Article No. jmbi.1999.3323 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 294, 1149–1158

JMB

C-H..O Hydrogen Bonds in Minor Groove of A-tracts in DNA Double Helices

Anirban Ghosh and Manju Bansal*

Molecular Biophysics Unit
Indian Institute of Science
Bangalore, 560012, India

Analysis of available *B*-DNA type oligomeric crystal structures as well as protein-bound DNA fragments (solved using data with resolution <2.6 Å) indicates that in both data sets, a majority of the (3′-Ade) H2..O2(3′-Thy/Cyt) distances in AA.TT and GA.TC dinucleotide steps, are considerably shorter than their values in a uniform fibre model, and are smaller than their optimum separation distance. Since the electropositive C2-H2 group of adenine is in close proximity of the electronegative keto oxygen atoms of both pyrimidine bases in the antiparallel strand of the double-helical DNA structures, it suggests the possibility of intra-base-pair as well as cross-strand C-H..O hydrogen bonds in the minor groove. The C2-H2..O2 hydrogen bonds within the A.T base-pairs could be a natural consequence of Watson-Crick pairing. However, the close cross-strand interactions between the bases at the 3′-ends of the AA.TT and GA.TC steps arise due to the local sequence-dependent geometry of these steps. While the base-pair propeller twist in these steps is comparable to the fibre model, some of the other local parameters such as base-pair opening angle and inter-base-pair slide show coordinated changes, leading to these shorter C2-H2..O2 distances. Hence, in addition to the well-known minor groove hydration, it appears that favourable C2-H2..O2 cross-strand interactions may play a role in imparting a characteristic geometry to AA.TT and GA.TC steps, as well as $A_n.T_n$ and $GA_n.T_nC$ tracts, which leads to a narrow minor groove in these regions.

Keywords: C-H..O hydrogen bonds; C-H..N hydrogen bonds; cross-strand hydrogen bonds; DNA minor groove; A-tract structure

*Corresponding author

Introduction

Sequence-dependent conformational heterogeneity in DNA structure has been under intense scrutiny, ever since it was first observed in the *B*-DNA type oligomer crystal structure reported by Dickerson's group (Drew *et al*., 1981). Several analyses of the available crystal structures have attempted to establish relationships between the observed variation in the base-pair parameters, of the ten dinucleotide sequences, and certain intrinsic features, such as differences in the size of the purine and pyrimidine bases, base-stacking energy or nature of the exocyclic groups (Dickerson & Drew, 1981; Bhattacharyya & Bansal, 1990; Mohanty & Bansal, 1991; Hunter, 1993; Gorin *et al*., 1995; Hunter & Lu, 1997; Suzuki *et al*., 1997; El Hassan & Calladine, 1996, 1997; Dickerson, 1998; Olson *et al*., 1998). Extrinsic forces arising from crystal lattice packing or due to tightly bound solvent molecules are expected to influence the finer details of the observed structure (DiGabriele *et al*., 1989; DiGabriele & Steitz, 1993; Dickerson *et al*., 1994; Shui *et al*., 1998a,b). In addition, binding of proteins or any other ligand to DNA in either of the two grooves, leads to certain well-defined changes in the DNA structure, though these vary from almost negligible effects in the case of some regulatory protein-bound DNA, to very large bending in the case of CAP, TBP and IHF complexes (Dickerson, 1998). We have recently analysed some common sequence-dependent features in DNA oligomer and nucleic acid-regulatory protein-bound complex structures and noticed that the AA and GA steps show remarkable similarity in both data sets. We found that these dinucleotide steps apparently take up geometries that are stabilized by favourable C-H..O interactions (hydrogen bonds) in the minor groove, similar to those reported recently in several other organic molecules (Taylor

E-mail address of the corresponding author:
mb@mbu.iisc.ernet.in

0022-2836/99/501149–10 $30.00/0

& Kennard, 1982; Desiraju, 1991, 1996; Steiner & Saenger, 1993) nucleic acid structures (Leonard *et al.*, 1995, Wahl & Sundaralingam, 1997; Auffinger & Westhof, 1998), proteins (Derewenda *et al.*, 1995; Bella & Berman, 1996; Chakrabarti & Chakrabarti, 1998) and protein-DNA complexes (Mandel-Gutfreund *et al.*, 1998).

Cross-strand interactions in the DNA major groove

Favourable cross-strand interactions, or bifurcated hydrogen bonds were first identified between the 6-amino group of adenine and the 4-keto oxygen of thymine, in the major groove of an A-tract containing dodecamer crystal structure (Nelson *et al.*, 1987). The high propeller twist in the AT base-pairs is held to be the cause, as well as a consequence of such hydrogen bonds, in this and other A-tract-containing structures (Edwards *et al.*, 1992; DiGabriele & Steitz, 1993; Shatzky-Schwartz *et al.*, 1997) while the occurrence of a narrow minor groove seems to be a more ubiquitous feature of all A + T-rich sequences (Yoon *et al.*, 1988; Bhattacharyya & Bansal, 1992). N-H..N type of major groove cross-strand hydrogen bonds have been reported between the 6-amino groups of both the 5′-adenine bases in AT.AT steps, as well as with the 4-amino group of cytosine, in the inosine-containing steps AI.CT (Sponer & Kypr, 1994; Shatzky-Schwartz *et al.*, 1997; Luisi *et al.*, 1998). However, such cross-strand interactions have not been examined for other sequences, which also have potential to form cross-strand hydrogen bonds in the major groove, viz. GG.CC, AC.GT, CA.TG sequences (N-H..O) and AG.CT, CG.CG sequences (N-H..N). The possibility of such hydrogen bonds in the minor groove has not been explored, even though an N-H..O hydrogen bond was mentioned as being present at the AG.AT steps in the A.G mismatch base-pair-containing decamer structure (Prive *et al.*, 1987). In addition, an N..O or N..N separation of less than 3.6 Å has generally been used to assign a hydrogen bond, while the H..O or H..N distance and the N-H..O (or N) angle, which would be chemically more meaningful indicators of hydrogen bond formation, have not been examined for these medium-resolution structures. We have fixed hydrogen atoms at ideal geometric locations for all the bases in 35 *B*-DNA-type oligomer crystal structures solved at better than 2.6 Å resolution and examined all possible H..O and H..N distances, as well as hydrogen bond angles (N-H..O/N), to check which cross-strand interactions can be truly classified as being hydrogen bonds, using a more stringent criterion than that used by earlier workers. We found that, as in the *B*-DNA fibre model structure (Chandrasekaran & Arnott, 1996), a majority of the N..O and N..N cross-strand distances in the major groove are slightly smaller than their equilibrium separation distances of 3.48 Å, as defined in the AMBER force-field (Cornell *et al.*, 1995), but the corresponding H..O and H..N distances are generally >2.8 Å, while the N-H..O/N angle is about 100° (Bansal & Ghosh, 1999). It was also seen that the average values of the cross-strand distances between the exocyclic groups at the 5′-ends of several base-paired dinucleotides, such as N6(A)..H62(A) in AT.AT (2.82(±0.18) Å), N6(A)..H42(C) in AG.CT (2.84(±0.26) Å), as well as O6(G)..H42(C) in GG.CC (2.92(±0.23) Å) steps in the oligomer crystal structures, are shorter than the average H62(A)..O4(T) distance (3.04 ± 0.23) Å in the well-documented AA.TT steps (for which it is 2.85 Å in the *B*-DNA fibre model poposed by Chandrasekaran & Arnott (1996)). This indicates that (i) an AT base-pair with large propeller twist is not a mandatory requirement for close interaction between the exocyclic groups in the major groove and (ii) a majority of the potential cross-strand hydrogen bonds in the major groove do not have very favourable geometries.

