Crystal and solution structures of the oligonucleotide d(ATGCGCAT)$_2$: a combined X-ray and NMR study

G.R.Clark+, D.G.Brown, M.R.Sanderson, T.Chwalinski, S.Neidle*, J.M.Veal[1], R.L.Jones[1], W.D.Wilson[1]*, G.Zon[2], E.Garman[3] and D.I.Stuart[3]
Cancer Research Campaign Biomolecular Structure Unit, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK, [1]Department of Chemistry, Georgia State University, Atlanta, GA 30303, [2]Applied Biosystems Inc., Foster City, CA 94404 USA and [3]Laboratory of Molecular Biophysics, University of Oxford, Oxford OX1 5QQ, UK

Received May 17, 1990; Revised and Accepted August 16, 1990

ABSTRACT

A combined crystal-structure determination and NMR analysis of the octanucleotide d(ATGCGCAT)$_2$ is reported. The X-ray analysis shows that the structure is A-form duplex in crystal state. The NMR study shows that in solution this sequence is B-type. The conformational results from each technique are presented in detail. The implications of these findings in terms of conformational flexibility and ligand binding are discussed.

INTRODUCTION

The DNA double helix is highly polymorphic. The major structural types of A, B and Z have all been observed in fibres of polynucleotides and in single crystals of oligonucleotides (1,2). The latter studies have also shown that polymorphism extends to the individual base level with sequence-dependent microstructure being a dominant property of DNA (3), which is related to the sequence-specific recognition of DNA by regulatory proteins (4,5). There is considerable current controversy as to the extent by which the intrinsic structural features of a sequence are responsible for specific recognition, rather than hydrogen-bonding from the bases. It is therefore important to determine the patterns of sequence-dependent structural features in a wide range of sequences and crystallographic situations.

The A-form has until recently received less attention as a major structural motif for recognition, even though its importance has long been acknowledged as the structural type adopted by DNA-RNA hybrids. There is now increasing evidence that it is important for the binding site of the TFIIIA transcription factor (6). The TFIIIA binding site, although B-form in solution (38), crystallises in the A-form (7), and this observation suggests that conversion of this sequence to the A-form may be important for recognition by TFIIIA. Some 15 different sequences have now been crystallised in the A-form, with crystal packing effects playing a role in determining some conformational detail (8,9). Almost all of these are of octanucleotides. The extent to which packing forces can actually force a particular sequence to adopt the A-form rather than any other is still an unresolved question, although it may well be that the tendency of octanucleotides to crystallise in just one of two space groups (tetragonal or hexagonal) is due to the need to adopt the A-form for efficient packing in the crystal. On the other hand, it is clear that the dodecanucleotide sequence d(CGCGAATTCGCG)$_2$ is in the B genus both in the crystal and in solution (10,11).

This study analyses the conformation of the alternating sequence d(ATGCGCAT)$_2$ with both Xray and NMR methods and compares the conformation found in the crystal with that determined in solution. We also report crystal-structure analyses of two crystal forms with and without bound spermine. This sequence is unusual in crystallising both in the presence and absence of this counter-ion, with both crystal forms diffracting to high resolution.

EXPERIMENTAL

Crystallography

The oligonucleotide was crystallised both in the presence and in the absence of spermine, using a hanging-drop technique and MPD as precipitant. Both crystallise in the space group $P4_32_12$, with crystals having bipyramidal habit. Intensity data were collected on a Xentronics area detector using a rotating-anode source with crystals kept in sealed quartz capillaries.

(1) With spermine. Cell dimensions are a,b = 42.526 and c = 24.920Å. Data were collected to 1.7Å resolution.

(2) Without spermine. Cell dimensions are a,b = 42.406 and c = 24.902Å. Data were collected to 1.5Å resolution.

Both structures were solved by molecular replacement methods, using the coordinates of the octamer sequence d(GGCCGGCC) (12) as a starting point. In both instances the asymmetric unit was a single strand of octamer. These were altered to the sequence d(ATGCGCAT) by means of the interactive graphics program GEMINI (13). Initial rigid-body and group refinements were with

* To whom correspondence should be addressed

+ Permanent address: Department of Chemistry, University of Auckland, Auckland, New Zealand

CORELS (14) followed by NUCLSQ (15). Both structures were assigned half-occupancies. Periodic checks were made to ensure the equivalence of the two strands, and symmetry-averaging was performed several times. At the conclusions of the refinements, no qualitative differences between strands were apparent and backbone torsion angles in the two strands of each structure differed by no more than 7°. Difference and 'omit' electron-density maps calculated with the PROTEIN package (16) were used to locate water and spermine molecules as well as to correct improperly-positioned portions of the oligonucleotides themselves. The spermine molecule was unequivocally visible in these maps at an electron level of 4σ with individual atoms being well resolved in the density. The maps were displayed on Silicon Graphics IRIS workstations using the TOM (17) version of FRODO. During the course of the refinement it was found that the largest observed structure amplitudes were systematically less than their calculated structure factors, and required local rescaling. This effect is due to the abnormally high intensities of these few reflections, which were beyond the linear response range of the area detector.

The current R factors for the two structures are: 17.7% for the spermine-free one, with 110 water molecules included, and 18.0% for the spermine-containing one, with 113 water molecules. The spermine molecules refined satisfactorily, with no evidence of disorder. Temperature factors for the oligonucleotide atoms range from 9.6 to 27.1Å^2 for the water molecules. Coordinate data will be deposited in the Brookhaven Database. Helical parameters have been calculated with the NEWHELIX program, using the Cambridge conventions (18). Details of the extensive water network will be reported elsewhere.

NMR Studies

The oligonucleotide was synthesized and the NMR sample prepared in D_2O as previously described (19). The strand concentration was 4 mM in phosphate buffer (0.01 M phosphate, 0.10 M NaCl, 0.1 mM EDTA, pH 7.0).

Pure absorption phase sensitive (20) correlated spectroscopy (COSY) and nuclear Overhauser spectroscopy (NOESY) datasets were acquired on a GE GN-600 spectrometer at 25°C. An acquisition time of 372 ms and a repetition delay time between scans of 1.6 s were employed for both experiments. For the COSY spectra 2048 complex points in t_2 and 512 points in t_1 were collected with 32 scans acquired per t_1 experiment. For the NOESY spectra 2048 complex points in t_2 and 256 points in t_1 were collected with 160 scans acquired per t_1 experiment. A mixing time of 100 ms was used. Additional NOESY experiments with a mixing time of 250 ms were also collected under similar conditions on a Varian VXR 400 at 15°C.

