A Systematic Method for Studying the Spatial Distribution of Water Molecules around Nucleic Acid Bases

Bohdan Schneider,‡ Dawn M. Cohen,* Leah Schleifer,‡ A. R. Srinivasan,‡ Wilma K. Olson,‡ and Helen M. Berman‡

Departments of ‡Chemistry and *Computer Science, Rutgers University, New Brunswick, New Jersey 08903 USA

ABSTRACT A new method to analyze the distribution of water molecules around the bases in DNA is presented. This method relies on the notion of a "hydrated building block," which represents the joint observed hydration around all bases of a particular type, in structures of a particular conformation type. The hydrated building blocks were constructed using atomic coordinates from 40 structures contained in the Nucleic Acid Database. Pseudoelectron densities were calculated for water molecules in each hydrated building block using standard crystallographic procedures. The electron densities were fitted to obtain "average building blocks," which represent bases with waters only at average or probable positions. Both types of building blocks were used to construct models of hydrated DNA oligomers. The essential features of the solvent structure around $d(CGCGAATTCGCG)_2$ in the B form and $d(CGCGCG)_2$ in the Z form were reproduced.

INTRODUCTION

It has been established for some time that there is a shell of tightly bound water molecules at the surface of DNA with properties significantly different from those of bulk water (Wang, 1955; Tunis and Hearst, 1968; Falk et al., 1970; Lavalle et al., 1990; Kubinec and Wemmer, 1992). It has remained difficult, however, to determine the locations of these bound waters and to determine how hydration structure is related to the sequence and conformation of DNA. Significant progress in that direction has been reported in a recent wide angle neutron scattering study of an A-DNA fiber, which shows localized sites of preferred hydration in the major groove (Langan et al., 1992).

Single-crystal x-ray crystallography offers another method for determining the positions of water molecules that are either localized in space or on a time average are found at particular sites in a specific crystal. Early studies focused on individual structures that displayed various intriguing water networks, including spines (Drew and Dickerson, 1981; Kopka et al., 1983; Chevrier et al., 1986) and pentagons (Neidle et al., 1980; Kennard et al., 1986) of hydration. However, it is only recently that it has become possible to examine a large database of nucleic acid structures (Berman et al., 1988; Schneider et al., 1992a; Westhof, 1987a) in order to decipher the principles that underlie the hydration of DNA.

In the analysis presented here, a new method has been developed that can quantify distributions of water molecules around DNA. It has already been shown (Schneider et al., 1992a) that the average positions of water molecules differ according to the conformational class of the DNA molecule. In the present study, a Fourier transform technique is used to represent the spatial density distributions of the water molecules around the bases in each conformational class. These densities are used to identify average hydration sites in the first hydration shell around DNA. In this way, hydration of DNA oligomers whose conformations are known can be described. This is a first step in being able to predict the possible networks of waters in specific sequences of DNA.

METHODS

Database search

The Nucleic Acid Database (NDB) (Berman et al., 1992) was used to search for coordinates of crystal structures of A, B, and Z oligodeoxynucleotides with at least two bases in a strand. DNA structures containing drugs were excluded from the search. Forty crystal structures containing a total of 644 bases met these criteria (Table 1). Of these, 586 are standard unmodified DNA bases that were used in this study. Twenty modified cytosines in Z-DNA structures were also considered.

The crystallographic resolution of the oligonucleotide structures included ranged from 0.8 to 3.0 Å. In most cases, the electron densities of light cations such as Mg^{2+} are rarely distinguishable from those of water molecules; however, any identified cations were omitted from this study.

Hydrated building blocks

For each of the bases in this study, all water molecules (modeled as oxygen atoms) within 3.2 Å of any heavy atom were calculated using the program BANG (Carrell, 1979). As shown previously (Schneider et al., 1992), the distance of 3.2 Å effectively distinguishes between potential hydrogen bonding contacts and van der Waals contacts. A water was counted once per base despite contacts to multiple atoms within the base. In cases in which a single water molecule contacted more than one base, it was counted once for each base contacted.

All the bases of a particular type were placed into a common Cartesian coordinate system with their associated water molecules as in our previous study (Schneider et al., 1992a). They were then separated by DNA conformational type. Occupancies were assigned to the water positions as $1/N$ where N is the number of bases in a particular DNA conformational class.

Received for publication 20 May 1993 and in final form 22 September 1993.

Address reprint requests to Dr. Helen M. Berman, Dept. of Chemistry, Wright and Rieman Laboratories, Rutgers University, P.O. Box 939, Piscataway, NJ 08855-0939.

Dr. Schneider's present address is Heyrovsky Institute of Physical Chemistry, Czech Academy of Sciences, 182 23 Prague, Czech Republic.

0006-3495/93/12/2291/13 $2.00

TABLE 1 Statistics of structures used in this study

NDB Code*	Sequence‡	R Factor§	Resolution‖	Water¶	Base Water**	Reference
ADDB01	CCGG	16.5	2.0	10.7	1.63	(Conner et al., 1984)
ADH007	GGGATCCC	16.6	2.5	0.6	0.38	(Lauble, et al., 1988)
ADH008	GCCCGGGC	17.1	1.8	4.2	1.38	(Heinemann, et al., 1987)
ADH023	GTACGTAC	17.0	2.4	7.0	0.25	(Courseille et al., 1990)
ADH024	GTACGTAC	18.4	2.2	6.5	0.63	(Takusagawa, 1990)
ADHP36	GCCCGGGC	16.0	2.1	3.6	1.88	(Heinemann et al., 1991)
ADJ022	ACCGGCCGGT	18.0	2.0	3.6	0.70	(Frederick et al., 1989)
ADL025	CCCCCGCGGGGG	17.7	2.0	4.5	0.29	(Verdaguer et al., 1991)
BDBP23	CG	10.9	0.8	3.0	2.00	(Han et al., 1990)
BCJ008	CCAAGATTGG	18.5	1.3	6.9	2.20	(Privé et al., 1987)
BDJ017	CCAGGCCTGG	16.0	1.6	4.2	1.3	(Heinemann and Alings, 1989)
BDJ019	CCAACGTTGG	16.0	1.4	7.7	2.30	(Privé et al., 1991)
BDJ025	CGATCGATCG	16.1	1.5	7.7	2.15	(Grzeskowiak et al., 1991)
BDJ031	CGATTAATCG	15.7	1.5	5.7	1.60	(Quintana et al., 1992)
BDJB27	CCAGGCCTGG	17.4	1.8	4.2	1.55	(Heinemann and Hahn, 1992)
BDL001	CGCGAATTCGCG	17.8	1.9	3.3	0.75	(Drew et al., 1981)
BDL002	CGCGAATTCGCG	15.1	2.7	3.5	0.46	(Drew et al., 1982)
BDL007	CGCATATATGCG	18.9	2.2	1.8	0.17	(Yoon et al., 1988)
BDL015	CGCAAAAATGCG	20.1	2.6	0.9	0.21	(DiGabriele et al., 1989)
BDL020	CGCGAATTCGCG	18.8	1.9	2.6	0.92	(Westhof, 1987b)
BDL022	CGCAAGCTGGCG	19.3	2.5	2.0	0.50	(Webster et al., 1990)
BDL028	CGTGAATTCACG	17.0	2.7	3.5	0.64	(Narayana et al., 1991)
BDL029	CGTGAATTCACG	15.8	2.5	1.5	0.64	(Larsen et al., 1991)
BDLB03	CGCGAATTCGCG	13.0	3.0	1.8	0.29	(Fratini et al., 1982)
BDLB04	CGCGAATTCGCG	17.3	2.3	4.8	0.83	(Fratini et al., 1982)
BDLB13	CGCGAATTCGCG	16.9	2.0	3.6	0.54	(Frederick et al., 1988)
BDLB26	CGCGAATTTGCG	18.5	2.0	2.9	0.96	(Leonard et al., 1990)
ZDB020	CG	13.6	0.8	3.2	2.25	(Ramakrishnan and Viswamitra, 1988)
ZDF001	CGCGCG	14.0	0.9	6.2	2.25	(Wang et al., 1979)
ZDF002	CGCGCG	17.5	1.0	7.0	3.17	(Gessner et al., 1989)
ZDF028	CGCGCG	19.8	1.2	5.7	3.33	(Kagawa et al., 1991)
ZDF029	CGCGCG	18.5	1.0	3.9	1.50	(Egli et al., 1991)
ZDFB04	CGCGCG	13.3	1.6	5.5	1.67	(Chevrier et al., 1986)
ZDFB05	CGCGCG	12.5	1.4	6.9	2.33	(Chevrier et al., 1986)
ZDFB10	CGUACG	20.9	1.3	7.2	2.50	(Geierstanger et al., 1991)
ZDFB11	CACGTG	21.7	1.3	6.9	1.92	(Coll et al., 1986)
ZDFB12	CGCGUG	17.2	1.5	4.8	1.67	(Coll et al., 1989)
ZDFB21	CGCGCG	19.0	1.9	5.0	1.92	(Ginell et al., 1990)
ZDFB24	CGUACG	20.8	1.3	5.0	1.50	(Zhou and Ho, 1990)
ZDFB31	CGUACG	13.8	1.3	7.0	2.67	(Schneider et al., 1992b)