Cross-strand interactions in the DNA minor groove

N-H..O and N-H..N hydrogen bonds

An analysis of all possible cross-strand interactions in the minor groove of oligomeric DNA crystal structures revealed that close N..O and N..N contacts (<3.5 Å) between the nitrogen atom of the 2-amino group of guanine and oxygen of the 2-keto oxygen atom of either cytosine or thymine, located towards the 3′-side, are occasionally seen for GG.CC, AG.CT and CG.CG steps, as reported earlier for the AG.AT steps in the decamer containing two A.G mismatch base-pairs (Prive *et al.*, 1987). However, the H..O and H..N distances are >2.8 Å in most cases and the corresponding hydrogen bond angle is <110°, indicating that these N-H..O/N hydrogen bonds have poor geometry (Desiraju, 1996), similar to that mentioned above for the major groove.

C-H..O hydrogen bonds

The most surprising finding of the detailed analysis of cross-strand interactions was the occurrence of short C2..O2 as well as H2..O2 cross-strand contact distances, in the AA.TT and GA.TC steps, between the adenine base on the 3′-end of the purine strand and the thymine or cytosine base on the 3′-end in the pyrimidine strand. All the C-H..O close contacts in the oligomer crystal data set, wherein the cross-strand H..O distance is <2.8 Å are given in Table 1, along with the corresponding C2-H2..O2 angle. The parameters for the potential C-H..O hydrogen bond between the C2-H2 group of adenine, involved in a cross-strand interaction, and the O2 of its Watson-Crick paired thymine base, are also listed, and the values are nearly the same as for the cross-strand interaction. It is obvious that when a 2.8 Å cutoff is applied for the cross-strand H2..O2 distance, the corresponding C2..O2 distance is also <3.3 Å in most cases, for

Table 1. Relevant parameters for the cross-strand C2H2..O2 hydrogen bonds with H2..O2 distance < 2.8 Å, in the minor groove of (A) AA.TT and (B) GA.TC steps in DNA oligomer crystal structures

	Cross-strand hydrogen bonds				Watson-Crick hydrogen bonds			
PDB name	Base numbers[a]	H2..O2 (Å)	C2..O2 (Å)	C2-H2..O2 (deg.)	Base numbers	H2..O2 (Å)	C2..O2 (Å)	C2-H2..O2 (deg.)
A. *AA.TT steps*								
Fibre[b]		AA.TT						
	2A..2T	3.33	4.05	125.0	2A..1T	3.03	3.71	121.2
Oligomer[c]		AA.TT						
Mean		2.81 (0.45)	3.43 (0.46)	116.5 (7.2)		2.82 (0.22)	3.54 (0.21)	123.8 (7.0)
1D98		[CGCAAAAAAGCG]						
	5A..9T	2.66	3.42	126.8	5A..8T	2.97	3.67	122.3
	6A..8T	2.56	3.36	129.0	6A..7T	3.00	3.86	135.4
	7A..7T	2.19	2.82	114.5	7A..6T	3.10	3.63	110.2
	9A..5T	2.29	2.99	120.5	9A..4T	3.15	3.99	134.9
1BDN		[CGCAAAAATGCG]						
	5A..9T	2.63	3.19	111.0	5A..8T	3.18	3.68	109.2
	6A..8T	2.63	3.43	129.6	6A..7T	3.03	3.69	119.2
	7A..7T	2.47	2.95	104.9	7A..6T	2.59	3.20	115.0
	8A..6T	2.61	3.17	110.9	8A..5T	3.00	3.74	125.4
1D89		[CGCGAAAAAACG]						
	6A..8T	2.74	3.33	113.2	6A..7T	2.63	3.30	119.2
	7A..7T	2.68	3.20	108.8	7A..6T	2.62	3.35	124.0
	8A..6T	2.71	3.28	112.6	8A..5T	2.75	3.48	124.4
	9A..5T	2.26	2.93	117.5	9A..4T	2.99	3.59	115.5
	10A..4T	2.57	3.25	119.8	10A..3T	2.60	3.34	124.8
1BNA		[CGCGAATTCGCG]						
	6A..8T	2.66	3.29	116.5	6A..7T	2.78	3.49	122.8
	8T..6A	2.74	3.32	112.9	7T..6A	2.61	3.32	121.8
355D		[CGCGAATTCGCG]						
	6A..8T	2.74	3.31	112.2	6A..7T	2.65	3.43	127.6
	8T..6A	2.64	3.24	114.0	7T..6A	2.69	3.45	125.9
1D29		[CGTGAATTCACG]						
	6A..8T	2.52	3.22	121.2	6A..7T	3.41	4.02	116.9
	8T..6A	2.24	2.92	118.3	7T..6A	3.12	3.71	114.9
1D65		[CGCAAATTTGCG]						
	5A..9T	2.78	3.57	129.1	5A..8T	3.17	3.80	117.8
	9T..5A	2.60	3.31	121.8	8T..5A	3.11	3.67	112.7
1D77		[CGCIAATTCGCG)						
	6A..8T	2.42	3.09	118.6	6A..7T	3.19	3.88	122.4
	8T..6A	2.54	3.24	121.4	7T..6A	2.79	3.56	127.3
265D		[CG5MeCGAATT5MeCGCG]						
	6A..8T	2.76	3.37	115.3	6A..7T	2.58	3.39	130.8
	8T..6A	2.73	3.30	112.4	7T..6A	2.55	3.38	132.6
270D		[CGCGAATT5MeCGCG]						
	8T..6A	2.62	3.19	111.8	7T..6A	2.77	3.49	123.4
4BNA		[CGCGAATT5BrCGCG]						
	6A..8T	2.70	3.32	115.1	6A..7T	2.68	3.33	117.6
	8T..6A	2.63	3.29	118.5	7T..6A	2.82	3.61	129.2
4DNB		[CGCGA6MeATTCGCG]						
	8T..6A	2.54	3.05	107.3	7T..6A	3.09	3.72	117.7
307D		[CAAAGAAAAG]						
	7A..5T	2.37	3.00	115.1	7A..4T	2.91	3.62	122.8
	8A..4T	2.30	2.91	113.3	8A..3T	2.85	3.51	119.3
	9A..3T	2.64	3.26	115.3	9A..2T	2.59	3.34	125.9
167D		[CCATTAATGG]						
	5T..7A	2.52	3.04	108.4	4T..7A	2.99	3.75	127.0
1D49		[CGATTAATCG]						
	5T..7A	2.66	3.22	111.6	4T..7A	2.54	3.32	128.1
252D		[CGCAATTGCG]						
	5A..7T	2.68	3.29	114.7	5A..6T	3.12	3.77	119.0
271D[d]		[CGCGAAUUCGCG]						
	6A..8U	2.76	3.33	112.3	6A..7U	2.63	3.45	131.0
	8U..6A	2.26	2.99	122.2	7U..6A	2.86	3.47	115.8
B. *GA.TC steps*								
Oligomer[c]		GA.TC						
Mean		2.95 (0.34)	3.54 (0.32)	114.2 (7.2)		2.84 (0.18)	3.56 (0.18)	123.9 (7.1)
1BNA		[CGCGAATTCGCG]						
	9C..5A	2.50	3.17	118.3	8T..5A	2.92	3.52	115.2
355D		[CGCGAATTCGCG]						
	9C..5A	2.75	3.35	114.7	8T..5A	2.69	3.46	127.2
1D29		[CGTGAATTCACG]						
	9C..5A	2.26	2.87	113.5	8T..5A	2.52	3.18	117.6
265D		[CG5MeCGAATT5MeCGCG]						
	5A..9C	2.69	3.17	106.5	5A..8T	2.55	3.32	126.7

