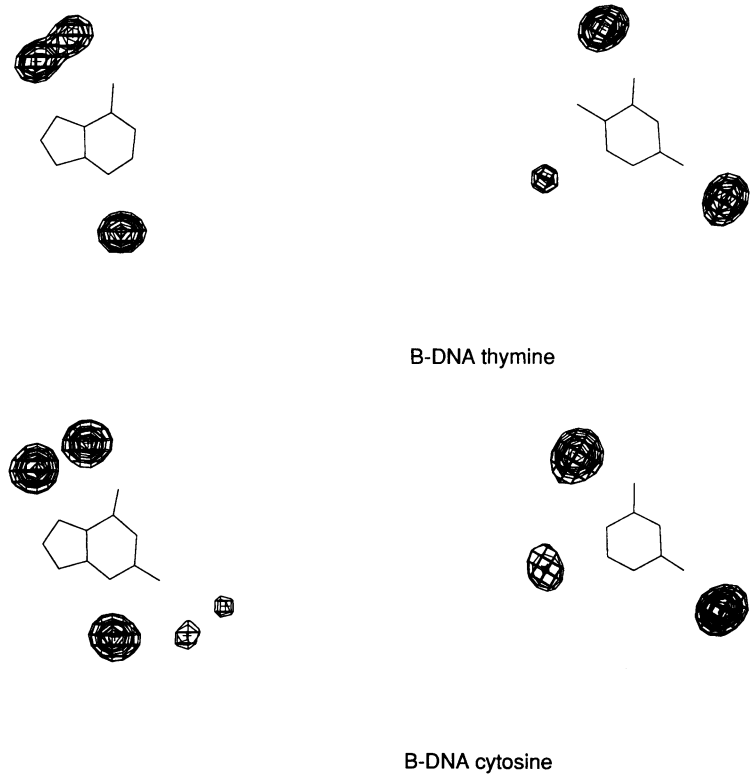

FIGURE 2 Pseudo-electron densities for the hydrated building blocks of adenine, guanine, thymine, and cytosine.

Guanine and adenine

The character of hydration of the two purines is very similar. The main hydration sites lie close to the base plane in good hydrogen bonding distances from their anchoring atoms. A significant difference between hydrations of the two bases was found in the distances between the W2 hydration sites to O6/N6. A larger distance in adenine probably reflects a different nature of HOH···O= and H_2O···HN— hydrogen bonds. Occupancies of all major hydration sites in guanine are very similar, whereas the adenine minor groove is hydrated more extensively than its major one.

Hydration of the guanine N2 atom has been a subject of controversy for some time; the present data confirm the previous observation (Schneider et al., 1992, 1993) that there is not a localized hydration site around this atom. The two distant and weak hydration sites observed around guanine N2 in the current study (S2, S3, Table 5) originate from hydration sites of neighboring pyrimidine O2 atoms (Fig. 3).

TABLE 6 Occupancies and temperature displacement factors of the current and previously determined hydration sites

		Current hydration sites‡		Previous hydration sites§	
	HS*	occ¶	Beq‖	occ¶	Beq‖
Guanine	S1	0.65	29	0.30	20
	S2	0.20	32	—	—
	S3	0.15	23	—	—
	W1	0.65	30	0.40	16
	W2	0.75	29	0.30	17
Adenine	S1	1.00	33	0.65	23
	W1	0.50	28	0.25	29
	W2	0.55	28	0.30	26
Cytosine	S1	0.75	36	0.15	20
	W1	0.90	29	0.45	14
	W2	0.35	36	0.05	15
Thymine	S1	0.80	36	0.70	21
	W1	0.80	30	0.40	22
	W2	0.35	36	—	—

*Names of hydration sites. See also Table 5.
‡Refinement of the current hydration sites.
§Refinement of the previously determined hydration sites (Schneider et al., 1993).
¶Occupancies of hydration sites; full occupancy = 1.00.
‖Equivalent values of anisotropic temperature displacement factors, in Å^2.

Cytosine and thymine

The hydration sites of pyrimidines are very similar to each other. Both minor groove hydration sites are out of the base plane, although more so for thymine. Hydration of the major groove is influenced by the thymine methyl group, which moves the W1 hydration site closer to the paired adenine (Table 5, the C4-N4/O4-W1 angles for thymine and cytosine are 116° and 144°). Out-of-plane, low-density hydration sites W2 probably represent a connection to hydration of the phosphate group. In both pyrimidines, minor and major groove peaks have similar occupancies (disregarding weak W2 sites).