* Structure identification used in the Nucleic Acid Database. The first letter gives conformational type.

‡ Sequence of oligonucleotide. Modified bases are underscored.

§ Crystallographic R factor: $R = \Sigma (F_{obs} - F_{calc})/\Sigma F_{obs}$ (%).

‖ Crystallographic resolution in Å at which the structure has been refined.

¶ Number of crystallographically ordered water molecules in the asymmetric unit divided by number of nucleotides in the asymmetric unit.

** Number of crystallographically ordered water molecules in the asymmetric unit within 3.2 Å of any base atom divided by number of nucleotides in the asymmetric unit.

For example, for the sample of 54 guanines in Z-DNA, the occupancy of each water atom was set to 1/54. A base with all of its associated water molecules is called a "hydrated building block" as illustrated in Fig. 1 *a*. The 12 possible hydrated building blocks are described in Table 2. No analyses of adenine and thymine building blocks in the A and Z conformations were carried out because of the limited sample.

Density representation of water positions

Electron densities were calculated for each hydrated building block using standard crystallographic procedures as implemented in the program X-PLOR (Brunger, 1990). A reverse Fourier transformation was used to calculate structure factors, $F(hkl)$, from the water positions. A second Fourier transform then converted these $F(hkl)$ values into electron densities. The calculations were done in two ways. First, the base and water molecules were included in the calculation to obtain the density shown in Fig. 1 *b*. Second, the base atoms were not included in the density calculation. This has the effect of enhancing the density of the water molecules, as shown in Fig. 1 *c*. Because each water molecule has partial occupancy as described above, the total electron density of the waters is equal to that of the average number of water molecules around the base. Since all bases of a given type are nearly identical, it is sufficient to consider a single base with the occupancy of its atoms set to 1.0, surrounded by water molecules of partial occupancy when computing density maps.

In doing these calculations, several parameters had to be considered including unit cell symmetry and dimensions, crystallographic resolution, temperature factors, and atom types. The space group was assumed to be triclinic P1. The unit cell dimensions used for this calculation were determined from the largest span of water coordinates in modeled bases or oligomeric sequences (see below for details about oligomeric sequences studied). On each side of the cell, an extra space of about 3 Å was left so that water molecules from two neighboring unit cells could not overlap. The cell dimensions for hydrated bases were $a = b = 15.0$ Å, and $c = 10.0$ Å. In all cases the angles of the unit cell were 90°. The c axis was approximately perpendicular to the base plane. The crystallographic resolution was set to 1.5 Å. This value is based on computer experiments with density maps of water molecules around guanine in B- and Z-DNA structures. All of the maps calculated at 2.8, 2.0, and 1.5 Å resolution

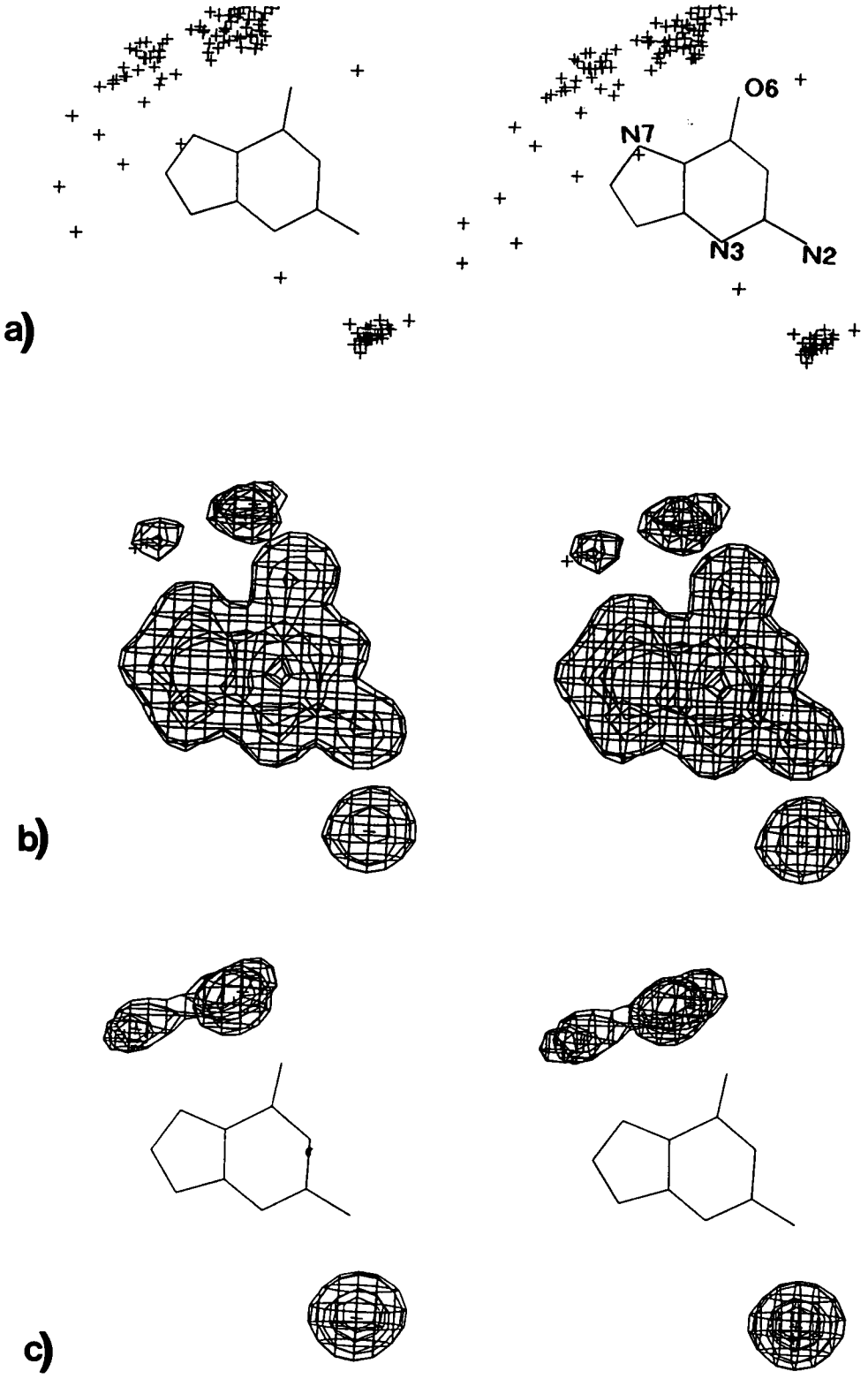

FIGURE 1 Comparison, in stereo, of three representations of water distributions around Z-DNA guanine bases. (*a*) The 136 point-like water molecules. (*b*) Electron density of guanine atoms with occupancies equal to 1.00 and 136 hydrogen-bonded water molecules found in 54 guanines. Every water molecule is modeled as an oxygen atom with an occupancy of 1/54. (*c*) Electron densities of the 136 water molecules alone.

were quite similar, but the 1.5 Å maps were sharpest and showed no spurious peaks. The temperature factor was set to a B value of 15. The density maps of water molecules in B- and Z-DNA guanine were calculated with B = 50, 25, and 15. The three corresponding maps were similar, but the map with B = 15 was easiest to interpret. As is usual in DNA crystallography, the water molecules were modeled as oxygen atoms.