Table 1. *(continued)*

270D		[CGCGAATT5MeCGCG]						
	5A..9C	2.80	3.34	110.6	5A..8T	2.85	3.68	132.5
4BNA		[CGCGAATT5BrCGCG]						
	5A..9C	2.53	3.35	131.5	5A..8T	3.13	3.87	125.9
	9C..5A	2.35	3.21	134.5	8T..5A	2.86	3.44	113.4
4DNB		[CGCGA^{6Me}ATTCGCG]						
	9C..5A	2.46	2.98	107.3	8T..5A	2.57	3.04	104.4
1D23		[CGATCGATCG]						
	5C..7A	2.64	3.21	112.0	4T..7A	2.71	3.46	125.0
	9C..3A	2.79	3.41	115.8	8T..3A	2.74	3.50	127.1
1D56		[CGATATATCG]						
	3A..9C	2.75	3.40	117.9	3A..8T	3.06	3.80	125.5
1D57		[CGATATATCG]						
	3A..9C	2.61	3.21	114.0	3A..8T	3.18	3.76	114.3
	9C..3A	2.51	3.27	126.2	8T..3A	2.96	3.66	122.5
271D[d]		[CGCGAAUUCGCG]						
	9C..5A	2.80	3.40	114.4	8U..5A	2.87	3.59	123.3
1D77[e]		[CGCIAATTCGCG]						
	9C..5A	2.55	3.31	125.7	8T..5A	2.82	3.50	120.2

The parameters for the C2H2..O2 hydrogen bond, involving the same C2-H2 group of adenine, with the corresponding Watson-Crick base-paired thymine base, are also listed.

All H2 atoms have been fixed in the plane of the adenine bases, using standard bond length and bond angle b(C-H) = 1.08 Å, ∠N1-C2-H2 = 120°. The dataset consists of 17 dodecamers and 18 decamer structures, solved using data at resolution <2.6 Å. (PDB code names: 1BNA, 355D, 1D98, 1DN9, 1BDN, 1D29, 1D65, 119D, 1D89, 249D, 271D, 194D, 4BNA, 4DNB, 1D77, 265D, 270D; 1BD1, 5DNB, 1D23, 1D49, 1D56, 1D57, 1CGC, 126D, 158D, 167D, 196D, 252D, 307D, 2D25, 1D60, 1D61, 1DA3 and 183D).

[a] The bases have been numbered from 1 to n in each strand, starting at the 5′ end.

[b] Fibre model structure for *B*-DNA (Chandrasekaran & Arnott, 1996).

[c] Mean values for all the AA.TT (53) and GA.TC (40) dinucleotide steps in oligomer crystal structure data set, along with the standard deviations. The mean geometrical parameters for AA.TT steps are indicated in the legend to Figure 1.

[d] This structure contains AA.UU steps in lieu of AA.TT and GA.UC steps in lieu of GA.TC.