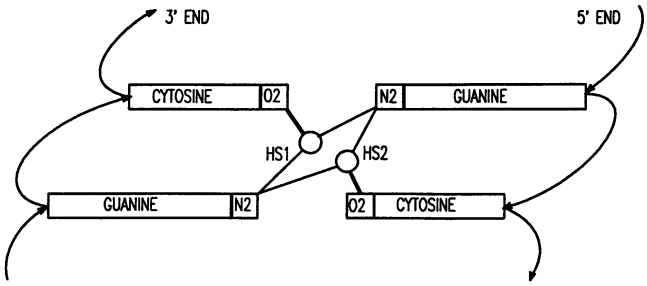

FIGURE 3 Explanation of the existence of two distant hydration sites at the guanine N2 atom. Some water molecules hydrating the pyrimidine atom O2 are also close to the guanine N2 atom. The effect is most apparent for the 5′-GC-3′ sequence. A view into the B-DNA minor groove shows that both guanine sites are in the 3′ direction from guanine, whereas the cytosine site is in the 5′ direction. HS1 and HS2 sites of both cytosines contribute to weak hydration of both guanines.

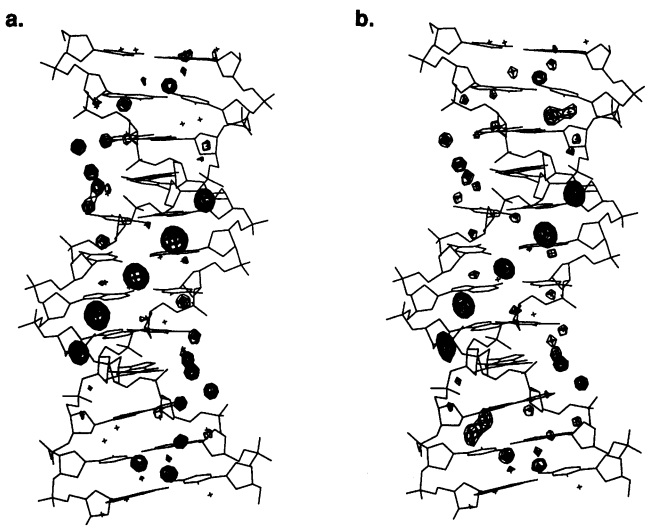

FIGURE 4 Comparison of the predicted hydration of the Dickerson-Drew dodecamer using (*a*) building blocks from all B-DNA structures and (*b*) decamer building blocks.

Refinement of base hydration sites

Important refinement parameters and comparison between the current and previous hydration sites are in Tables 4 and 6.

The hydration sites of guanine and cytosine in A- and Z-DNA conformations were also refined. The results are fully comparable to the refinement characteristics of B-DNA hydration sites (Table 4). The positions of the hydration sites are very similar before and after the refinement.

Hydration of oligonucleotides

The hydration of the Dickerson-Drew dodecamer 5′-d(CGCGAATTCGCG)-3′ (Drew et al., 1981) was predicted using the hydrated building blocks created from the 14 B-DNA decamer structures. A comparison of the dodecamer structure built with the decamer building blocks and with building blocks from all B-DNA structures is shown in Fig. 4. The hydration sites observed for the isolated bases are mostly reproduced at each dodecamer step. Every base has one hydration site in the minor groove. Pyrimidines have one site and purines have two sites in the major groove. The distances between refined hydration sites and the anchoring base atoms are within their 3 esds from these distances in the isolated bases. A significant exception is observed in the minor groove of the central AATT region, where water densities of neighboring bases fuse together and create five hydration sites. As is shown in Fig. 5, the sites are formed by waters of C21 and A5, A6 and T20, T7 and T19, T8 and A18, and A17 and C9. These five sites are the most populated, with occupation factors close to 1. The last four sites correspond to waters of the spine of hydration observed experimentally (Drew and Dickerson, 1981), but the site formed between C21 and A5 has no experimental counterpart. In the crystal environment, the N2 atom of residue G4 is involved in a packing interaction with the O3′ atom of G12. Even when the distance between the two atoms is rather long (3.75 Å) and N2 G4 does not directly interact with the predicted hydration site (see Fig. 5), the presence of the symmetry-related DNA molecule may block binding of a water molecule to this potential hydration site.