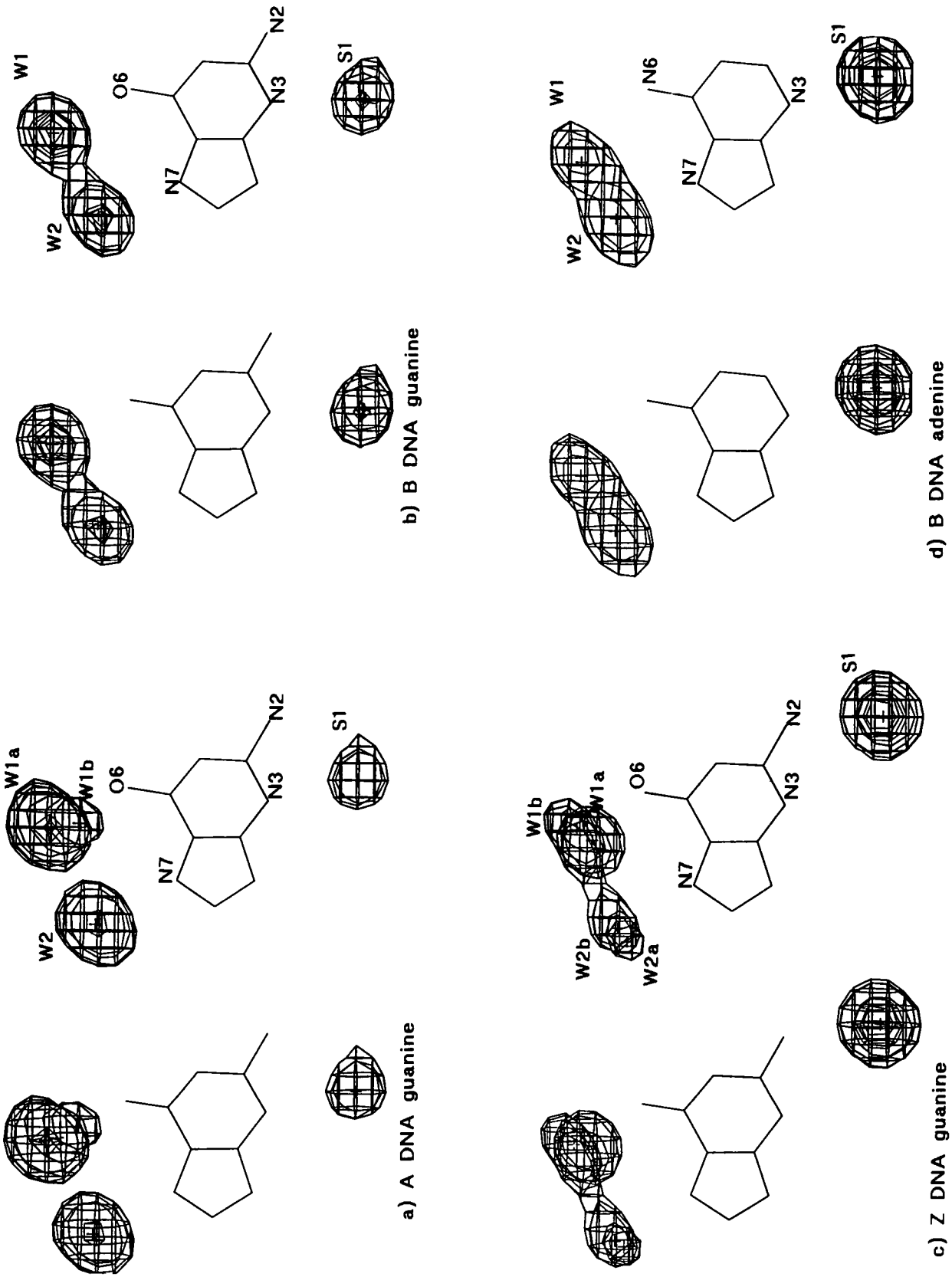

FIGURE 2 Density representations (as in Fig. 1 *c*) of hydration sites of purines. The minor groove hydration sites are labeled S and the major groove sites W. All figures are shown in stereo. (*a*) Guanine in A-DNA; (*b*) guanine in B-DNA; (*c*) guanine in Z-DNA; (*d*) adenine in B-DNA.

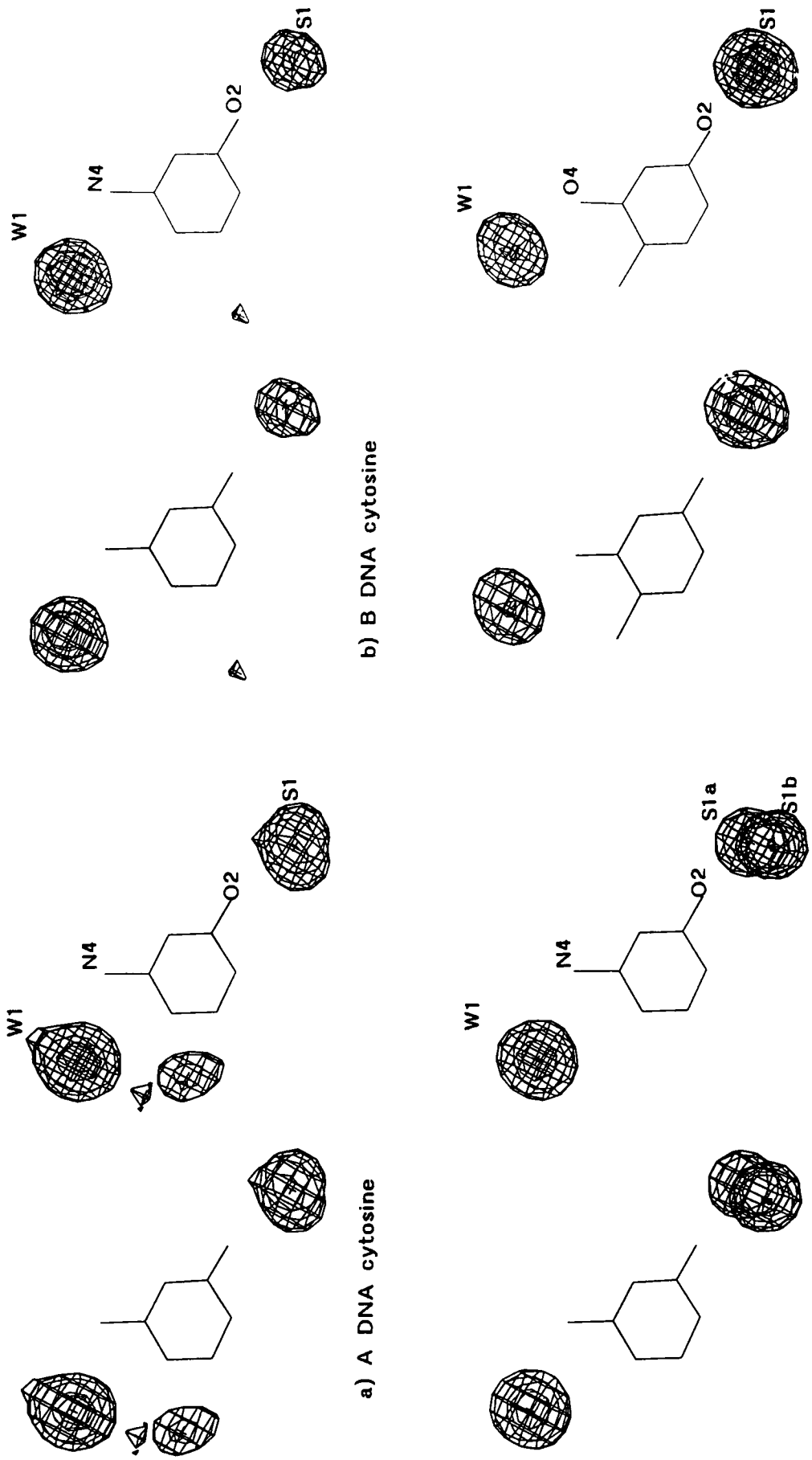

FIGURE 3 Density representations (as in Fig. 1 *c*) of hydration sites of pyrimidines. The minor groove hydration sites are labeled S and the major groove sites W. All figures are shown in stereo. (*a*) Cytosine in A-DNA; (*b*) cytosine in B-DNA; (*c*) cytosine in Z-DNA; (*d*) thymine in B-DNA.