Following acquisition, both NOESY and COSY datasets were transferred to a Silicon Graphics Inc. IRIS workstation and processed using Hare Research FTNMR software. For the COSY experiment a sine bell function shifted 10° was used to apodize and improve the resolution of the data in both dimensions. The data was zero-filled to 4K in the t_2 dimension and 2K in the t_1 dimension prior to Fourier transformation. For the NOESY experiment, a line broadening of 2 Hz was applied in the t_2 dimension, and in the t_1 dimension the data was apodized with a sine-squared function phase shifted 90° and zero-filled to 2K points. For both COSY and NOESY data sets, the first t_1 experiment was multiplied by 0.5 prior to Fourier transformation (21). The spectra were referenced relative to TMS. Measurements of coupling constants (COSY spectra) and evaluation of cross peak intensities (NOESY spectra) were made on non-symmetrized spectra.

RESULTS

Crystallography

The two crystal structures have very similar conformational features for the octanucleotide, and therefore only one, the spermine-free, will be discussed in detail here.

The structure is an A-form double helix, (Figure 1) with an average helix rotation of 32.4°, 11.1 residues per turn and a mean rise per residue of 3.2Å. Step 4−5, as Figure 2 shows, does have some slight inter-strand stacking, which is clearly related to the extended backbone conformation at this point (see below). There is some variation in helical twist values along the sequence, with a maximum of 37.3° between steps 3 and 4, and a minimum of 26.4° between steps 4 and 5. Table 1 shows that there is a marked alternation in some base-pair parameters along the sequence, with pyrimidine-3′,5′-purine steps having consistently larger roll and smaller tilt angles. The large tilt angles of the purine-3′,5′-pyrimidine steps have alternation in sign along the helix. All individual base pairs have propeller twist, with an average of −10.0(2.0)°. There is no significant difference between the values for AT and GC basepairs. The purine-3′,5′-pyrimidine steps are well stacked, with six-membered rings almost entirely over-lapping each other. By contrast, the pyrimidine-3′,5′-purine steps are mostly completely unstacked (Figure 2) with virtually no intra or inter-strand overlap. The structure is notable for the consistency of these features along the whole of its length.

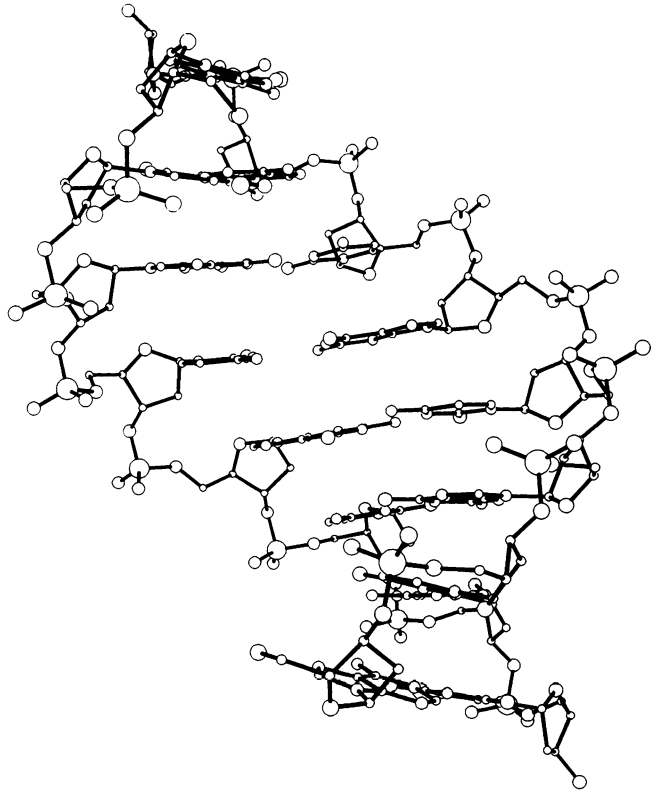

Figure 1. A view of the d(ATGCGCAT)$_2$ double helix.

Individual backbone angles generally show relatively small variation along each chain. There is a notable exception at residue thymidine-4 of strand one and its complementary adenosine on strand two, where the usually *gauche*$^-$ value for angle α, around the P-05′ bond, is now intermediate between *gauche*$^+$ and *trans* (140°) (Figure 2). Angle γ, around the C5′-C4′ bond, has a pure *trans* value here of 186°, rather than the *gauche*$^+$ elsewhere in the structure. Sugar puckers cluster around the C3′ *endo* conformation, with an average pseudorotation angle of 15°. The maximum deviations from this average occur around guanosine-5, with a pseudorotation angle of −16°, corresponding to C3′*endo* C2′*exo* pucker.

The minor groove in the spermine-free structure is remarkably constant in width, (9.7Å); there is thus no groove discontinuity around the centre of the helix. It is not really possible to describe the major groove width in such a short length of A-helix, although the P2-P10 distance of 8.5Å can be taken as some indication of it. There is no significant change in either groove width for to the spermine structure.

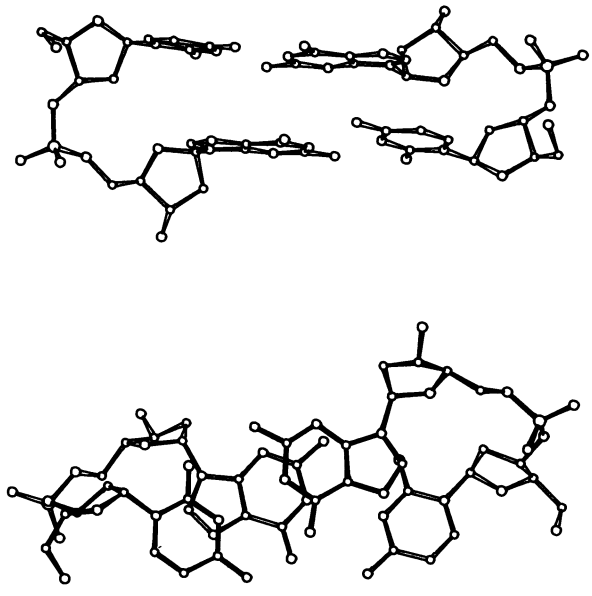

Figure 2. Views of the central C4-G5 base-pair step.