It is interesting to note that half of the predicted minor groove hydration sites are anchored to the backbone O4′ atoms. This agrees with compilation of water bridges in B-DNA structures in which a N3/O_2-water-O4′ bridge was found to be the most frequent bridge in the minor groove (unpublished results).

In the dodecamer crystal structure, 19 of 81 crystallographic water molecules have distances shorter than 3.40 Å from any base atom. Of these 19, only two do not have counterparts among the predicted hydration sites; one is attached to the T19 methyl group, and the other is attached to N4 of cytosine 13 at the end of the nucleotide strand. The root mean square deviation between the 17 crystal waters and corresponding hydration sites is 0.57 Å.

To further test the differences in possible hydration of the A-T versus G-C pairs, hydration sites were also determined for a model dodecamer structure with the same conformation as observed in the crystal of $d(CGCGAATTCGCG)_2$ (Drew et al., 1981) but with the central AATT part replaced by the GGCC sequence. The guanine hydrated building blocks were inserted instead of adenine blocks, cytosine blocks were inserted instead of thymine blocks, and water densities were calculated around the modified dodecamer $d(CGCGGGCCCGCG)_2$. The positions and densities of hydration sites in the central part of the dodecamer are almost identical whether calculated for the AATT or GGCC sequence. Distances between the corresponding sites in the minor groove vary between 0.27 and 0.42 Å after superposition. Their densities are very similar on the absolute scale, except for the density of the central site between T7 and T18 (20 σ) and C7 and C18 (28 σ).

Both sets of hydrated building blocks predict 55 hydration sites around the dodecamer. As in the case of the

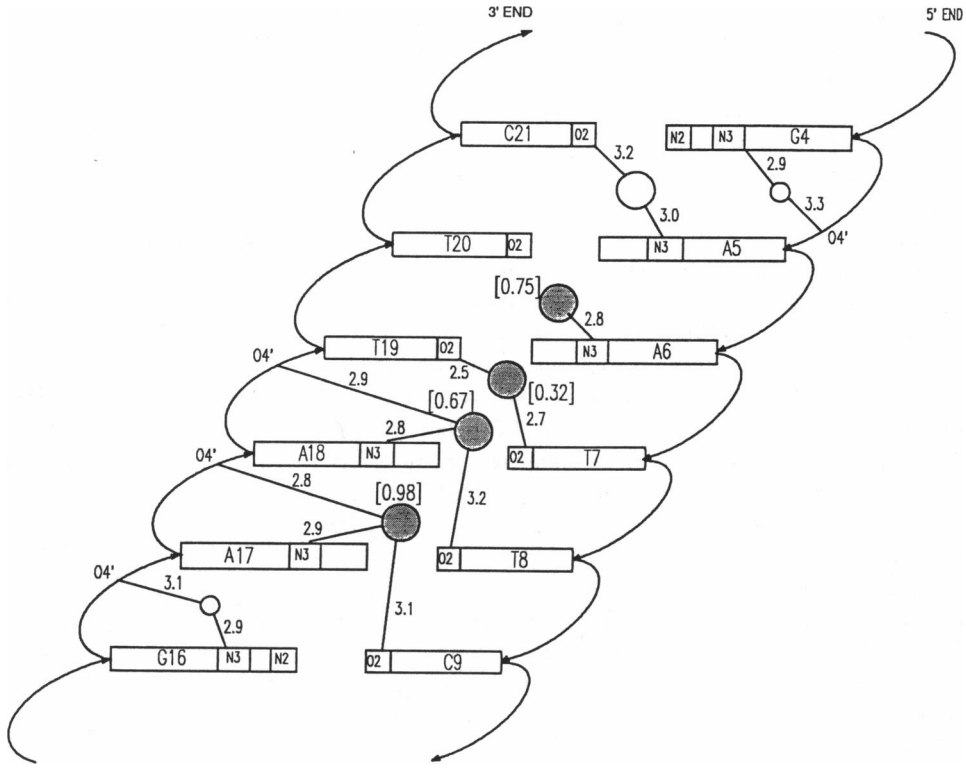

FIGURE 5 Connectivity of the hydration sites in the central AATT region of the Dickerson-Drew dodecamer as predicted by the current building blocks. Hydration sites with full occupancy are shown as large spheres, small spheres are hydration sites with occupancies around 0.5. The four shaded sites match four experimentally determined water molecules (Drew et al., 1981), which form the spine of hydration. The distances between the predicted and experimental water positions are given in square brackets. All contacts of the hydration sites shorter than 3.40 Å are labeled. Note tight contacts to backbone's atoms O4′.

isolated bases, the current hydration sites are very close to the previous ones, with a root mean square deviation of 0.3 Å, and have systematically higher occupancies. In addition, they match the 17 crystal waters with a root mean square deviation of 0.6 Å. Higher temperature factors of hydration sites predicted with the current building blocks (see Table 4, note ‖) are a corollary of the fact that the more diverse decamer structures were used to predict hydration of the dodecamer, whereas in the case of the previous building blocks the dodecamer hydration was predicted using the dodecamer structures.

