Structural Variation in d(CTCTAGAG). Implications for Protein–DNA Interactions†

William N. Hunter, Beatrice Langlois D'Estaintot, and Olga Kennard*
University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.
Received August 1, 1988; Revised Manuscript Received November 7, 1988

ABSTRACT: Single-crystal X-ray diffraction techniques have been used to characterize the structure of the self-complementary DNA oligomer d(CTCTAGAG). The structure was refined to an R factor of 14.7% using data to 2.15-Å resolution. The tetragonal unit cell, space group $P4_32_12$, has dimensions a = 42.53 and c = 24.33 Å. The asymmetric unit consists of a single strand or four base pairs. Two strands, related by a crystallographic dyad axis, coil about each other to form a right-handed duplex. This octamer duplex has a mean helix rotation of 32°, 11.3 base pairs per turn, an average rise of 3.1 Å, C3′-endo furanose conformations, a shallow minor groove, and a deep major groove. Such averaged parameters suggest classification of the octamer as a member of the A-DNA family. However, the global parameters tend to mask variations in conformational parameters observed at the level of the base pairs. In particular, the central TpA (=TpA) step displays extensive interstrand purine–purine overlap and an unusual sugar–phosphate backbone conformation. These structural features may be directly related to certain sequence-specific protein–DNA interactions involving nucleases and repressors.

Protein–DNA interactions regulate many cellular events. DNA functioning in these biological processes requires flexibility and structural variation. The sequence-dependent properties of DNA make an important contribution to the high degree of specificity exhibited in many protein–DNA interactions (Saenger, 1984). A detailed knowledge of the relevant protein and DNA structures is essential for a full understanding of the mechanism of interaction at the molecular level. Three-dimensional information can be obtained by single-crystal X-ray diffraction techniques (Blundell & Johnson, 1976). These methods have greatly improved our understanding of DNA conformation, which was previously based on the relatively inaccurate models derived from fiber diffraction studies (Saenger, 1984). More recently, diffraction studies on DNA binding proteins, some complexed with the relevant DNA fragments, have provided exciting new insights into cellular processes [reviewed by Ollis and White (1987)].

Our interest is in understanding the factors controlling DNA conformation and the relationship with sequence and in looking for specific features that could play a role in recognition by proteins. We have carried out a structural analysis of the deoxyoligonucleotide d(CTCTAGAG) and now present the refined crystal structure with 35 solvent molecules located per strand. Full experimental details are given. All important structural parameters are listed and, where appropriate, comparisons are made with previously determined structures. The structure of d(CTCTAGAG) displays a variation in local conformational parameters that may be related to the sequence-specific protein–DNA interactions involving micrococcal and S1 nuclease and certain repressors. These are discussed in detail.

Materials and Methods

Synthesis, Crystallization, and Data Collection. The oligomer d(CTCTAGAG) was synthesized by phosphoramidite methodology (Applied Biosystems 381A synthesizer) and then purified by ion-exchange and reverse-phase high-pressure liquid chromatography. Our experience indicates that a high degree of purity is required for successful crystallization. With the present compound, crystals were grown at 18 °C by vapor diffusion from droplets sitting in Corning glass depression plates (McPherson, 1982). Well formed tetragonal bipyramids reached a usable size over a period of 3–4 weeks from solutions containing about 6 mg/mL octamer, 60 mM sodium cacodylate buffer (pH 6.8), 25 mM magnesium chloride, 1 mM spermine hydrochloride, and 7% (v/v) 2-methyl-2,4-pentanediol (MPD) diffused against 50% (v/v) MPD.

Precession photography indicated space groups $P4_12_12$ or $P4_32_12$. Unit cell dimensions from diffractometer measurements (Syntex $P2_1$ Cu Kα radiation, equipped with a helium-filled long arm) gave a = 42.52 and c = 24.33 Å. A single crystal of dimension 0.25 × 0.25 × 0.15 mm was mounted and sealed in a glass capillary and used to collect intensity data in the 2θ range 2.5–42°. This range corresponds to an upper resolution limit of 2.15 Å. Intensities were corrected for Lorentz and polarization factors, absorption, and time-dependent decay. Symmetry equivalent reflections were averaged. In total, 2320 measurements reduced to 1378 unique reflections with R_{sym} = 7.5%. Approximately 70% of the data has $I > \sigma(I)$.