TABLE 2 Statistics of hydrated building blocks

	Conformational Type			
	A-DNA	B-DNA	Z-DNA	TOTAL
Guanine				
Structures*	8	19	11	38
Bases‡	37 (1)	124 (2)	54 (12)	215 (15)
Waters§	36	124	136	
Adenine				
Structures	4	18	2	24
Bases	7 (0)	78 (2)	3 (5)	88 (7)
Waters	[2]	89	[5]	
Cytosine				
Structures	8	19	9	36
Bases	36 (2)	115 (6)	44 (20)	195 (28)
Waters	28	74	93	
Thymine				
Structures	4	18	1	23
Bases	7 (0)	79 (0)	2 (0)	88 (0)
Waters	[2]	83	[1]	

* Number of structures containing the specified base in this conformation.
‡ Number of bases in this conformation. The numbers in parentheses are the numbers of modified bases in this conformation.
§ Number of water molecules used to calculate an electron density map. The numbers in square brackets are too small to be used in the electron density calculation.

Average building blocks

In the next stage of the analysis, the electron density maps that were calculated with the bases removed were displayed using FRODO (Jones, 1978). The peak centers for each density were determined visually. As seen in Figs. 2 and 3, the shapes of the electron densities are such that it was relatively straightforward to do this. These fitted points represent "averaged" water positions and are referred to in the text as either "fitted water positions" or "hydration sites." Bases with water molecules fitted into the electron density peaks are called "average building blocks." An average building block contains the base atoms and a few (two to five) hydration sites, whereas a hydrated building block contains the base atoms and 28–136 water molecules.

Density of water molecules in DNA oligomers

In order to study the utility of the density method in predicting hydration structure, models of hydration of various oligonucleotides were constructed by substituting the hydrated building blocks for the bases in a DNA molecule. A hydrated B-DNA dodecamer with the sequence, d(CGCGAATTCGCG)$_2$, was built using the conformations observed in the crystal (Drew and Dickerson, 1981) and in the fiber (Chandrasekaran and Arnott, 1989). A model was also built for the sequence d(CGCGCG)$_2$ in the Z conformation observed in the crystal (Wang et al., 1979; Gessner et al., 1989). Electron density maps were calculated for models with hydrated building blocks, and the resultant densities were fit with water molecules. For the dodecamer, cell dimensions were set to $a = b = 35.0$ Å, and $c = 60.0$ Å, while for the Z-DNA hexamer $a = b = 25.0$ Å, and $c = 35.0$ Å, with all angles equal to 90°.

The same oligomers were also constructed using the averaged building blocks, in order to study their usefulness in making direct predictions of hydration structure, without the intervening density calculation step. In addition, the oligomers were constructed, using the *incorrect* conformational type averaged building blocks, in order to find out the importance of differences between the blocks.

RESULTS

Table 2 shows the number of water molecules contained in the hydrated building blocks as a function of base and conformational type. The following discussion describes the hydration of guanine and cytosine in A-, B-, and Z-DNA structures and of adenine and thymine in B-DNA structures. Fig. 2 illustrates the hydration densities of purines and Fig. 3 the hydration densities of pyrimidines calculated as electron densities of waters in the hydrated building blocks. The geometries and relative electron densities of the hydration sites for each base are given in Table 3. In that table and in the discussion that follows, the positions of hydration sites "above" and "below" base planes are described with respect to the 5′ and 3′ termini of the oligonucleotide strand. In Figs. 2 and 3, hydration sites above the plane of the base are in the direction of the 3′ terminus for A- and B-DNA, whereas they are in the direction of the 5′ terminus for Z-DNA. A hydration site is considered to be planar when the torsion angle connecting it to the base is within 10° from either 0° or 180°. We have adopted the convention of Seeman (Seeman et al., 1976) whereby hydration sites in the minor groove are labeled "S" and those in the major groove "W."

Hydration of bases

Guanine

A-DNA (Fig. 2 *a*, Table 3)

The minor groove has one low density, slightly unspherical hydration site. It contacts both N2 and N3 atoms and is centered in the base plane. The major groove has two main hydration sites that interact with O6 and N7. Both of these have the same density, are spherical, and lie close to the base plane (torsion angles 6° for W1a and −169° for W2). A second site of lower density contacts O6 only and is out of the base plane in the 3′ direction. The distances between the three major groove peaks are all favorable for forming hydrogen bonds with the bases and with each other.

B-DNA (Fig. 2 *b*, Table 3)

In the minor groove, there is a single out-of-plane hydration site which interacts only with the N3 atom. The guanines used to calculate the maps are from dodecamers and decamers. These two classes of structures differ in that many of the guanine N2 atoms in dodecamers are involved in packing and are blocked from hydration while such interactions are not observed in decamers. Nevertheless, the N2 of the guanine is almost completely unhydrated, indicating that its lack of hydration is *not* a packing effect. The major groove contains two hydration sites that lie close to the base plane at a distance favorable for forming water-water as well as water-base hydrogen bonds. The O6 site has a slightly higher density than the N7 site.

Z-DNA (Fig. 2 *c*, Table 3)

The minor groove has one hydration site that interacts solely with the N2 atom. In this conformation, the N3 atom is close to the sugar moiety and is, thus, blocked from hydration. The N2 peak is much higher in density than the major groove peaks and lies out of the base plane in the 5′ direction. The Z-DNA major groove hydration differs markedly from that of both A- and B-DNA. There

TABLE 3 Geometrical properties of the hydration sites in the average building blocks

Atoms			Hydration					
A	B	C	Site*	Distance‡	Angle‡	Torsion‡	Direction§	Height‖
				Guanine				
	A-DNA							
N1	C2	N2	S1	3.1	87	−179	plane	9
N1	C2	N3	S1	2.7	106	179	plane	9
C5	C6	O6	W1a	2.6	143	6	plane	19
C5	C6	O6	W1b	3.0	112	63	3′	6
N9	C8	N7	W2	2.7	117	−169	3′	17
		W1a	W1b	2.5				
		W1a	W2	3.2				
	B-DNA							
N1	C2	N3	S1	2.9	115	160	5′	17
C5	C6	O6	W1	2.7	140	9	plane	25
N9	C8	N7	W2	2.4	120	−177	plane	19
		W1	W2	2.8				
	Z-DNA							
N1	C2	N2	S1	3.1	123	158	5′	27
C5	C6	O6	W1a	2.8	114	50	5′	10
C5	C6	O6	W1b	2.5	137	−31	3′	7
N9	C8	N7	W2a	2.8	103	−155	5′	7
N9	C8	N7	W2b	2.8	114	160	3′	5
		W1a	W1b	2.9				
		W2a	W2b	2.1				
		W1a	W2a	3.2				
		W1b	W2b	3.1				
				Adenine				
	B-DNA							
N1	C2	N3	S1	2.7	102	170	plane	28
C5	C6	N6	W1	2.9	122	−6	plane	10
N9	C8	N7	W2	2.6	122	−173	plane	13
		W1	W2	2.0				
				Cytosine				
	A-DNA							
N1	C2	O2	S1	2.7	156	26	5′	15
N3	C4	N4	W1	2.8	111	−176	plane	22
N3	C4	C5	W2a	3.1	124	116	3′	7
N3	C4	C5	W2b	2.9	128	−138	5′	4
		W1	W2b	3.1				
	B-DNA							
N1	C2	O2	S1	2.9	158	55	5′	12
N3	C4	N4	W1	2.8	116	180	plane	34
C2	N1	C6	W2	2.8	138	−148	3′	4
	Z-DNA							
N1	C2	O2	S1a	2.8	140	−58	5′	16
N1	C2	O2	S1b	2.9	148	65	3′	13
		S1a	S1b	3.0				
N3	C4	N4	W1	2.9	113	172	plane	22
	Z-DNA, modified cytosines							
N1	C2	O2	S1a	3.0	133	−73	5′	12
N1	C2	O2	S1b	3.0	150	67	3′	14
		S1a	S1b	3.5				
N3	C4	N4	W1	2.9	142	178	plane	26
				Thymine				
	B-DNA							
N1	C2	O2	S1	2.7	160	73	5′	28
N3	C4	O4	W1	2.6	145	−176	plane	16