The spermine molecule in the second crystal structure was unequivocally located in the minor groove of the duplex, in the central GCGC region (Figure 3). It adopts a folded rather than an extended conformation, which closely follows the contours of the minor groove such that the -(CH_2)- groups are close to sugar residues. The terminal NH_3^+ groups of the spermine are close to the phosphate groups of cytosine 6 of strand one and cytosine 14 of strand two.

NMR Analysis

Assignments: The 1D spectrum of the d(ATGCGCAT) duplex at 25°C is shown in Figure 4. A single signal was found for all protons indicating a single conformation was present, or if multiple conformations were present, they were in fast exchange. Resonances belonging to the H1′, H2′, H2″, H3′ and aromatic protons of each nucleotide in the duplex were assigned from examination of the NOESY spectra according to standard procedures (22). The specific strategy for making the assignments has been reported previously (19). Assignments were confirmed from examination of the COSY spectra and are collected in Table 3. There were no unusual chemical shifts in this sequence, and all signals were located in the expected regions for a DNA duplex (22). All purine H8 resonances were significantly downfield of pyrimidine H6 resonances (Figure 4), an initial indicator of a right-handed B-form duplex (23,24).

Cosy Spectra: Critical features for establishing the DNA conformational family of an oligonucleotide by NMR methods are the individual deoxyribose conformations, which are most accurately determined from vicinal coupling constants (COSY experiments), and the sugar to base spatial relationships which can be determined from cross peaks in NOESY spectra (22). For each base the multiplet fine structures of COSY cross peaks arising between the H1′-H2′, H1′-H2″, H3′-H2′ and H3′-H2″ sugar protons were examined for qualitative features and for determination of vicinal coupling constants and sums of coupling constants. All measurements were made from cross sections of the 2D COSY spectra running parallel to $_2$ due to higher

TABLE 1. Base-pair parameters for the spermine-free d(ATGCGCAT) double helix, calculated with the NEWHELIX program.

Base pair	Roll (°)	Tilt (°)	Buckle (°)	Slide (Å)	Propeller twist(°)	Helical twist(°)	X Displacement(Å)	Rise per residue(Å)	Inclination (°)
A-T (1)			6		−8		−3.5		13
	8	−6		−1		31		3.4	
T-A (2)			−1		−11		−3.5		9
	13	−1		−2		31		3.2	
G−C (3)			1		−9		−3.5		7
	2	8		−1		37		3.0	
C−G (4)			−4		−12		−3.2		9
	11	0		−2		26		3.1	
G−C (5)			4		−10		−3.1		9
	4	−7		−1		35		3.0	
C−G (6)			−3		−9		−3.5		8
	13	−1		−2		32		3.1	
A−T (7)			3		−13		−3.4		8
	8	5		−1		33		3.4	
T−A (8)			−8		−7		−3.4		12
Mean	8	0	1	−1	−10	32	−3.4	3.2	10
Standard deviation	4	6	5	0	2	2	0.1	0.2	2
A−DNA	10.7	22.0	0.0	−1.5	−6.0	32.7	−4.4	2.56	19

resolution along this axis. Analysis of the 3′ terminus T8 deoxyribose was limited due to the nearly complete overlap of the 2′ and 2″ resonances. Representative cross peaks and cross sections through cross peaks are shown in Figures 5–7.

The multiplet structure of individual cross peaks in the COSY spectra were initially compared to simulations (25,40) which show the dependence of cross peak structure on sugar conformation. Visual inspection immediately defined all sugars under the conditions of the NMR experiments (excepting T8, see below) as having an average conformation which is either predominantly or exclusively C2′*endo*/C3′*exo*. Well resolved quartets were observed for the multiplet structure of H1′-H2′ and H1′-H2″ cross peaks (Figure 5), and all H1′-H2″ cross peaks had significant intensity. Conversely, H3′-H2″ cross peaks were either not observed or of minimal intensity. Cross sections parallel to $_2$ and through H3′ of H3′-H2′ cross peaks had a multiple structure containing six peaks with one exception, A1, which had seven resolvable peaks (Figure 6c). Consequently, A1, based on comparison to simulations (25) had the least C2′*endo*/C3′*exo* character (excepting T8), a not surprising observation since it is the 5′ terminus. Nonetheless, even A1 was strongly C2′*endo*/C3′*exo* in nature under the solution conditions of the NMR experiments.

Evaluation of coupling constants: A specific conformation of a deoxyribose residue can be represented in terms of two parameters, P and τ_m, where P is a pseudorotation angle, and τ_m is the amplitude of sugar pucker (26). Assuming a static model for a deoxyribose moiety in solution, a sugar conformation (P, τ_m) is specified by values for the vicinal coupling constants ($^3J_{H-H}$) through a modified Karplus equation (27,28). A C2′*endo*/C3′*exo* conformation, characteristic of B-form DNA, has a P value range of 180 +/− 90° whereas a C3′*endo*/C2′*exo* conformation characteristic of A-form DNA, has a P value range of 0 +/− 90° (29). A dynamic model in which a sugar ring rapidly interconverts between pure C2′*endo*/C3′*exo* and C3′*endo*/C2′*exo* conformations is also relevant. For such a model, the observed vicinal coupling constants now represent time weighted averages of the coupling constants for two extreme conformations, the pure south (C2′*endo*/C3′*exo*) and north (C3′*endo*/C2′*exo*) conformations:

$$J_{obs} = X_S J_S + X_N J_N$$

where X^S and X^N represent the fractions south and north, and J_{obs}, J_S, and J_N represent the time-averaged, pure south and pure

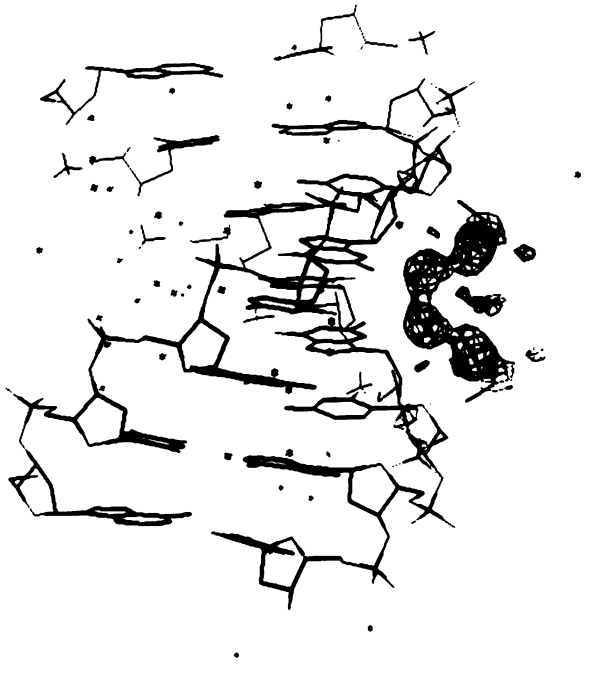

Figure 3. View, taken from the computer graphics screen, of the $2F_o$-F_c map showing density corresponding to the spermine molecule. The contouring is at the 3σ level. The minor groove of the helix is on the right-hand side.