During the refinement of the current hydration sites around the dodecamer, the stoichiometric number of water molecules decreased from the initial value of 30.2 to 29.0 at the end of the refinement. Most individual occupancies were between 0.4 and 0.6. Smaller values were obtained for the major groove hydration sites. The hydration sites representing waters in the spine of hydration had occupancies between 0.90 and 1.0 in the current and the previous models.

DISCUSSION AND CONCLUSIONS

A significant extension of the method of density representation (Schneider et al., 1993) is presented here. The method allows a prediction of positions as well as occupancies and distributions of the hydration sites. Although the refinement of the hydration sites is not the same as the refinement of atomic models against experimentally derived data, the final refinement characteristics are very similar to those of biological molecules. Because we have established that the hydration of the DNA bases in B-type conformations appears to depend on the chemical nature of the base, the hydration around longer sequences can be predicted using these building blocks. Thus, these hydration sites can be used as reliable starting positions in theoretical calculations and computer experiments.

The current predictions are based solely on the B-DNA decamer structures. The positions of the current and previously determined hydration sites are very similar, despite the fact that the sample of mostly dodecamer structures used to determine the previous hydration sites was biased by a limited variability of nucleotide sequences and by the low number of packing arrangements available. It indicates that the derivation of the building blocks is not sensitive to sequence context or packing. The limiting distance for selection of attached water molecules was different for construction of the current (3.40 Å) and previous (3.20 Å) building blocks, so that both distances serve well to discriminate between hydrogen bonding and van der Waals (second shell) water molecules.

The analysis of hydration of the Dickerson-Drew dodecamer (Drew et al., 1981) showed that hydration of an oligomeric B-DNA can be predicted. The geometry of hydration around many bases in the helix is very similar to that around the isolated bases, but the local deformations of the DNA geometry may cause redistribution of water densities so that the hydration sites are in different positions than in the isolated bases. This is the case in the central AATT region of the dodecamer where the spine of hydration is formed.

It was demonstrated here that the spine of hydration can be formed by both AATT and GGCC sequences and therefore does not reflect unique features of hydration of adenine and thymine but rather the ability of the AATT sequence to form a narrow minor groove (Privé et al., 1991). The question remains as to whether the GGCC sequence is able to adopt such a conformation so that the spine of hydration can form.

This study demonstrates that in crystals there is similar hydration for CG and TA pairs. This is in contrast to studies by Tunis and Hearst (1968) and Mrevlishvili (1981), who have detected more extensive hydration around TA, and the measurements by Chalikian et al. (1994), who have observed more extensive hydration around CG. The results of different methods are difficult to compare quantitatively. Differences between the observables for poly(dC-dG) and poly(dT-dA) probably reflect not only differences in hydration per se but also differences in conformation of these polymers. Chemical intuition would lead one to suggest that CG could be hydrated more than TA, because there is one more hydrophilic atom (guanine N2) in the former and one more hydrophobic group in the latter pair (thymine methyl). However, in crystals we show conclusively that the guanine N2 does not have an ordered hydration site. Furthermore, statistical analyses of the sample set show that guanine and adenine in the minor groove and cytosine and thymine in the major groove are hydrated to the same extent.

In a stimulating paper by Chalikian et al. (1994), the authors state that the expected number of waters of hydration is between 19 and 24 per nucleotide pair. They stress that the actual number is method-dependent, but their densimetric and ultrasonic measurement generally agrees with the older estimates, which are between 12 and 20 water molecules per nucleotide pair. The calculations presented here permit us to extend their conclusions about the number of bound water molecules. In principle, these bound water molecules may or may not be ordered. Our calculations allow an estimation of the number of ordered water molecules in the first hydration shell of bases. Three purine and two pyrimidine fully occupied hydration sites give the upper limit of five ordered water molecules attached to the bases in each nucleotide pair. The upper limit of the number of water molecules hydrogen-bonded to the sugar-phosphate backbone can be estimated as four molecules per phosphate group and one per O5′ and O3′ atoms, yielding 12 water molecules per nucleotide pair. The O4′ atom has the common hydration sphere with the minor groove atoms O2 and N3, so that no water molecules are counted for this sugar atom. The base plus backbone ordered hydration accounts maximally for 17 water molecules per nucleotide pair. Thus, there are two to seven water molecules fewer than the estimate by Chalikian et al. These remaining water molecules could form a disordered and highly mobile part of the tightly bound hydration shell of B-DNA that would not be seen as single sites in crystals, but may be seen as spatial fluctuations of electron density after a proper refinement of good-quality crystal data (Badger and Caspar, 1991).