* Hydration sites are labeled S in the minor grooves and W in the major grooves and uniquely numbered. When there is more than one hydration site associated with one base atom then the sites are labeled by the same number with added a, b, etc. The same labeling is used in Figs. 2 and 3.

‡ Distances C-Site are in Ångstrom; angles B-C-Site and torsion angles A-B-C-Site are in degrees.

§ Direction indicates whether a hydration site lies above, below, or in the base plane. A hydration site is described as nonplanar if its torsion angle deviates more than ±10° from either 0° or 180°. The direction is described with respect to the 5′ or 3′ terminus of the oligonucleotide chain. Note the difference between the right-handed A- and B-DNA and left-handed Z-DNA.

‖ Heights of peaks are listed in sigma units. Sigma is defined as an estimated standard deviation of the density in a map:

$$\sigma = (\sum |x(i) - \langle x \rangle|^2/(n-1))^{1/2}$$

where $x(i)$ is the density at a grid point i of the map and $\langle x \rangle$ is the average density of the map. The map has n points.

are four peaks that form an approximately planar trapezoid perpendicular to the base plane. The distance between peaks W1a and W1b would permit hydrogen bonding whereas W2a and W2b are too close to be occupied by two water molecules at the same time. The latter sites may thus be either a part of a disordered system or else represent hydration sites populated by distinct and as yet unknown subclasses of Z-DNA.

Adenine

B-DNA (Fig. 2 *d*, Table 3)

The minor groove N3 contains one hydration site with density much higher than either site in the major groove.The minor groove site is centered slightly out of the plane in the 5′ direction (torsion 170°) but less than in the case of guanine in B-DNA (torsion 160°). The major groove has two hydration sites whose positions resemble those of the B-DNA guanine. However, W1 is pushed away from N6 and closer to N7. It also changes its direction relative to the base plane (torsion 9° in guanine, −6° in adenine). Since the position of the N7 hydration site (W2) is very similar in both bases, the distance between the W1 and W2 in adenine is too short to be a hydrogen bond. The explanation of the closeness of these two sites is similar to that offered in the case of the Z-DNA peaks (see above).

Cytosine

A-DNA (Fig. 3 *a*, Table 3)

The minor groove has one out-of-plane ellipsoidal hydration site whose major axis is perpendicular to the base plane. The major groove has three hydration sites. The N4 hydration site (W1) lies in the base plane and has the highest density of all the peaks. There are two weaker out of plane sites (W2a and W2b), one in the 3′ and the other in the 5′ direction, close to the C5 atom.

B-DNA (Fig. 3 *b*, Table 3)

The single minor groove peak deviates more from the base plane in the 5′ direction than the corresponding cytosine minor groove site in A-DNA (torsion angle 55° vs. 26°). The major groove is hydrated at two sites. The very dense N4 hydration site lies in the base plane and has higher density than the minor groove site. The other major groove site, near C5, is of marginal density and centered out of the base plane.

Z-DNA (Fig. 3 *c*, Table 3)

Unlike cytosine in A- and B-DNA, the minor groove is hydrated by two dense peaks whose centers lie symmetrically above and below the base plane. The distance between the peaks is such that they can form a hydrogen bond. The major groove has one spherical hydration site near N4 in the base plane, with only slightly higher density than the minor groove peaks.

For Z-DNA modified cytosines (C5-methylated or -brominated), the major groove peak W1 is shifted away from the C5 modifier relative to the unmodified cytosine: The C4-N4-W1 angle is 142° in modified and 113° in unmodified Z-DNA cytosines. The position of the W1 site is very similar to that of the W1 hydration site found for B-DNA thymine, which also has a hydrophobic group attached to the C5 atom. Surprisingly, the positions of the minor groove peaks S1a/b are also changed in the C5-modified cytosines. They deviate more from the base plane and are further from the O2 atom than in unmodified Z-DNA cytosines (torsion angles −73° and 67° for modified and −58° and 65° for unmodified cytosines). S1a and S1b are separated by 3.5 Å in modified Z-DNA cytosines and are, thus, too far apart to form a hydrogen bond.

Thymine

B-DNA (Fig. 3 *d*, Table 3)

The minor groove atom O2 is hydrated at one site (S1) that is more out of the plane than the comparable S1 peak in B-DNA cytosine (torsion angle 73° in thymine, 55° in cytosine). Unlike the B-DNA cytosine, the planar major groove hydration site W1 has a lower density than the minor groove site S1. Waters at the W1 site are apparently repelled by the thymine methyl group, so that the W1 position is different compared to the B-DNA cytosine W1 site.

Models of hydrated oligonucleotides

As described under "Methods," models were built for selected oligonucleotides, to determine if it would be possible to use the building blocks to predict water structures as a function of conformation and sequence.

$d(CGCGAATTCGCG)_2$

The crystal structure of this dodecamer (Drew et al., 1981) forms a double helix in the B conformation with a "spine of hydration" in the minor groove (Drew and Dickerson, 1981). The portions of this water network involving the DNA bases consist of two types of water bridges. One connects the O2 atom of a thymine in one strand to the N3 atom of an adenine one step down in the 3′ direction of the other strand; the other water connects the O2 of a thymine in one strand with the O2 of a thymine one step down in the 3′ direction of the other strand (Fig. 4 *a*). In the crystal structure, second shell water molecules "zip" the spine by making water-water-water bridges. Examination of these outer waters, however, is beyond the scope of this study.

To test the utility of hydrated building blocks for predicting water structure, two models of the dodecamer were built with conformations identical to that observed in the crystal structure (Drew et al., 1981). One model was built by substituting the DNA bases with the "hydrated building blocks." The density map of this structure is shown in Fig. 5. The peaks in this map were fitted, and it was found that each of the water molecules in the crystal had a predicted counterpart; the average distance between the predicted and observed peaks is 0.64 Å. The highest peaks are in the central