[e] This structure contains IA.TC steps in lieu of GA.TC steps.

cross-strand as well as Watson-Crick atom pairs, an indication of three-centre C-H..O hydrogen bonds being formed, since both H..O and C..O distances are considerably smaller than the equilibrium separation distance values, as assigned in the AMBER force-field (Cornell *et al.*, 1995). Both sets of C2-H2..O2 angles have similar values of about 120° (Table 1), while the H2..O2-C2 angle also lies between 110° and 130° for both interactions. However, the azimuthal angle defining the elevation of H2..O2 direction with respect to the plane of the thymine base is obviously different in the two cases, being very small (<20°) for the intra-base-pair C-H..O hydrogen bond, and about 60° for the cross-strand hydrogen bond. As a representative example, the middle six base-pairs in the d(CGCGAATTCGCG) dodecamer structure, solved recently at 1.4 Å resolution (Shui *et al.*, 1998a) are shown in Figure 1, with H2..O2 cross-strand distances in the minor groove of AA.TT and GA.TC steps, as well as the H2..O2 distances in the A.T Watson-Crick base-pairs being indicated. All the H2..O2 distances are between 2.6 and 2.9 Å. The C2-H2..O2 interaction in Watson-Crick base-pairs can be regarded as being a ''passive'' consequence of the geometry of the base-pair (Wahl & Sundaralingam, 1997) while the cross-strand interaction can be regarded as being ''voluntary'', since the observed H2..O2 distances are considerably smaller than the H2..O2 separation of 3.3 Å in the latest *B*-DNA fibre model, which has a high propeller twist of −15° (Chandrasekaran & Arnott, 1996). It can be seen from the legends to Figure 1 and Table 1 that the A.T base-pairs in the crystal structures apparently undergo distortion (in their opening angle) so as to facilitate the formation of good C-H..O hydrogen bonds within the AT Watson-Crick pair itself, while deviations in the inter-base-pair parameters (particularly ''slide'') in AA.TT steps lead to an additional cross-strand C-H..O hydrogen bond with a neighbouring pyrimidine on the 3′-side. It is interesting to note that this cross-strand interaction occurs without any roll movement towards the minor groove, i.e. negative values for the roll parameter, as defined by the Cambridge nomenclature (Dickerson *et al.*, 1989). The hydrogen bond geometry of these C-H..O interactions (as described by the H..O distances and C-H..O angles) is better than that of most cross-strand N-H..O and N-H..N hydrogen bonds.

An even clearer picture emerges if we examine the complete range of cross-strand C-H..O distances in all the AA.TT and GA.TC steps in the oligomer crystal data set, as well as a data set consisting of 20 crystal structures of DNA oligomers complexed with regulatory proteins (resolution <2.6 Å), wherein the DNA does not undergo large distortion. The overall distribution of the C..O and H..O distances is shown in Figure 2(a)-(d) (as fractional frequency of occurrence) and it is seen that the cross-strand H..O distance has a maximum between 2.6 and 2.9 Å, while the maxi-

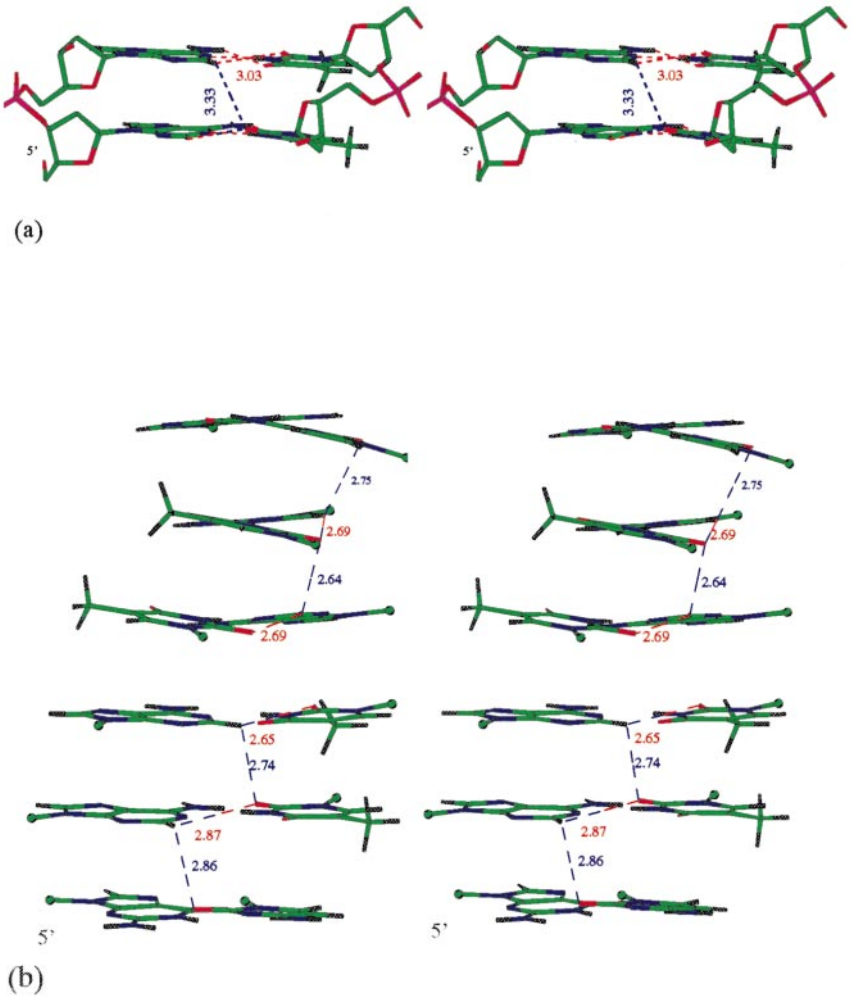

Figure 1. (a) Stereodiagram showing a view into the minor groove of an AA.TT dinucleotide step with the *B*-DNA fibre model geometry (Chandrasekaran & Arnott, 1996). The H2..O2 distance in the A.T Watson-Crick base-pair, as well as between the 3′-end adenine base in strand 1 and the 3′-end thymine base in strand 2 are indicated. The intra-base-pair propeller twist is −15° and opening angle is 0.3° in this structure while the inter-base-pair parameters roll and slide are 2.2° and 0.56 Å, respectively, as calculated using a modified NUPARM package (Bansal *et al.*, 1995). (b) Stereodiagram showing the middle six base-pair GAATTC.GAATTC in a high-resolution (1.4 Å) dodecamer structure (PDB code 355D, Shui *et al.*, 1998a,b). The cross-strand H2..O2 distances in the minor groove, for AA.TT and GA.TC are shown in blue, while the H2..O2 distances for the Watson-Crick A.T base-pairs are shown in red. All the H2..O2 distances in the AA.TT steps are <2.87 Å in this structure. The average propeller twist and opening angle for the A.T base-pairs in this structure are −16° and 5.6°, while the roll angle is −0.2° and slide is −0.46 Å for the AA.TT steps. The mean propeller and opening angle for all the 53 AA.TT steps in the oligomer data set included in Table 1 are −16.2(±4.5)° and 3.8(±3.5)° while the roll angle and slide values are 0.9(±3.9)° and −0.18(±3.4) Å, respectively. The stereodiagrams have been generated using MOLSCRIPT (Kraulis, 1991).