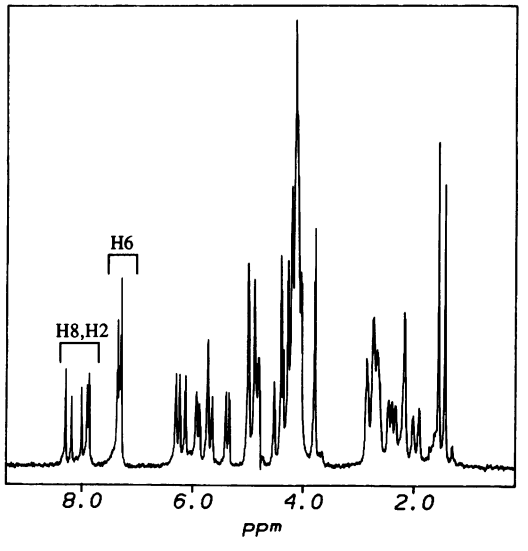

Figure 4. 600 MHz proton NMR spectrum of the d(ATGCGCAT)$_2$ duplex at 25°C in phosphate buffer, pH 7.0.

TABLE 2. Backbone torsion angles and sugar pucker parameters for the spermine−free d(ATGCGCAT)$_2$ double helix.

Residue	α	β	γ	δ	ε	ξ	P	τ_m
A1	–	–	33	83	217	196	8	38
T2	270	176	78	76	209	196	17	43
G3	287	166	66	83	203	197	26	36
C4	308	166	44	82	192	205	21	36
G5	141	190	186	96	228	185	−16	34
C6	275	178	64	76	208	194	23	47
A7	310	166	53	81	212	196	20	39
T8	235	173	28	94	–	215	21	29

TABLE 3. Proton chemical shifts (ppm) for the d(ATGCGCAT)$_2$ duplex at 25°C in phosphate buffer, pH 7.0.

Nucleotide proton	A1	T2	G3	C4	G5	C6	A7	T8
H1′	6.23	5.72	5.93	5.73	5.88	5.65	6.30	6.13
H2′	2.68	2.20	2.64	2.02	2.62	1.92	2.75	2.17
H2~″	2.84	2.47	2.73	2.40	2.72	2.34	2.87	2.17
H3′	4.88	4.88	5.00	4.86	4.99	4.81	5.01	4.52
H2	8.01	–	–	–	–	–	7.88	–
H5	–	–	–	5.34	–	5.40	–	–
CH3	–	1.45	–	–	–	–	–	1.56
H6	–	7.36	–	7.33	–	7.29	–	7.30
H8	8.19	–	7.91	–	7.88	–	8.30	–

north vicinal coupling constants, respectively (29). The two models yield qualitatively similar results (25,30) and, consequently, the vicinal coupling constants and sums of coupling constants between protons of a deoxyribose define its conformation as predominantly S, C2'*endo*/C3'*exo*, or N, C3'*endo*/C2'*exo* (29).

The vicinal coupling constants were experimentally extracted from the splitting patterns of COSY cross peaks. $^3J_{1'-2'}$, $^3J_{1'-2''}$, $^3J_{2'-3'}$, and sums of vicinal coupling constants- $\Sigma1'$, $\Sigma2'$, $\Sigma2''$, where

$$\Sigma1' = {}^3J_{1'-2'} + {}^3J_{1'-2''}$$
$$\Sigma2' = {}^3J_{1'-2'} + {}^3J_{2'-3'} + {}^3J_{2'-2''}$$
$$\Sigma2'' = {}^3J_{1'-2''} + {}^3J_{2''-3'} + {}^3J_{2'-2''}$$

were evaluated for bases 1−7 of the oligonucleotide duplex in solution and are summarized in Table 4. The calculated values confirmed that the duplex in solution had an average conformation which was predominantly C2'*endo*/C3'*exo*. As noted, the expected quartet structures for cross sections of cross peaks between H1'-H2' and H1'-H2'' were observed, allowing accurate determination of $^3J_{1'-2'}$, $^3J_{1'-2''}$, $\Sigma1'$ (Figures 6a,6b). The quartet structures reflect the contributions of both active (antiphase) and passive (same phase) couplings to the appearance of a cross peak (31). $\Sigma1'$ was greater than 14.2 Hz for all seven sugars with $^3J_{1'-2'}$ ranging from 8.9 to 10.1 Hz and $^3J_{1'-2''}$ ranging from 5.1 to 5.4 Hz. Values for c1' greater than 13.3 Hz and $^3J_{1'-2'} > {}^3J_{1'-2''}$ are considered indicative of a time averaged predominantly south (C2'*endo*/C3'*exo*) conformation (29).

Measurements of $\Sigma2'$ and $\Sigma2''$ were made on cross sections through H3'-H2' and H1'-H2'' cross peaks, respectively (Figures 6c,6d). These measurements were subject to more error due to incomplete resolution of the expected octet structures. This was particularly true for $\Sigma2''$ as only quartets were observed for the H1'-H2'' cross peak sections (6d). This observation was an indication of a predominant S conformation for the duplex as it suggested a small value (<2.7Hz) for $^3J_{2''-3'}$, another characteristic of a C2'-*endo*/C3'-*exo* conformation (25,29). The magnitude of the value of $\Sigma2'$ relative to $\Sigma2''$ for a given deoxyribose, and the approximate values of $^3J_{2'-3'}$ (5−6 Hz), and $^3J_{2''-3'}$ (<3Hz) were again fully consistent with a predominantly, if not exclusively, S conformation (Table 4) for the oligonucleotide duplex in solution.