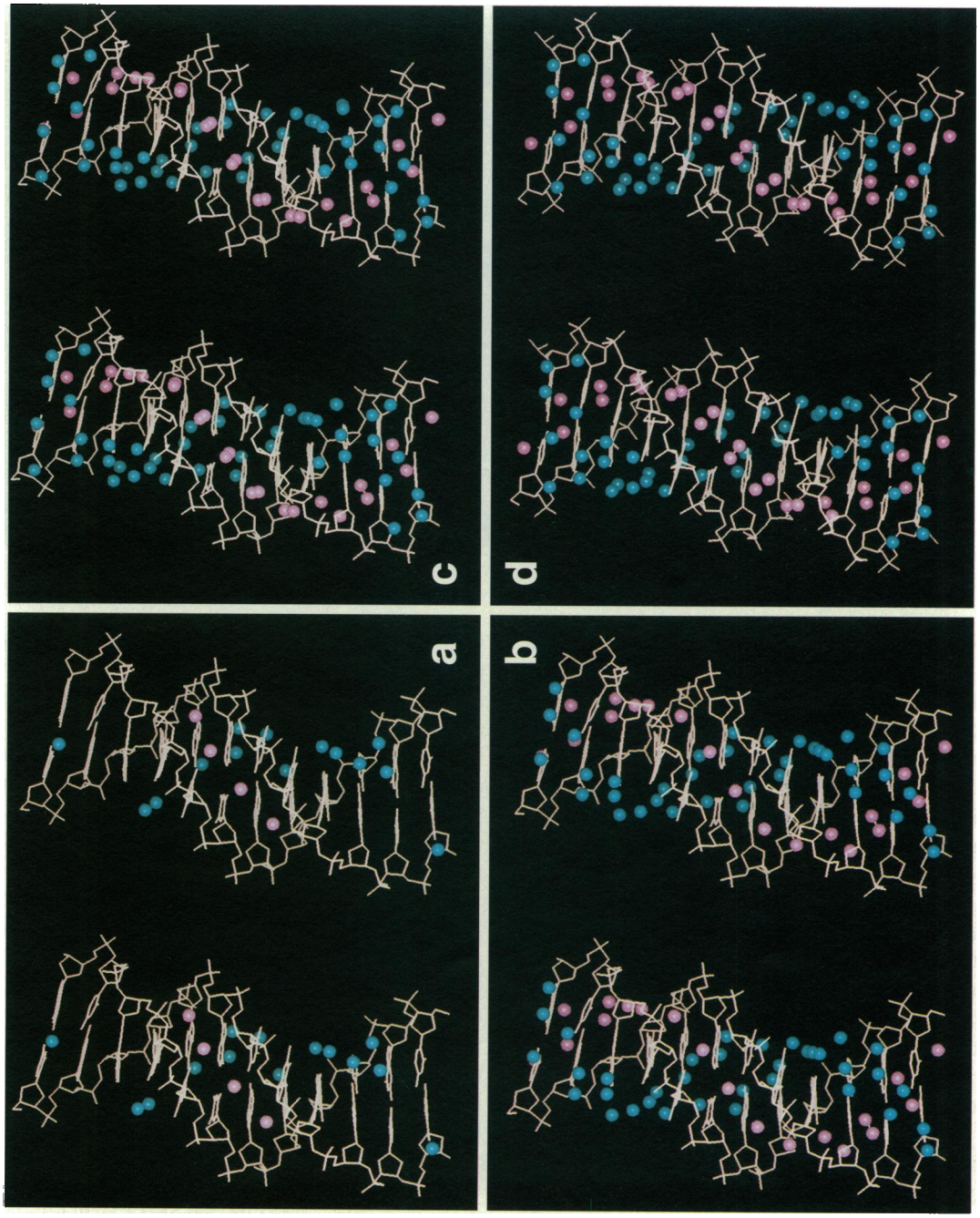

FIGURE 4 Comparison of different models of hydration in d(CGCGAATTCGCG)$_2$ (in stereo). (*a*) Crystallographically determined water positions around the dodecamer crystal structure (Drew et al., 1981). Only water molecules within 3.2 Å from any base atom are drawn. (*b*) Density-fitted hydration sites around the dodecamer in the crystal conformation. (*c*) Hydration site positions are calculated by substituting the average building blocks for bases in the crystal conformation. (*d*) Density-fitted water positions around the fiber model (Chandrasekaran and Arnott, 1989). The DNA is colored grey, the minor groove water molecules magenta, and the major groove water molecules cyan.

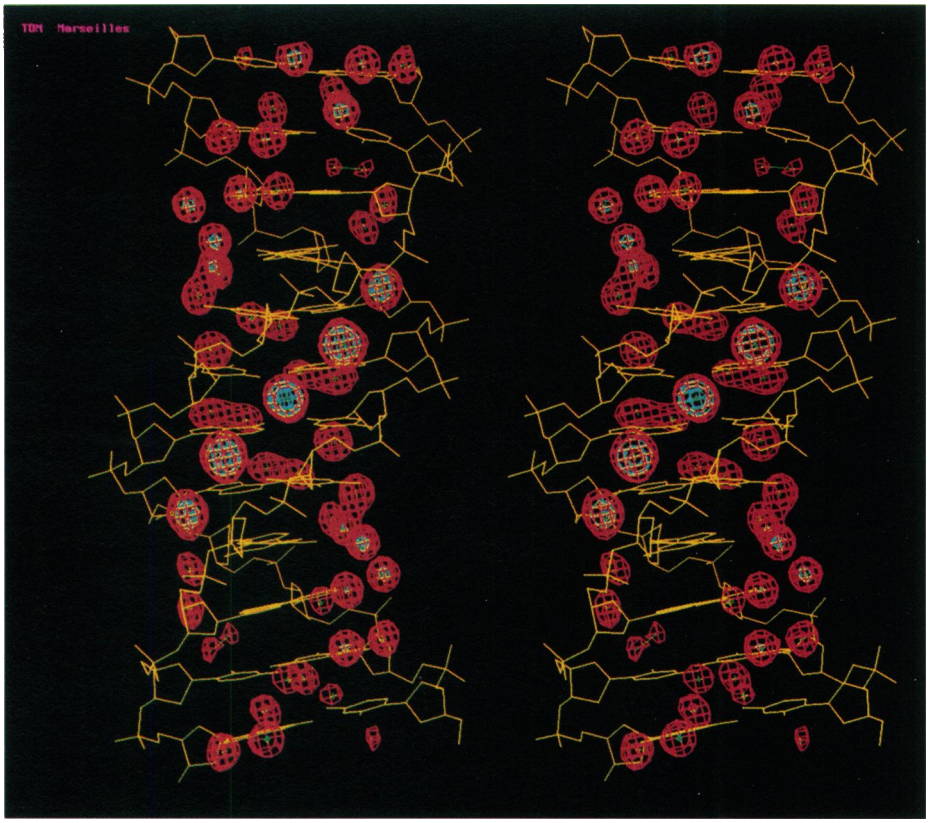

FIGURE 5 The electron density of modeled hydration around the dodecamer crystal structure (Drew et al., 1981). Densities were drawn at 8 (red) and 24 (cyan) sigma levels. DNA is in yellow.

AATT region of the minor groove corresponding to the spine of hydration observed in the crystal. There are 14 additional peaks of lower densities in the minor groove, including double rows of hydration sites at the flanking CGCG sequences (Fig. 4 *b*). Some of these sites are precluded because of packing in this particular crystal form. However, they apparently exist in other crystal forms or could in solution. In the major groove, all 14 water molecules that were observed in the crystal are reproduced in the model. The additional peaks may represent low occupancy sites in the crystal or sites available in other crystal forms, and they could certainly exist in solution.

The other model of the hydrated dodecamer was built with the average building blocks. For models made with the average building blocks, no density was calculated. The resulting hydration sites are nearly identical to those fitted in the density map, with one major exception: the single water bridge O2(T)-water-O2(T) observed in the spine of hydration of the crystal structure is replaced by doublets of hydration sites separated by 0.6 Å (Fig. 4 *c*). This formation of doublets reflects the accuracy with which the simple averaged blocks can reproduce individual features of a particular crystal structure.

Another model of the dodecamer sequence was built using hydrated building blocks, in which the B-DNA has the conformation found in fibers (Chandrasekaran and Arnott, 1989). The positions of density-fitted hydration sites are shown in Fig. 4 *d*. In this model, instead of single hydration sites along the spine of hydration, there are doublets of hydration sites 1.5–1.8 Å apart. Nevertheless, the principal features of the first hydration sphere are preserved. It is clear that small conformational changes in the fiber structure would allow these peaks to fuse into the single hydration sites observed in the crystal. Incorporation of the average building blocks into the fiber conformation yielded almost identical hydration sites as those obtained when the densities of the hydrated building blocks were fitted.