mum occurs between 3.1 and 3.5 Å for the C..O distance, for AA.TT and GA.TC steps in DNA oligomers as well as protein-complexed structures. These distances in the *B*-DNA fibre model are 3.3 and 4.0 Å, respectively, confirming that the C-H..O interactions in the minor groove are highly favoured at AA.TT and GA.TC steps and probably contribute towards bringing the two strands closer together. In order to compare the relative importance of these minor groove interactions with the more commonly discussed major groove hydrogen bond interactions in the AA.TT steps, we compared the relative frequency of N6-H62..O4 and C2-H2..O2 hydrogen bonds, as judged by identical distance criteria, i.e. percentage of N6..O4, C2..O2 or H62..O4, H2..O2 distances, which are shorter than any chosen cut-off distance. Plots of this type are shown in Figure 3 and it is clear that the choice of cut-off distance does not affect the general conclusion that, in AA.TT steps very large numbers of H2..O2 distances are shorter than the H62..O4 distances. Thus, only about 13% of H62..O4 distances occur within a cut-off value of 2.8 Å (Figure 3(a)), the value for this cross-strand distance in the fibre model, while 66% of the H2..O2 distances in AA.TT steps are shorter than this value (Figure 3(b)). Even a cut-off distance of 3.0 Å gives only 45% frequency of occurrence for the H62..O4 distance, while it is 81% for the H2..O2 distance in DNA oligomers for AA.TT steps and 60% for GA.TC steps (as seen in Figure 3(b) and (c), respectively) A similar trend is observed for the protein-DNA complex data set, though both steps show a tendency towards slightly longer distances (also shown in Figure 3(a)-(c)). The frequencies of occurrence of N6..O4 and C2..O2 distances in the AA.TT steps, within any cut-off limit, are not significantly different (Figure 3(d) and (e)), with the frequency of C2..O2 being higher by about 10% in the whole range (3.0-3.8 Å). However, it should be noted that these trends are quite striking, since they are opposite to that expected from the fibre

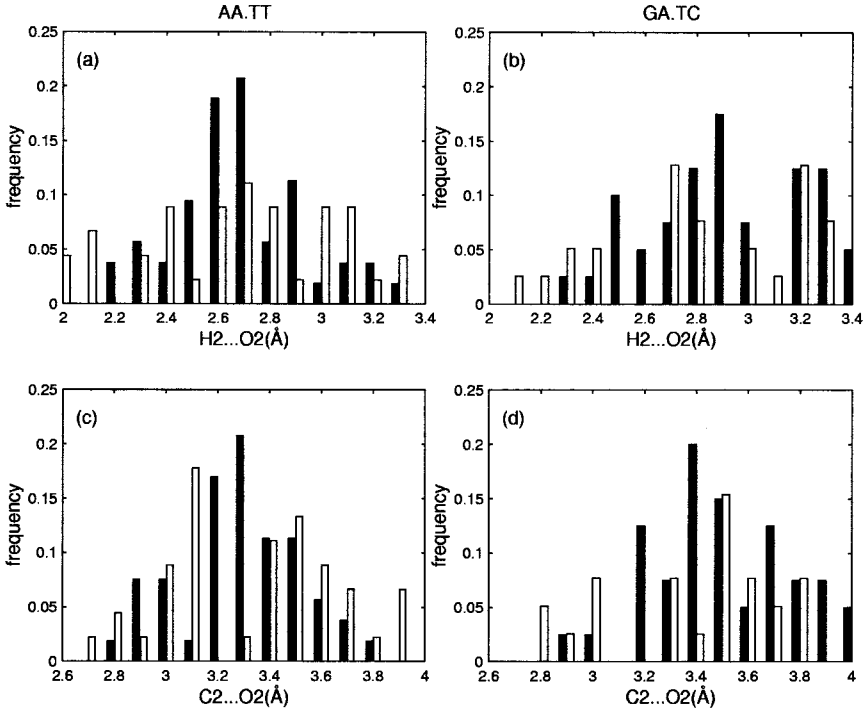

Figure 2. Histograms showing the frequency of occurrence of H2..O2 (top) and C2..O2 (bottom) cross-strand distances in the range 2.0-3.4 Å and 2.6-4.0 Å, respectively, in the minor grooves of (a) and (c) AA.TT, (b) and (d) GA.TC steps. The data for *B*-DNA type oligomer crystal structures are shown as filled bars and the corresponding values for regulatory protein bound DNA structures are represented by open bars. Frequency of occurrence has been normalized, so that data sets of different sizes can be compared. There are 53 AA.TT steps in the oligomer data set and 45 in the DNA-protein data set, while there are 40 GA.TC steps in the oligomers and 39 in the complex data set. The oligomer data set includes the 35 structures listed in the footnote to Table 1. The protein-DNA complex data set includes 20 structures (containing 24 molecules in the asymmetric units) solved with a data resolution of <2.6 Å (PDB codes; 1AAY, 2BOP, 1LMB, 1HCR, 1TRO, 1LAT, 1ZAA, 1LLI, 1MEY, 1TUP, 2DGC, 1TSR, 1TRR, 1HCQ, 1PDN, 1PER, 1RPE, 1UBD, 2OR1 and 3CRO).

model structure, wherein the N6..O4 distance is 3.1 Å, while the C2..O2 distance is 4.0 Å. Thus, a considerable reduction in C2..O2 and H2..O2 cross-strand distances is consistently observed for both AA.TT and GA.TC steps in the two data sets (Figure 3(e) and (f)), the oligomer set consisting of decamer and dodecamer sequences, as well as the DNA-protein complex structures, with DNA of varying length and sequence, crystallized in different lattices. On the contrary, the cross-strand H62..O4 distance in the crystal structures is longer than that in the fibre model. While this distance can be marginally reduced if the amino hydrogen atoms of adenine are rotated out of the base plane, it does not significantly alter the above observations. A similar analysis for the intra-base-pair H2..O2 distance reveals that in 51% of the A.T Watson-Crick base-pairs, this distance is <2.8 Å, while in 80% of base-pairs the distance is <3.0 Å, the value for the fibre model structure.