We thank Jaroslav Vojtechovsky and Jordi Bella for valuable discussions. The help of Julie Woda is gratefully acknowledged.

This research was supported by the grant Agency of the Czech Republic to B.S. (grant 204/95/0036) and by the National Institutes of Health to H.M.B. (grant GM21589). The Nucleic Acid Database is supported by the National Science Foundation (grant BIR 9305135).

REFERENCES

Badger, J., and D. L. D. Caspar. 1991. Water structure in cubic insulin crystals. *Proc. Natl. Acad. Sci. USA.* 88:622–626.

Baikalov, I., K. Grzeskowiak, K. Yanagi, J. Quintana, and R. E. Dickerson. 1993. The crystal structure of the trigonal decamer C-G-A-T-C-G-6 meA-T-C-G: a B-DNA helix with 10.6 base-pairs per turn. *J. Mol. Biol.* 231:768–784.

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

Berman, H. M., A. Sowri, S. L. Ginell, and D. L. Beveridge. 1988. A systematic study of patterns of hydration in nucleic acids. I. Guanine and cytosine. *J. Biomol. Struct. Dyn.* 5:1101–1110.

Brünger, A. T. 1992. X-Plor, Version 3.1. A System for X-Ray Crystallography and NMR. Yale University Press, New Haven, CT.

Carrell, H. L. 1979. BANG, A Program to Calculate Distances and Angles. Fox Chase Cancer Center, Philadelphia.

Chalikian, T. V., A. P. Sarvazyan, G. E. Plum, and K. J. Breslauer. 1994. Influence of base composition, base sequence, and duplex structure on DNA hydration: apparent molar volumes, and apparent molar adiabatic compressibilities of synthetic, and natural DNA duplexes at 25°C. *Biochemistry.* 33:2394–2401.

Chen, Y. Z., and E. W. Prohofsky. 1993. Synergistic effects in the melting of DNA hydration shell: melting of the minor groove hydration spine in poly(dA)·poly(dT) and its effect on base pair stability. *Biophys. J.* 64:1385–1393.

Drew, H. R., and R. E. Dickerson. 1981. Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* 151:535–556.

Drew, H. R., R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. 1981. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA.* 78:2179–2183.

Falk, M., A. G. Poole, and C. G. Goymour. 1970. Infrared study of the state of water in the hydration shell of DNA. *Can. J. Chem.* 48:1536–1542.

Goodsell, D. S., M. L. Kopka, D. Cascio, and R. E. Dickerson. 1993. Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. USA.* 90:2930–2934.

Grzeskowiak, K., K. Yanagi, G. G. Privé, and R. E. Dickerson. 1991. The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. The effect of base pair reversals. *J. Biol. Chem.* 266:8861–8883.

Heinemann, U., and C. Alings. 1989. Crystallographic study of one turn of G/C-rich B-DNA. *J. Mol. Biol.* 210:369–381.

Heinemann, U., C. Alings, and M. Bansal. 1992. Double helix conformation groove dimensions and ligand binding potential of a G/C stretch in B-DNA. *EMBO J.* 11:1931–1939.

Heinemann, U., and M. Hahn. 1992. C-C-A-G-G-C-m5C-T-G-G: helical fine structure, hydration, and comparison with C-C-A-G-G-C-C-T-G-G. *J. Biol. Chem.* 267:7332–7341.

Johnson, C. K. 1976. ORTEPII. Report ORNL-5138. Oak Ridge National Laboratory, Oak Ridge, TN.

Jones, T. A. 1978. FRODO: a graphic model building and refinement system for macromolecules. *J. Appl. Cryst.* 11:268–272.