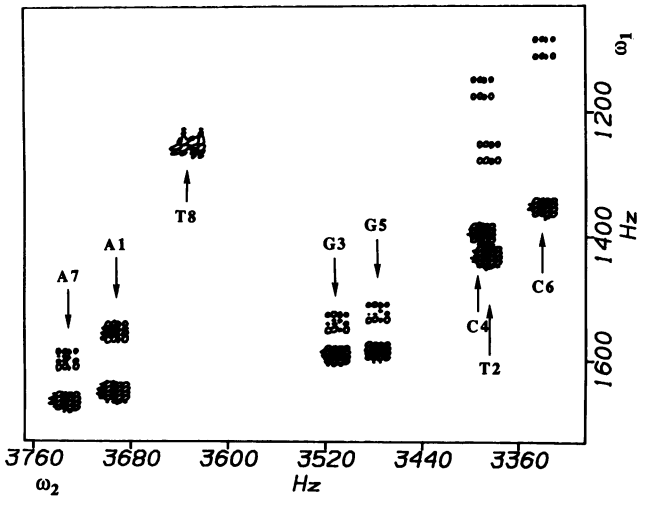

Figure 5. Expanded contour plot of the phase sensitive COSY spectrum of the d(ATGCGCAT)$_2$ duplex at 25°C showing H1' to H2' or H2'' cross peaks. For each residue, with the exception of T8, the H1' to H2' cross peak is downfield of the H1' to H2'' cross peak.

For all seven deoxyriboses the data were most consistent with a τ_m value of 40° and P values in the range of 180°. X_S was >=85% for all bases excepting T8 (see table 4). The results, particularly the values for $^3J_{2'-3'}$, are most consistent with the internal deoxyriboses being relatively static and the terminal A1 and T8 deoxyriboses being in a more dynamic state.

T8, although inaccessible to full analysis, had a time-averaged conformation which was distinct from the other seven deoxyriboses and more C3'*endo*/C2'*exo* in nature relative to the other seven sugars. This conclusion was drawn from the observation that a cross section through H3' of the T8 H3'-H2'/H2' overlapped cross peaks had four resolvable peaks whereas analogous cross sections through the other H3' resonances had only two resolvable peaks. This observation indicated that the T8 deoxyribose had a significantly larger value for $\Sigma3'$ ($^3J_{2'-3'} + {}^3J_{2''-3'} + {}^3J_{3'-4'}$) relative to other sugars, and in particular larger values for $^3J_{2''-3'}$ and $^3J_{3'-4'}$ which for a pure C2'*endo*/C3'*exo* conformation are small (1−2 Hz). c3' was at least 16.5 Hz and c1' was estimated to be 14.0 Hz (c1' and c3' are in principle unaffected by overlapping 2'/2'' resonances). These values are consistent with a value for X_S in the range of 0.7.

NOESY Spectra: COSY results provide definitive criteria for distinguishing deoxyribose conformation and making a distinction between the A and B conformational families in solution. There are, however, several critical sugar to base proton distance differences between the helical forms which provide additional information and support for classification of the conformational type (22). The base H8 or H6 to H3' distance is short (<2.5Å) in A-form but long (>3.5Å) in B-form DNA while the H8 or H6 to H2' distance is long in the A-form but short in the B conformation. In addition, for the A-form there is a short distance between the H8 or H6 of one residue and the H2' of the residue on the 5' side. In the B-form, the distance between the H8 or H6 of one residue and the H2'' of its 5' neighbour is short.

A complete 600 MHz proton NOESY spectrum for the d(ATGCGCAT)$_2$ duplex is shown in Figure 7a. The intense cross peaks in the aromatic to H2'/H2'' region and lack of such peaks in the aromatic to H3' region are readily seen. Intraresidue H8 or H6 to H2' cross peaks had strong intensity for all residues except A1 (Figure 7b). Interresidue cross peaks with moderate intensity were A1 H2'' to T2 H6, G3 H2'' to C4 H6, G5 H2'' to C6 H6, and A7 H2'' to T8 H6. The H8 to 5' H2'' cross peaks were also observed but had weak intensity. In contrast, intra-residue H8 or H6 to H3' cross peaks and inter-residue H6 or H8 to 5' H2' cross peaks were either not observed or had very weak intensity. The exception, as with the COSY analysis, was T8 which had an intra-residue H6 to H3' cross peak of moderate intensity and an inter-residue H2' to A7 H8 of weak to moderate intensity. The relative intensities of cross peaks in the 400 MHz NOESY spectrum were completely consistent with those obtained at 600 MHz even with the differences in temperature (25 vs 15°C) and the longer mixing time (100 vs 250 ms). Consequently the NOESY spectra supported a predominantly B-form conformation for all residues in agreement with the conclusions drawn from the COSY spectra.

DISCUSSION

The studies reported here show clearly that the octamer sequence d(ATGCGCAT)$_2$ is highly polymorphic, adopting the A

conformation in the crystal and is fully B in solution. Its behaviour is thus close to that of other octamer sequences with A form in the crystal (for example, 8,9). We conclude that crystal packing forces the sequence into the energetically less favourable A form: there is no evidence of any significant A population in solution. It is unsurprising that the NMR experiments find evidence for fraying of the ends of the duplex, since they are AT base pairs. It is more surprising that there is no indication of this effect in the crystal, and indeed that this sequence diffracts better than the majority of A-type octamers studied. Taken as a whole, the study shows that this particular sequence can, under appropriate environmental conditions, adopt an A structure. The excess of purine-3′,5′-pyrimidine over pyrimidine-3′,5′-purine sequences here, in a 4:3 ratio, suggests that even though the sequence is an alternating one, this factor prevents it from forming a Z-type helix.

The *transoid* values for backbone angles α and γ at the 4th base pair represent significant deviations from standard A-form double helices. This feature has also been reported at the same point in the tetragonal crystal structure of the alternating sequence d(GTGTACAC)$_2$ (32) (though not in the hexagonal form (3)), as well as in all other octanucleotide duplexes with a central pyrimidine-3′,5′-purine sequence (Table 5). All crystallise in the same tetragonal space group. This strongly supports the hypothesis that the conformational differences between tetragonal and hexagonal forms (which also produce differences in helix rise, inclination and groove width (8,9)), are environmental rather than sequence-dependent (8).