As a test of the uniqueness of the building blocks, A- and Z-DNA averaged building blocks were inserted into the fiber and crystal structures of the B-DNA dodecamer. It is important to note that when B-DNA building blocks are used, there are no steric clashes (contacts less than 2.4 Å) between the hydration centers and any atom of the DNA molecule. On the other hand, the use of A-DNA building blocks in the fiber model produces steric clashes between hydration sites and both stacked bases and backbone. Clashes with bases involve the O4 atoms of thymine and the O6 atoms of guanine, while those with backbone include the deoxyribose sugar C2′ atoms. The water molecules involved in these clashes are in major groove hydration sites attached to guanine O6 and cytosine C5. There are more poor contacts when the crystal conformation is used. These contacts, which come as close as 1.5 Å, also involve major groove heteroatoms and the sugar C2′ atoms.

When the Z-DNA building blocks are incorporated in the B-DNA fiber model, all steric clashes involve the sugar atoms contacting the minor groove S1a hydration sites asso-

ciated with the O_2 atom of cytosine. There are additional bad contacts with major groove base atoms when the crystal conformation is used.

d(CGCGCG)$_2$

The crystal structure of this hexamer consists of a double helix in the left-handed Z-DNA conformation (Wang et al., 1979; Gessner et al., 1989) with a network of water molecules in the minor groove. In this case, the network is composed of O2-water-O2 bridges between the cytosines. The hydration sites associated with the guanines are isolated and thus do not contribute to the network.

Two models of the hydrated Z-DNA that were built produced results analogous to those for the B-DNA dodecamer. In the model in which hydrated building blocks were substituted for the bases and hydration sites are fitted into calculated electron densities, the minor groove network of water molecules connecting the O2 atoms of the cytosines was reproduced. In the model made with the averaged building blocks, the minor groove hydration sites appear as doublets about 0.6 Å apart. The differences between the model constructed with the hydrated building blocks and averaged building blocks are greater in the case of Z-DNA than of B-DNA.

DISCUSSION

A method to obtain average water positions from experimentally determined single crystal structures has been described. Although this method is quite similar to one recently described for studying protein hydration (Pitt et al., 1993), it differs in that it uses Fourier transforms to produce the average densities. The identified positions that are extracted from these densities are interpreted as sites of dominant hydration. By sampling all bases in all nucleic acid crystal structures, it is possible to generalize hydration patterns observed in individual crystal structures and in some important cases to test the relevance of crystal packing on hydration. For example, this study shows that the guanine N2 atom in B-DNA is not hydrated regardless of the packing of B-DNA double helices in different crystal forms.

The analysis demonstrates that there are clear differences in the distributions of hydration sites around guanines and cytosines in the A-, B-, and Z-DNA helical conformations. Such a comparison cannot yet be made for adenine and thymine because of the lack of appropriate A- and Z-DNA structures. While it is not surprising that the water distributions in the left-handed structures are not the same as those in the right-handed structures, it is noteworthy that there are distinctive differences between the spatial arrangements of hydration sites between the right-handed A and B conformations. That these differences are significant is best illustrated by the fact that A-DNA building blocks are incompatible with B-DNA helices.

Not only are the hydration sites positioned differently, but their densities are different in different DNA conformational types. The density of a hydration site is a measure of the probability of water occurring at this site. High densities correspond to sites of preferred hydration. A quantitative way of correlating densities of hydration sites to probabilities of water occurrence is being developed.

Some systematic differences in hydration densities may be related to groove dimensions. The hydration sites in the narrow, sterically more restricted minor groove in B- and Z-DNA have a higher density than the corresponding sites in the open minor groove of A-DNA. Similarly, the hydration sites in the deep major groove in A-DNA are better defined than those in the shallow major groove of B-DNA and in the convex major groove surface of Z-DNA.

There are some open issues associated with this method of analyzing water distributions. The method predicts more hydration sites than there are water positions in the actual crystal structures. Because it uses the information about water positions in many crystal structures, it is not surprising that the water positions in any one crystal structure are a subset of predicted (possible) hydration sites. It is possible that the occupancies of these were too low and therefore unobserved in some crystal structures. In other cases, water positions may be predicted in areas that are involved in packing in a specific crystal. Thus, these water positions might be valid for the DNA in another crystal space group or in solution. In some instances, the method identifies hydration sites that are too close to be occupied simultaneously by two water molecules as in the case of hydration of the B-DNA adenine major groove atoms N6 and N7. Three possible factors may contribute to this apparent contradiction. i) Both positions are correct, and each is only partially occupied by dynamically disordered water. ii) Crystallographically determined water positions have positional errors that are propagated in the determination of hydration sites by the density method. The models made with the averaged building blocks indicate that if the discrepancies between the predictions and the crystallographically observed patterns are to be accounted for by this mechanism, the error propagated from the individual water positions is about 0.5 Å. iii) There are factors other than conformational class and base type codetermining hydration around the bases. These factors are currently under investigation using pattern recognition and machine learning techniques (Cohen et al., 1993).

The method will be useful in situations where the detailed conformation of a particular DNA sequence is known as is illustrated by the modeling of the B-DNA dodecamer crystal conformation. This approach may find ultimate utility in the last stages of crystallographic refinement to help interpret experimental difference maps when solvent must be positioned. However, until there are sufficient data available, such an approach must be used with caution. It also appears to predict a plausible model from which the hydration structure around canonical DNA may be inferred. When a considerable new body of structures

become available, it will be instructive to repeat this study to obtain a more detailed understanding of the hydration of bases.

Thus, we have provided an initial framework for studying the interrelationship between DNA conformation and hydration, solvent-induced transitions in the DNA double helix, and DNA-protein and DNA-drug recognition. Work is in progress to achieve these goals.

We thank Benson Yang and Greg Kreymer for helping with the calculations and Debbie Smith and Jordi Bella for helpful discussions.
This research was supported by National Institutes of Health Grants GM21589 (H.M.B.) and GM20861 (W.K.O.). Support for D.M.C. was provided in part from a Molecular Biophysics Training Grant GM 08319. The Nucleic Acid Database is supported by National Science Foundation Grant BIR N012772.

REFERENCES

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

Brunger, A. T. 1990. X-PLOR Manual (Version 2.1): A System for Crystallography and NMR. Yale University Press, New Haven, CT. 1–291.

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

Chandrasekaran, R., and S. Arnott. 1989. The structures of DNA and RNA helices in oriented fibers. *In* Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology, New Series, Group VII: Biophysics, Vol. 1. Subvol. b. Crystallographic and Structural Data. W. Saenger, editor. Springer-Verlag, Berlin. 31–170.

Chevrier, B., A. C. Dock, B. Hartmann, M. Leng, D. Moras, M. T. Thuong, and E. Westhof. 1986. Solvation of the left-handed hexamer d(5BrC-G-5BrC-G-5BrC-G) in crystals grown at two temperatures. *J. Mol. Biol.* 188:707–719.

Cohen, D. M., C. Kulikowski, and H. M. Berman. 1993. Knowledge based generation of machine learning experiments: learning with DNA crystallography data. *In* Proceedings of the First International Conference Intelligent Systems on Molecular Biology. L Hunter, D. Searls, and J. Shavlik, editors. AAAI/MIT Press. 92–100.

Coll, M., D. Saal, C. A. Frederick, J. Aymami, A. Rich, and A. H.-J. Wang. 1989. Effects of 5-fluorouracil/guanine wobble base pairs in Z-DNA: molecular and crystal structure of d(CGCGFG). *Nucleic Acids Res.* 17: 911–923.

Coll, M., A. H.-J. Wang, G. A. van der Marel, J. H. van Boom, and A. Rich. 1986. Crystal structure of a Z-DNA fragment containing thymine/2-aminoadenine base pairs. *J. Biomol. Struct. Dyn.* 4:157–172.

Conner, B. N., C. Yoon, J. L. Dickerson, and R. E. Dickerson. 1984. Helix geometry and hydration in an A-DNA tetramer: ^{I}C-C-G-G. *J. Mol. Biol.* 174:663–695.

Courseille, C., A. Dautant, M. Hospital, B. L. D'Estaintot, G. Precigoux, D. Molko, and R. Téoule. 1990. Crystal structure analysis of an A-DNA octamer d(GTACGTAC). *Acta Crystallogr.* A46:FC9-FC12.