Cross-strand C2-H2..O2 hydrogen bonds are found also in AA.UU and GA.UC steps, present in a Drew-Dickerson type dodecamer sequence (as seen in Table 1, PDB code 271D), inosine containing steps IA.TC in a similar dodecamer (Table 1B, PDB code 1D77; Xuan & Weber, 1992), as well as IIICCC.IIICCC tracts of a decamer crystal structure (PDB code 286D; Shatzky-Schwartz *et al.*, 1997), suggesting that such cross-strand hydrogen bonds may be a general feature of all dinucleotide steps containing a base with a potential C-H donor group and another with a keto oxygen atom as an acceptor, in the minor groove of DNA double helices.

C-H..N hydrogen bonds

A few of the CA.TG steps also show close cross-strand contacts between the C2-H2 of adenine and the 2-amino nitrogen atom of guanine, with H2..N2 distances <2.8 Å and C2-H2..N2 angles ~110°, suggesting the presence of C-H..N hydrogen bonds. These short distances correspond to CA.TG steps with large twist, negative roll and positive slide, which is associated with a B_{II} geometry of the backbone (Bhattacharyya & Bansal, 1990; Nagaich *et al.*, 1994). These geometries are generally found in structures with CAA fragments, with the neighbouring AA.TT steps taking up an unusual geometry with large positive roll, small

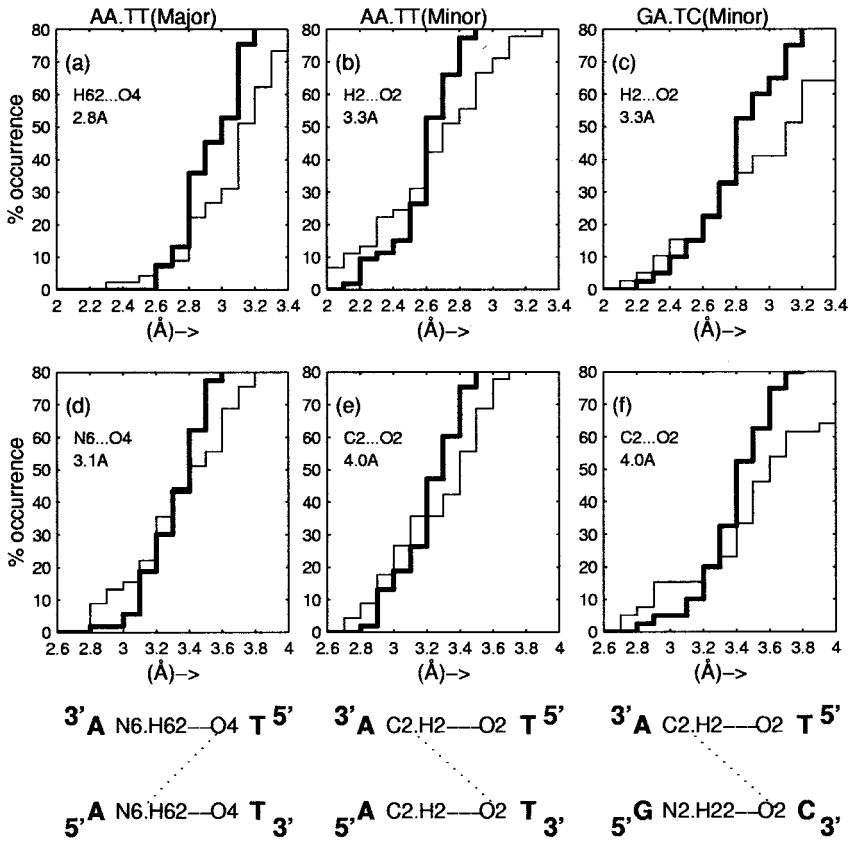

Figure 3. Ladder plots showing the percentage of doublet steps that form cross-strand hydrogen bonds in the AA.TT and GA.TC steps, which are shorter than a particular cut-off distance (indicated along the *x*-axis). In each part of the Figure, the thick line corresponds to the oligomer data and the thin line to the protein-complexed DNA data set. (a) H62...O4 and (d) N6...O4 for the N6H62(Ade)...O4(Thy) cross-strand hydrogen bond in the major groove of AA.TT steps, as shown at the bottom. (b) and (c) H2..O2 distances, (e) and (f) C2..O2 distances for cross-strand C2H2(Ade)...O2(Thy/Cyt) hydrogen bonds in the minor groove of AA.TT and GA.TC steps, respectively, as shown at the bottom. For example, 66 % of AA.TT steps form C2-H2..O2 cross-strand hydrogen bonds with H2(3′-Ade)..O2(3′-Thy) distance <2.8 Å, while only about 13 % form N-H..O hydrogen bonds in the major groove, for the same cut-off distance for H62(5′-Ade)..O4(5′-Thy). The distances corresponding to the fibre model structure are given in each part of the Figure, to highlight the considerable reduction in H2..O2 and C2..O2 minor-groove distances in AA.TT and GA.TC steps in the crystal structures, as compared to the fibre model, whereas the trend is exactly opposite in the case of major-groove distances in AA.TT steps.

twist and low positive slide (for example in structures with PDB codes 5DNB, 158D, 1D61, 1D65, 307D) and these are the only AA.TT steps characterized by large cross-strand C2..O2 distances (>3.6 Å).

Discussion

The mean separation between the C2H2 group of adenine and O2 of thymine in the Watson-Crick A.T base-pairs in oligomer crystal structures is found to be shorter than that in the fibre model of *B*-DNA. As seen in Table 1, as well as Figure 1, the mean H2..O2 distance in oligomers is 2.8 Å, while it is 3.0 Å for the fibre model, even though the base-pair propeller twist is nearly the same in both cases. This reduction can be attributed to a difference of 3° in the intra-base-pair opening angle for AT base-pairs, which brings the exocyclic atoms in the minor groove closer together, while the distance between N6 and O4 in the major groove increases to 3.1 Å from 2.95 Å in the fibre model. The reduction in the cross-strand C2H2..O2 distance between the 3′-end adenine base on one strand and the 3′-end thymine or cytosine base on the opposite strand in AA.TT and GA.TC steps is even more striking. In 90 % of the AA.TT steps, the H2..O2 distance is <3.3 Å, the value in the fibre model, while the mean value for the crystal structures is 2.8 Å, a reduction of 0.5 Å. This arises primarily due to a change in the inter-base-pair slide parameter, from 0.56 Å in the fibre model to −0.18 Å for the AA.TT steps in the oligomer crystal structures, while the average of all the dinucleotide sequences is 0.33 Å. It is interesting to note that the propeller twist and the roll angle show only minor differences (~1°) from the fibre model. In the case of GA.TC steps, the mean propeller twist has a value of −13° and the slide is 0.0 Å, but a twist value of 38° leads to the mean cross-strand H2..O2 distance being 2.95 Å.