The d(ATGCGCAT)$_2$ crystal structure provides further confirmation that the central step in tetragonal octamers is always unwound (Table 5); the study also shows that a central CpG step is not required to be unwound by as much as 10°, as has recently been suggested (33). The slight compensatory overwinding, by 4°, in the immediately flanking step, has been noted in other

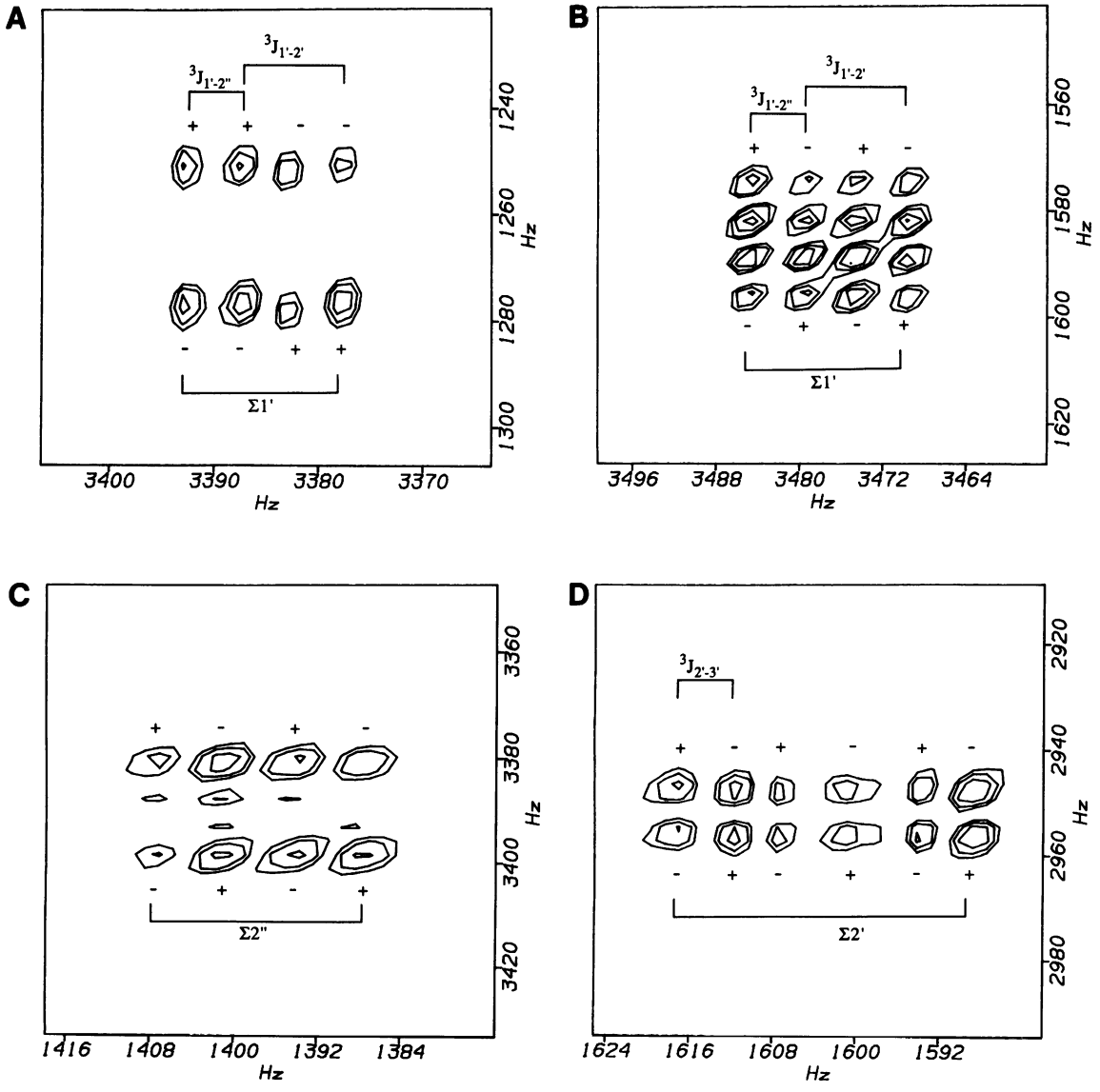

Figure 6. Contour plot of representative cross peaks observed in the phase sensitive COSY spectrum showing the distances used to calculate the coupling constants and sums of coupling constants. The $_2$ axis is the abscissa. A: T2 1′H-2′H; B: G3 1′H-2″H; C: A1 H3′-H2′; D: C4 1′H-2″H—note that opposite sides of the diagonal (B vs D) are inequivalent due to different levels of resolution along ω_1 and ω_2.

structures (32,35), although it is not a universal feature (34).

The marked alternation of roll, tilt and buckle (Table 1) which extend to the ends of the helix, may be a general feature of such

TABLE 4. Calculated H-H vicinal coupling constants and sums of vicinal coupling constants for the d(ATGCGCAT)$_2$ duplex at 25°C in phosphate buffer, pH 7.0.

Nucleotide	$^3J_{1'-2'}$	$^3J_{1'-2''}$	Σ1′	Σ2′	Σ2′	$^3J_{2'-3'}$	P	%S
A1	8.9	5.4	14.2	28.0	20.1	5.0	195	85
T2	9.8	5.2	15.0	30.1	19.3	5.9	175	100
G3	10.0	5.3	15.2	29.7	19.4	5.4	175	100
C4	9.2	5.4	14.5	29.3	20.0	5.7	180	95
G5	10.1	5.1	15.3	29.5	19.6	5.4	170	100
C6	9.4	5.4	14.7	29.6	20.3	5.9	185	95
A7	9.0	5.4	14.4	28.6	20.1	5.4	190	90
T8	–	–	14.0	–	–	–	–	70

P and %S were evaluated based on predicted coupling constants (28), assuming P_S=171°, P_N=9°, and τ_m=40°.

TABLE 5. Selected conformational and helical features at the central step (A4-B5), as reported in various A-DNA octanucleotide crystal structures.

Sequence	Central-step backbone angles		Helical Twist	Roll	Ref
	α	β			
d(GTACGTAC)	157	176	23	8	33
d(GTGTACAC)	141	186	28	6	32
d(GCCCGGGC)	−171	143	23	2	34
d(CTCTAGAG)	−204	166	21	6	35
d(CCCCGGGG)	−204	177	25	1	36
d(GGGCGCCC)	150	182	24	3	37
d(ATGCGCAT)	141	186	26	11	This paper

A form perfectly-alternating sequences; this alternation is not reflected either in backbone torsion angles (Table 2), or more surprisingly, in the highly-constant minor-groove width. If the A-form is accessible for some sequences of genomic DNA, then it may well be that specific protein recognition takes advantages of such differences in base-pair geometry, rather than solely groove-width differences. The minor groove-width of 9.7Å in d(ATGCGCAT)$_2$ is typical for A-DNA oligomers (36), and quite close to the value for A-DNA in a fibre, of 11.1Å (2). The major-groove width is much larger than the fibre value of 2.7Å. This increased size is a common feature of such octamer structures. The width approaches the 11.6Å for fibrous B-form DNA (2), again indicating that the structure is not a classic A-type. This is also seen in the low inclination values (Table 1).