DiGabriele, A. D., M. R. Sanderson, and T. A. Steitz. 1989. Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natl. Acad. Sci. USA.* 86: 1816–1820.

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., S. Samson, and R. E. Dickerson. 1982. Structure of a B-DNA dodecamer at 16K. *Proc. Natl. Acad. Sci. USA.* 79:4040–4044.

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.

Egli, M., L. D. Williams, Q. Gao, and A. Rich. 1991. Structure of the pure-spermine form of Z-DNA (magnesium free) at 1-Å resolution. *Biochemistry* 30:11388–11402.

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.

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

Frederick, C. A., G. J. Quigley, M.-K. Teng, M. Coll, G. A. van der Marel, J. H. van Boom, A. Rich, and A. H.-J. Wang. 1989. Molecular structure of an A-DNA decamer d(ACCGGCCGGT). *Eur. J. Biochem.* 181: 295–307.

Frederick, C. A., G. J. Quigley, G. A. van der Marel, J. H. van Boom, A. H.-J. Wang, and A. Rich. 1988. Methylation of the *Eco*RI recognition site does not alter DNA conformation: the crystal structure of d(CGCGAm^{6}ATTCGCG) at 2.0 Å resolution. *J. Biol. Chem.* 263: 17872–17879.

Geierstanger, B. H., T. F. Kagawa, S.-L. Chen, G. J. Quigley, and P. S. Ho. 1991. Base specific binding of copper (II) to Z-DNA: the 1.3-Å single crystal structure of d(m^{5}CGUAm^{5}CG) in the presence of $CuCl_2$. *J. Biol. Chem.* 266:20185–20191.

Gessner, R. V., C. A. Frederick, G. J. Quigley, A. Rich, and A. H.-J. Wang. 1989. The molecular structure of the left-handed Z-DNA double helix at 1.0-Å atomic resolution. *J. Biol. Chem.* 264:7921–7935.

Ginell, S. L., S. Kuzmich, R. A. Jones, and H. M. Berman. 1990. Crystal and molecular structure of a DNA duplex containing the carcinogenic lesion O^6-methylguanine. *Biochemistry.* 29:10461–10465.

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.

Han, F., W. Watt, D. J. Duchamp, L. Callahan, F. J. Kezdy, and K. Agarwal. 1990. Molecular structure of deoxycytidyl-3′-methylphosphonate (R_P) 5′-deoxyguanidine, d[Cp(CH3)G]. A neutral dinucleotide with Watson-Crick base pairing and a right handed helical twist. *Nucleic Acids Res.* 18:2759–2767.

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., 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.

Heinemann, U., H. Lauble, R. Frank, and H. Bloecker. 1987. Crystal structure analysis of an A-DNA fragment at 1.8 Å resolution: d(GCCCGGGC). *Nucleic Acids Res.* 15:9531–9550.

Heinemann, U., L.-N. Rudolph, C. Alings, M. Morr, W. Heikens, R. Frank, and H. Bloecker. 1991. Effect of a single 3′-methylene phosphonate linkage on the conformation of an A-DNA octamer double helix. *Nucleic Acids Res.* 19:427–433.

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

Kagawa, T. F., B. H. Geierstanger, A. H.-J. Wang, and P. S. Ho. 1991. Covalent modification of guanine bases in double stranded DNA: the 1.2-Å Z-DNA structure of d(CGCGCG) in the presence of $CuCl_2$. *J. Biol. Chem.* 266:20175–20184.

Kennard, O., W. B. T. Cruse, J. Nachman, T. Prange, Z. Shakked, and D. Rabinovich. 1986. Ordered water structure in an A-DNA octamer at 1.7 Å resolution. *J. Biomol. Struct. Dyn.* 3:623–647.

Kopka, M. L., A. V. Fratini, H. R. Drew, and R. E. Dickerson. 1983. Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* 163:129–146.

Kubinec, M. G., and D. E. Wemmer. 1992. NMR evidence for DNA bound water in solution. *J. Am. Chem. Soc.* 114:8739–8740.

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

Larsen, T. A., M. L. Kopka, and R. E. Dickerson. 1991. Crystal structure

analysis of the B-DNA dodecamer CGTGAATTCACG. *Biochemistry.* 30:4443–4449.

Lauble, H., R. Frank, H. Bloecker, and U. Heinemann. 1988. Three-dimensional structure of d(GGGATCCC) in the crystalline state. *Nucleic Acids Res.* 16:7799–7816.

Lavalle, N., S. A. Lee, and A. Rupprecht. 1990. Counterion effects on the physical properties and the A-transition to B-transition of calf-thymus DNA films. *Biopolymers.* 30:877–887.

Leonard, G. A., J. Thomson, W. P. Watson, and T. Brown. 1990. High-resolution structure of a mutagenic lesion in DNA. *Proc. Natl. Acad. Sci. USA.* 87:9573–9576.

Narayana, N., S. L. Ginell, I. M. Russu, and H. M. Berman. 1991. Crystal and molecular structure of a DNA fragment: d(CGTGAATTCACG). *Biochemistry.* 30:4449–4455.

Neidle, S., H. M. Berman, and H. S. Shieh. 1980. Highly structured water network in crystals of a deoxydinucleoside-drug complex. *Nature. (Lond.).* 288:129–133.

Pitt, W. R., J. Murray-Rust, and J. Goodfellow. 1993. Aquarius 2: knowledge-based modeling of solvent sites around proteins. *J. Comp. Chem.* 14:1007–1018.

Privé, G. G., U. Heinemann, S. Chandrasegaran, L. 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. 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.

Ramakrishnan, B., and M. A. Viswamitra. 1988. Crystal and molecular structure of the ammonium salt of the dinucleoside monophosphate d(CpG). *J. Biomol. Struct. Dyn.* 6:511–523.

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

Schneider, B., S. L. Ginell, R. Jones, B. Gaffney, and H. M. Berman. 1992b. Crystal and molecular structure of a DNA fragment containing a 2-aminoadenine modification: the relationship between conformation, packing, and hydration in Z-DNA hexamers. *Biochemistry.* 31:9622–9628.

Seeman, N. C., J. M. Rosenberg, and A. Rich. 1976. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA.* 73:804–808.

Takusagawa, F. 1990. The crystal structure of d(GTACGTAC) at 2.25 Å resolution: are the A-DNA's always unwound approximately 10 degrees at the C-G steps? *J. Biomol. Struct. Dyn.* 7:795–809.

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

Verdaguer, N., J. Aymami, D. Fernandez-Forner, I. Fita, M. Coll, T. Huynh-Dinh, J. Igolen, and J. A. Subirana. 1991. Molecular structure of a complete turn of A-DNA. *J. Mol. Biol.* 221:623–635.

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

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

Webster, G. D., M. R. Sanderson, J. V. Skelly, S. Neidle, P. F. Swann, B. F. Li, and I. J. Tickle. 1990. Crystal structure and sequence-dependent conformation of the A·G mis-paired oligonucleotide d(CGCAAGCTG-GCG). *Proc. Natl. Acad. Sci. USA.* 87:6693–6697.

Westhof, E. 1987*a*. Hydration of oligonucleotides in crystals. *Int. J. Biol. Macromol.* 9:186–192.

Westhof, E. 1987*b*. Re-refinement of the B-dodecamer d(CGCGAA-TTCGCG) with a comparative analysis of the solvent in it and in the Z-hexamer d(5BrCG5BrCG5BrCG). *J. Biomol. Struct. Dyn.* 5: 581–600.

Yoon, C., G. G. Privé, D. S. Goodsell, and R. E. Dickerson. 1988. Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc. Natl. Acad. Sci. USA.* 85:6332–6336.

Zhou, G. W., and P. S. Ho. 1990. Stabilization of Z-DNA by demethylation of thymine bases: 1.3-Å single-crystal structure of d(m^5CGUAm^5CG). *Biochemistry.* 29:7229–7236.