It is also interesting to note that, as seen from Figure 3, the C2..O2 and H2..O2 cross-strand distances in the minor groove are, in general, shorter than the cross-strand N6..O4 and H62..O4 distances in the major groove of AA..TT steps. In fact, the mean value of both these cross-strand distances in the major groove are considerably larger than in the fibre model structure. Thus, it is clear that there are correlated changes in the local geometry of AA.TT and GA.TC steps in the crystal structures, leading to short H2..O2 distances on the minor groove side, within the A.T base-pairs as

well as between the bases at the 3′-ends of these dinucleotide steps. A significantly large number of these dinucleotide steps have H2..O2 as well as C2..O2 distances that are shorter than their equilibrium separation distances of 2.86 Å and 3.57 Å, respectively (Cornell *et al*., 1995), while the C-H..O angles are about 120°. This suggests that the C2-H2..O2 interactions are not merely electrostatic in nature but can be regarded as being weak hydrogen bonds. In addition, the base-pair energy calculated for an A.T base-pair, using the AMBER force-field, is about 1.2 kcal/mol less than the value obtained using *ab initio* calculations, by the same group, while they are within 0.6 kcal/mol for the G.C base-pair (Gould & Kollman, 1994). If one includes the C-H..O hydrogen bond energy, which is generally considered to be in the range between 1 and 2 kcal/mol (Desiraju, 1991), then the agreement for the A.T base-pair energy is substantially improved. It should be mentioned, however, that the standard definition of a hydrogen bond (Steiner & Saenger, 1993) requires C2 of adenine to be more electronegative than H2, while in the currently used force-fields, which have been developed taking into consideration only the strong hydrogen bond interactions, C2 is more electropositive than H2 (Cornell *et al*., 1995). Since recent molecular dynamics studies on RNA indicate that some C-H..O interactions are dynamically stable (Auffinger & Westhof, 1998), it is suggested that the charge assignment for the C-H groups probably needs refinement, taking polarization effects into consideration (Wiberg *et al*., 1991).

The $A_n.T_n$ and $GA_n.T_nC$-containing regions of DNA oligomers show some features, such as a narrow minor groove, that are independent of the crystal lattice packing and the overall bending characteristics of the molecule (Drew *et al*., 1981; Nelson *et al*., 1987; Di Gabriele & Steitz, 1993; Dickerson *et al*., 1994; Han *et al*., 1997) as well as binding of α-helices in the major groove of DNA molecules, in the protein-DNA complexes (Aggarwal *et al*., 1988; Mondragon & Harrison, 1991; Rodgers & Harrison, 1993; Feng *et al*., 1994; Gewirth, & Sigler, 1995; Houbaviy *et al*., 1996). The Hin recombinase-bound DNA structure (Feng *et al*., 1994) is the only exception, it has a normal average minor groove width in spite of the presence of a $T_5.A_5$ tract. Hence, in addition to the favourable cross-strand interactions in the major groove and the effect of tightly bound solvent molecules in the minor groove, cross-strand C-H..O interactions could help stabilize the oligo(A) tract structure with its characteristic features, such as a narrow minor groove.

Acknowledgements

This work was supported, in part, by financial assistance from the Department of Biotechnology, India.

References

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science,* **242**, 899-907.

Auffinger, P. & Westhof, E. (1998). Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* **8**, 227-236.

Bansal, M. & Ghosh, A. (1999). Cross-strand C-H..O and N-H..N Interactions in *B*-DNA duplex structure: hydrogen bonds or short contacts?. In *Perspectives in Structural Biology* (Vijayan, M. & Yathindra, N., eds), pp. 353-364, University Press, Hyderabad.

Bansal, M., Bhattachayya, D. & Ravi, B. (1995). NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *Comput. Appl. Biosci.* **11**, 281-287.

Bella, J. & Berman, H. M. (1996). Crystallographic evidence for C-H...O = C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **264**, 734-742.

Bhattacharyya, D. & Bansal, M. (1990). Local variability and base sequence effects in DNA crystal structures. *J. Biomol. Struct. Dynam.* **8**, 539-571.

Bhattacharyya, D. & Bansal, M. (1992). Groove width and depth of *B*-DNA structures depend on local variation in slide. *J. Biomol. Struct. Dynam.* **10**, 213-226.

Chakrabarti, P. & Chakrabarti, S. (1998). C-H...O hydrogen bond involving proline residues. *J. Mol. Biol.* **284**, 867-873.

Chandrasekaran, R. & Arnott, S. (1996). The structure of *B*-DNA in oriented fibres. *J. Biomol. Struct. Dynam.* **13**, 1015-1027.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). The occurrence of C-H...O hydrogen bonds in proteins. *J. Mol. Biol.* **252**, 248-262.

Desiraju, G. R. (1991). The C-H...O hydrogen bond in crystals: what is it? *Acc. Chem. Res.* **24**, 290-296.

Desiraju, G. R. (1996). The C-H...O hydrogen bond: structural implications and supramolecular design. *Acc. Chem. Res.* **29**, 441-449.

Dickerson, R. E. (1998). Sequence-dependent *B*-DNA conformation in crystals and in protein complexes. In *Structure, Motion, Interaction & Expression of Biological Macromolecules* (Sarma, R. H. & Sarma, M. H., eds), pp. 17-36, Adenine Press, New York.

Dickerson, R. E. & Drew, H. R. (1981). Structure of a *B*-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol.* **149**, 761-786.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Skelnar, H., Soumpasis, D. M., Tung, C.-S., von Kitzing, E., Wang, A. H.-J. & Zhurkin, V. B. (1989). Definitions and nomenclatures of nucleic acid structure parameteres. *J. Mol. Biol.* **205**, 787-791.

Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994). ''... the tyranny of the lattice...''. *Proc. Natl Acad. Sci. USA,* **91**, 3579-3583.

DiGabriele, A. D. & Steitz, T. A. (1993). A DNA dodecamer containing an adenine tract crystallizes in a

unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024-1039.

DiGabriele, A. D., Sanderson, M. R. & Steitz, T. A. (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., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1981). Structure of a *B*-DNA dodecamer: conformation and dynamics. *Proc. Natl Acad. Sci. USA,* **78**, 2179-2183.

Edwards, K. J., Brown, D. G., Spink, N., Skelly, J. V. & Neidle, S. (1992). Molecular structure of the *B*-DNA dodecamer d(C-G-C-A-A-A-T-T-T-C-G-C)$_2$: an examination of propeller twist and minor-groove water structure at 2.2 Å. *J. Mol. Biol.* **226**, 1161-1173.

El Hassan, M. A. & Calladine, C. R. (1996). Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.* **259**, 95-103.

El Hassan, M. A. & Calladine, C. R. (1997). Conformational characteristics of DNA: emperical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Phil. Trans. Roy. Soc. ser. A,* **355**, 43-100.

Feng, J.-A., Johnson, R. C. & Dickerson, R. E. (1994). Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science,* **263**, 348-355.

Gewirth, D. T. & Sigler, P. B. (1995). The basis for half-site specificity explored through a non-cognate steroid receptor-DNA complex. *Nature Struct. Biol.* **2**, 386-394.

Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995). *B*-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**, 34-48.

Gould, I. R. & Kollman, P. A. (1994). Theoretical investigation of the hydrogen bond strengths in guanine-cytosine and adenine-thymine base-pairs. *J. Am. Chem. Soc.* **116**, 2493-2499.

Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K. & Dickerson, R. E. (1997). Structure of a DNA analog of the primer for HIV-1 RT second strand synthesis. *J. Mol. Biol.* **269**, 811-826.

Houbaviy, H. B., Usheva, A., Shenk, T. & Burley, S. K. (1996). Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl Acad. Sci. USA,* **93**, 13577-13582.

Hunter, C. A. (1993). Sequence-dependent DNA structure. The role of base stacking interactions. *J. Mol. Biol.* **230**, 1025-1054.

Hunter, C. A. & Lu, X.-J. (1997). DNA base-stacking interactions: a comparison of theoretical calculations with oligonucleotide X-ray crystal structures. *J. Mol. Biol.* **265**, 603-619.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Leonard, G. A., McAuley-Hecht, K., Brown, T. & Hunter, W. N. (1995). Do C-H...O hydrogen bonds contribute to the stability of nucleic acid base pairs? *Acta Crystallog. sect. D,* **51**, 136-139.

Luisi, B., Orozco, M., Sponer, J., Luque, F. J. & Shakked, Z. (1998). On the potential role of the amino nitrogen atom as a hydrogen bond acceptor in macromolecules. *J. Mol. Biol.* **279**, 1123-1136.

Mandel-Gutfreund, Y., Margalit, H., Jernigan, R. L. & Zhurkin, V. B. (1998). A role for CH...O interactions in protein-DNA recognition. *J. Mol. Biol.* **277**, 1129-1140.

Mohanty, D. & Bansal, M. (1991). DNA polymorphism and local variation in base-pair orientation: a theoretical rationale. *J. Biomol. Struct. Dynam.* **9**, 127-142.

Mondragon, A. & Harrison, S. C. (1991). The phage 434 Cro/O_R 1 complex at 2.5 Å resolution. *J. Mol. Biol.* **219**, 321-334.

Nagaich, A. K., Bhattacharyya, D., Brahmachari, S. K. & Bansal, M. (1994). CA/TG sequence at the 5′ end of oligo(A)-tracts strongly modulates DNA curvature. *J. Biol. Chem.* **269**, 7824-7833.

Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). The structure of an oligo(dA)-oligo(dT) tract and its biological implications. *Nature,* **330**, 221-226.

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

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

Rodgers, D. W. & Harrison, S. C. (1993). The complex between phage 434 repressor DNA-binding domain and operator site O_R3: structural differences between consensus and non-consensus half sites. *Structure,* **1**, 227-240.

Shatzky-Schwartz, M., Arbuckle, N. D., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T. E., Luisi, B. F. & Shakked, Z. (1997). X-ray and solution studies of DNA oligomers and implication for the structural basis of A-tract-dependent curvature. *J. Mol. Biol.* **267**, 595-623.

Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998a). The *B*-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry,* **37**, 8341-8355.

Shui, X., Sines, C. S., McFail-Isom, L., VanDerveer, D. & Williams, L. D. (1998b). Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry,* **37**, 16877-16887.

Sponer, J. & Kypr, J. (1994). Close mutual contacts of the amino groups in DNA. *Int. J. Biol. Macromol.* **16**, 3-6.

Steiner, T. & Saenger, W. (1993). Role of C-H...O hydrogen bonds in the coordination of water molecules. Analysis of neutron diffraction data. *J. Am. Chem. Soc.* **115**, 4540-4547.

Suzuki, M., Amano, N., Kakinuma, J. & Tateno, M. (1997). Use of a 3D structure data base for understanding sequence-dependent conformational aspects of DNA. *J. Mol. Biol.* **274**, 421-435.

Taylor, R. & Kennard, O. (1982). Crystallographic evidence for the existence of C-H...O, C-H...N, and C-H...Cl hydrogen bonds. *J. Am. Chem. Soc.* **104**, 5063-5070.

Wahl, M. C. & Sundaralingam, M. (1997). C-H...O hydrogen bonding in biology. *Trends Biochem. Sci.* **22**, 97-102.

Wiberg, K. B., Waldron, R. F., Schulte, G. & Saunders, M. (1991). Lactones. ! X-ray crystallographic studies of nonanolactone and tridecanolactone: nature of CH..O nonbonded interactions. *J. Am. Chem. Soc.* **113**, 971-977.

Xuan, J.-C. & Weber, I. T. (1992). Crystal structure of a *B*-DNA dodecamer containing inosine, d(CGCIAATTCGCG), at 2.4 Å resolution and its comparison with other *B*-DNA dodecamers. *Nucl. Acids Res.* **20**, 5457-5464.

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

Edited by B. Honig

(Received 24 March 1999; received in revised form 25 October 1999; accepted 25 October 1999)