Langan, P., V. T. Forsyth, A. Mahendrasingam, W. J. Pigram, S. A. Mason, and W. Fuller. 1992. A high angle neutron fiber diffraction study of the hydration of the A conformation of the DNA double helix. *J. Biomol. Struct. Dyn.* 10:489–503.

Langley, R. 1971. Practical Statistics. Dover, New York.

Lipanov, A., M. L. Kopka, M. Kaczor-Grzeskowiak, J. Quintana, and R. E. Dickerson. 1993. The structure of the B-DNA decamer C-C-A-A-C-I-

T-T-G-G in two different space groups: conformational flexibility of B-DNA. *Biochemistry.* 32:1373–1389.

Milton, J. G., and W. C. Galley. 1986. Evidence for heterogeneity in DNA-associated solvent mobility from acridine phosphorescence spectra. *Biopolymers.* 25:1673–1684.

Mrevlishvili, G. M. 1981. Relationship between hydration of natural DNA and its CG content. *Dokl. Akad. Nauk USSR.* 260:761–764.

Murray-Rust, P., and J. P. Glusker. 1984. Directional hydrogen bonding to sp^2- and sp^3-hybridized atoms and its relevance to ligand-macromolecule interactions. *J. Am. Chem. Soc.* 106:1018–1025.

Pitt, W. R., J. Murray-Rust, and J. M. Goodfellow. 1993. AQUARIUS2: knowledge-based modelling of solvent sites around proteins. *J. Comp. Chem.* 14:1007–1019.

Privé, G. G., U. Heinemann, S. Chandrasegaran, L. S. Kan, M. L. Kopka, and R. E. Dickerson. 1987. Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science.* 238:498–504.

Privé, G. G., K. Yanagi, and R. E. Dickerson. 1991. Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *J. Mol. Biol.* 217:177–199.

Quintana, J. R., K. Grzeskowiak, K. Yanagi, and R. E. Dickerson. 1992. The structure of a B-DNA decamer with a central T-A step: C-G-A-T-T-A-A-T-C-G. *J. Mol. Biol.* 225:379–395.

Rentzeperis, D., D. W. Kupke, and L. A. Marky. 1993. Volume changes correlate with entropies and enthalpies in the formation of nucleic acid homoduplexes: differential hydration of A and B conformations. *Biopolymers.* 33:117–125.

Roe, S. M., and M. M. Teeter. 1993. Patterns for prediction of hydration around polar residues in proteins. *J. Mol. Biol.* 229:419–427.

Rosenfield, R. E., Jr., S. M. Swanson, E. F. Meyer, Jr., H. L. Carrell, and P. Murray-Rust. 1984. Mapping the atomic environment of functional groups: turning 3D scatter plots into pseudo-density contours. *J. Mol. Graph.* 2:43–46.

Schneider, B., D. M. Cohen, and H. M. Berman. 1992. Hydration of DNA bases: analysis of crystallographic data. *Biopolymers.* 32:725–750.

Schneider, B., D. M. Cohen, L. Schleifer, A. R. Srinivasan, W. K. Olson, and H. M. Berman. 1993. A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophys. J.* 65:2291–2303.

Schreiner, L. J., J. C. MacTavish, M. M. Pintar, and A. Rupprecht. 1991. NMR spin grouping and correlation exchange analysis. *Biophys. J.* 59:221–234.

Sheldrick, G. M. 1993. SHELXL-93. A Program for Refinement of Crystal Structures. University of Göttingen, Göttingen.

Sponer, J., and P. Hobza. 1994. Nonplanar geometries of DNA bases. Ab initio second-order Môller-Plesset study. *J. Phys. Chem.* 98:3161–3164.

Sussman, J. L. 1984. BMF, A Program for Calculating the Best Molecular Fit. The Weizmann Institute, Rehovot, Israel.

Tunis, M.-J., and J. E. Hearst. 1968. On the hydration of DNA. II. Base composition dependence of the net hydration of DNA. *Biopolymers.* 6:1345–1353.

Umehara, T., S. Kuwabara, S. Mashimo, and S. Yagihara. 1990. Dielectric study on hydration of B-, A-, and Z-DNA. *Biopolymers.* 30:649–656.

Wang, J. H. 1955. The hydration of deoxyribonucleic acid. *J. Am. Chem. Soc.* 77:258–260.

Yuan, H., J. Quintana, and R. E. Dickerson. 1992. Alternative structures for alternating poly(dA-dT) tracts: the structure of the B-DNA decamer C-G-A-T-A-T-A-T-C-G. *Biochemistry.* 31:8009–8021.