The binding of spermine in the minor groove of the crystal structure is in striking contrast to the major-groove binding reported in the d(GTGTACAC)$_2$ crystal structure (32), where extensive spermine-base interactions were observed, and the spermine is in an extended conformation. In the present structure, a folded spermine molecule is exclusively bound, albeit weakly, to phosphates. Other interactions in the minor groove are hydrophobic.

The NMR results demonstrate that the d(ATGCGCAT)$_2$ sequence is of the general B-family in solution under our experimental conditions. This conclusion is based in (i) the relative magnitude of the chemical shifts of the purine H8 and pyrimidine H6 resonances; (ii) the coupling constants and sums of coupling constants of the deoxyribose protons determined from COSY spectra; and (iii) the base to deoxyribose protons distances determined from NOESY spectra. Consequenctly, the d(ATGCGCAT)$_2$ sequence joins other sequences such as the TFIIIA binding site whose crystal structure is in the A family but whose NMR structure is in the B family (7,38).

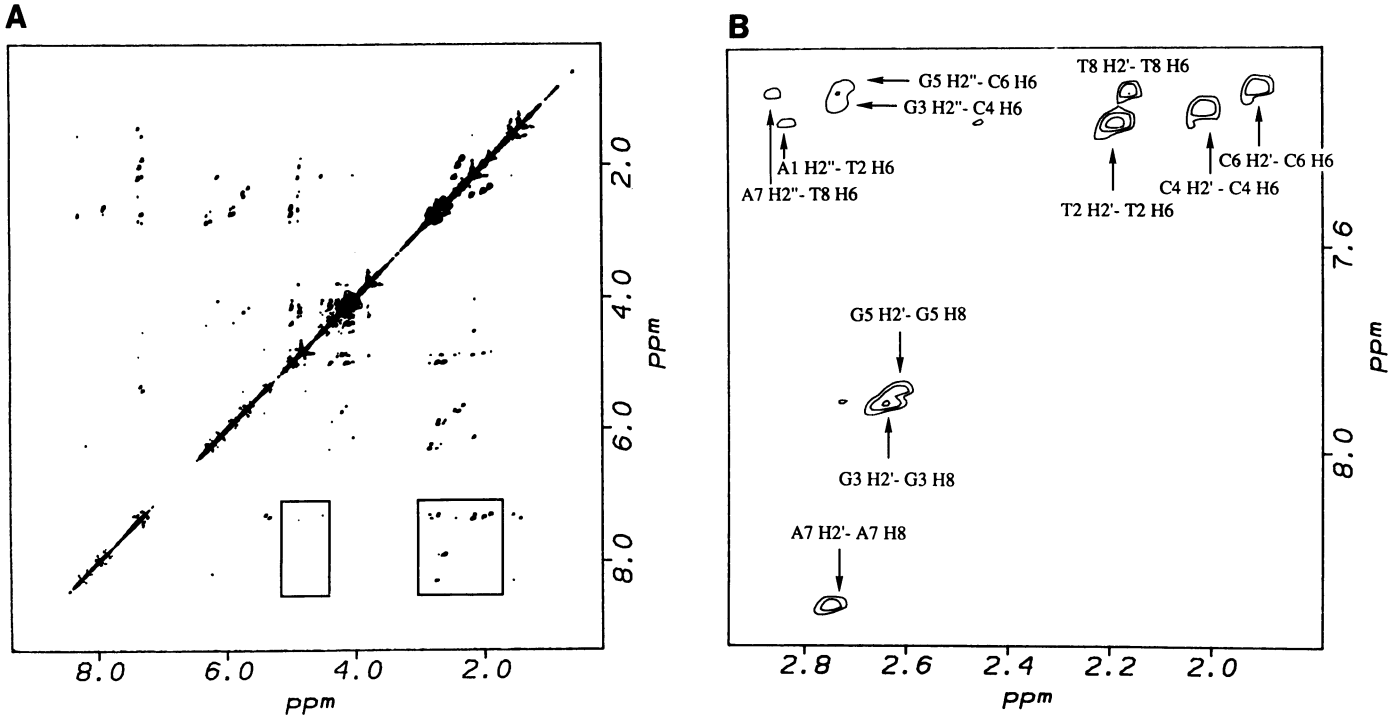

Figure 7. A: Symmetrical phase sensitive NOESY contour plot of the d(ATGCGCAT)$_2$ duplex at 25°C. The squares encompass the aromatic to H3′ (left) and aromatic to H2′/H2″ (right) regions. B: Expanded view of region containing H6 or H8 to H2′ or H2″ cross peaks.

The use of X-ray crystallography has proved invaluable in demonstrating that local DNA conformation is sequence dependent. However, the increasing inventory of sequences which crystallize in the A- or Z-form but occur in solution under physiological conditions in the B-form, raises the question as to the biological relevance of crystal structures beyond demonstrating sequence dependent variability. We argue that crystal structures, especially in conjunction with NMR studies, are of value for two primary reasons. First, the observation of distinct morphologies for a sequence in crystal and under solution conditions demonstrates its conformational flexibility. This dynamic property is difficult to establish except through comparative studies and may be of critical importance for DNA regulatory events. The second and related property of crystal structures is that they may mimic structures adopted by sequences upon ligand binding. These events often result in release of H_2O molecules, and dehydration catalyzes the B => A transition. For example, a conformational shift has been reported for binding of the chromomycin dimer to the sequence $d(TTGGCCAA)_2$ in solution (39), with the G.C rich minor groove site having A-like character.

Initial NMR studies of the $d(ATGCGCAT)_2$ sequence in its 2:1 complex with actinomycin D (ActD) indicate that ActD binds at both 5'GC3' sites in the sequence and forms a unique symmetric complex with the oligomer (19). The four large cyclic peptide groups of the two bound ActD are crowded into the minor groove, and the NMR results indicate that several of the deoxyribose residues of the complex are shifted to an N-type conformation as expected for the A-form helix in DNA. Such a conformational shift would result in a wider and more shallow minor groove and could relieve steric clash of the large peptide groups with the DNA residues in the minor groove. More detailed NMR studies of this complex are currently underway in an effort to more accurately evaluate the extent of conformational transition of the $d(ATGCGCAT)_2$ sequence upon binding of ActD, and we have also initiated crystallization experiments on the 2:1 complex.

ACKNOWLEDGEMENTS

We are grateful to the Cancer Research Campaign for the award of a Senior Visiting Fellowship (to GRC) and a Research Studentship (to DGB). The studies in Atlanta were supported by NIH grant NIAID AI-27196, and the Sutton-Atlanta collaboration by a NATO grant (to SN and WDW). We thank the following for sending us unpublished coordinates: U. Heinemann, W. N. Hunter, A. H.-J. Wang and M. Sundaralingam. R. E. Dickerson is thanked for provision of his NEWHELIX program, and E. Westhof for much help in the use of the NUCLSQ refinement programs.

REFERENCES

1. Shakked, Z. and Rabinovich, D. (1986) *Prog. Biophys. Molec. Biol.*, **47**, 159–195.
2. Chandrasekaran, R. and Arnott, S. (1989). In: *Landolt-Bornstein, Group VII, Vol. 1b*, Saenger, W, ed. Springer-Verlag, Berlin.
3. Yoon, C.; Priv, G. G., Goodsell, D. S. and Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6332–6336.
4. Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. and Sigler, P. B. (1988) *Nature*, **335**, 321–329.
5. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. and Harrison, S. C. (1988) *Science*, **242**, 899–907.
6. Rhodes, D. and Klug, A. (1986) *Cell*, **46**, 123–132.
7. McCall, M., Brown, T., Hunter, W. N. and Kennard, O. (1986) *Nature*, **322**, 661-664.
8. Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F. and Rabinovich, D. (1989) *Nature*, **342**, 456–460.
9. Jain, S. and Sundaralingam, M. (1989) *J. Biol. Chem.*, **264**, 12780–12784.
10. Drew, H. R. and Dickerson, R. E. (1981) *J. Mol. Biol.*, **151**, 535–556.
11. Hare, D. R., Wemmer, D. E., Chon, S. H., Drobny, G. H. and Reid, B. R. (1983) *J. Mol. Biol.*, **171**, 319–336.
12. Wang, A. H.-J., Fujii, S., van Boom, J. H. and Rich, A. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3968–3972.
13. GEMINI, a molecular modelling program written for Silicon Graphics IRIS workstations by Dr. A. Beveridge at the Institute of Cancer Research. Available from Hampden Data Systems Ltd., Foxcombe Court, Wyndyke Furlong, Abingdon Business Park, Abindgon, Oxon OX14 1DZ.
14. Sussman, J. L., Holbrook, S. R., Church, G. M. and Kim, S.-H. (1977) *Acta Crystallogr.*, **33**, 800–804.
15. Westhof, E., Dumas, P. and Moras, D. (1985) *J. Mol. Biol.*, **184**, 119–145.
16. TOM, a version of FRODO (T. A. Jones, in *Computational Crystallography* (1982), ed. Sayre, D., Clarendon Press, Oxford, pp. 303–310), amended to run on a Silicon Graphics IRIS workstation by C. M. Cambillau (1988).
17. Steigemann, W. (1974). Dissertation Technische Universität, Munich, West Germany.
18. (1989) *EMBO J.*, **8**, 1–4.
19. Scott, E. V., Zon, G., Marzilli, L. G. and Wilson, W. D. (1988) *Biochemistry*, **27**, 7940–7951.
20. States, D. J., Haberkorn, R. A. and Ruben, D. J. (1982) *J. Mag. Res.*, **48**, 286-292.
21. Otting, G., Widmer, H., Wagner, G. and Wuthrich, K. (1986) *J. Mag. Res.*, **66**, 187–193.
22. Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids (John Wiley and Sons, Inc. New York).
23. Chou, S., Flynn, P. and Reid, B. (1989) *Biochemistry*, **28**, 2422–2435.
24. Wolk, S., Thurmes, W. N., Ross, W. S., Hardin, C. C. and Tinoco, I. (1989) *Biochemistry*, **28**, 2452–2459.
25. Widmer, H. and Wuthrich, K. (1987) *J. Mag. Res.*, **74**, 316–336;
26. Altona, C. and Sundaralingam, M. (1972) *J. Am. Chem. Soc.*, **94**, 8205–8212.
27. Haasnoot, C. A. G., de Leeuw, F. A. A. M., de Leuw, H. P. M. and Altona, C. (1981) *Org. Mag. Reson.*, **15**, 43–52.
28. Haasnoot, C. A. G., de Leeuw, F. A. A. M. and Altona, C. (1980) *Tetrahedron*, **36**, 2783–2792.
29. Rinkel, L. J. and Altona, C. (1987) *J. Biomol. Struct. Dyn.*, **4**, 621–649.
30. Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A. and Wuthrich, K. (1989) *Biochemistry*, **28**, 1462–1471.
31. Aue, W. P., Bartholdi, E. and Ernst, R. R. (1976) *J. Chem. Phys.*, **64**, 2229–2246.
32. Jain, S., Zon, G. and Sundaralingam, M. (1989) *Biochemistry*, **28**, 2360–2364.
33. Takusagawa, F. (1990) *J. Biomol. Structure and Dynamics*, **7**, 795–809.
34. Heinemann, U., Lauble, H., Frank, R. and Blocker, H. (1987) *Nucleic Acids Res.* **15**, 9531–9550.
35. Hunter, W. N., Langlois D'Estaintot, B. and Kennard, O. (1989) *Biochemistry* **28**, 2444–2451.
36. Haran, T. E., Shakked, Z., Wang, A. H.-J. and Rich, A. (1987) *J. Biomol. Structure and Dynamics*, **5**, 199–217.
37. Rabinovich, D., Haran, T., Eisenstein, M. and Shakked, Z. (1988) *J. Mol. Biol.*, **200**, 151–161.
38. Aboul-ela, F., Varani, G., Walker, G. T. and Tinoco, I. (1988) *Nucleic Acids Res.*, 16, 3559–3571.
39. Gao, X. and Patel, D. (1988) *Biochemistry*, 27, 1744–1751.
40. Schmitz, U., Zon, G. and James, T. L. (1990) *Biochemistry*, **29**, 2357–2368.