Structure Solution and Refinement. Cell dimensions and symmetry suggested that the crystal structure of d(CTCTAGAG) was isomorphous with several other oligomers whose structures have already been fully characterized (Conner et al., 1984; Haran et al., 1987; Heinemann et al., 1987; Jain et al., 1987; Rabinovich et al., 1988; Wang et al., 1982a). The initial model selected for the structure analysis was that of d(GGCCGGCC) in space group $P4_32_12$ (Wang et al., 1982a). The subsequent analysis showed that the choice of space group was correct and that the model was a satisfactory starting point for the refinement.

The starting model was refined against the experimental data as a rigid body by using a modified version of SHELX (Sheldrick, 1976). Initially only data in the 10–5-Å range were included in the calculations. The resolution was gradually

†This study was funded by an MRC program grant to O.K.

0006-2960/89/0428-2444$01.50/0

extended to 9–3 Å, where convergence was obtained with an R factor of 38% for 513 reflections with $I > \sigma(I)$. Electron density ($2F_o - F_c$) and difference density ($F_o - F_c$) maps were calculated and displayed on an Evans and Sutherland PS350 computer graphics system using the FRODO program (Jones, 1978). The required sequence changes were introduced and the model coordinates manipulated to fit the electron density. The refinement procedure was continued with the restrained least-squares method of Hendrickson and Konnert (1981) using NUCLSQ, a program modified for nucleic acid components (Westhof et al., 1985).

A search for solvent molecules was initiated as soon as the resolution was extended to the limit of the available data. Solvent positions were identified from electron and difference density maps. Our selection criteria were that a well-shaped peak in the difference density map, greater that 2 standard deviations in height, was within 2.2–3.3 Å of plausible hydrogen-bonding partners and that upon subsequent refinement reasonable geometry, electron density, and thermal parameters were observed. Each solvent position was treated as an oxygen atom of fixed occupancy, and the thermal parameters were allowed to refine freely, thus compensating for partial occupancy. No unequivocal assignments of cations could be made at this limited resolution.

The refinement was terminated after the location of 35 solvent molecules (70 per duplex). The final R factor was 14.7% for the 931 reflections with $I > \sigma(I)$ in the range 8–2.15 Å. On a qualitative basis, the correctness of the structure is indicated by an excellent fit of atomic positions to electron density. The overall root-mean-square error in atomic positions can be estimated by the method of Luzzati (1952) and is calculated to be approximately 0.2 Å. It should be noted that negligible restraints were employed on sugar conformations and none on sugar–phosphate torsion angles and that every effort was made to ensure that artifacts from the initial model were not carried through this analysis. We have used a similar weighting scheme for the refinement in the analysis of a number of oligonucleotide structures and found it to be suitable to let the sugar conformations refine to some minimum determined by the diffraction data, not imposed by any preconceived ideas about which conformation should be present. Refined coordinates are being deposited with the Cambridge Crystallographic Database.

Results

Duplex Structure and Crystal Packing. The asymmetric unit in the crystal structure of d(CTCTAGAG) is a single strand or 4 base pairs. Two strands coil about each other to form a right-handed double helix with Watson–Crick base pairing. A crystallographic dyad, which traverses the unit cell diagonal, is coincident with the duplex dyad. In the duplex, nucleotides are numbered 1–8 on strand 1, 9–16 on strand 2 in the 5′ to 3′ direction. The Watson–Crick base pairs are therefore C1·G16, T2·A15, etc.

The overall helix structure can be considered as A-type DNA. The sugar conformations are generally C3′-endo; there is a deep major groove and a shallow minor groove. The average rise per base pair is 3.1 Å, with an average rotation per base pair value of 32°, which results in 11.3 base pairs/turn of helix. The average displacement of bases from the helix axis is 3.6 Å. In Figures 1 and 2 we present several views of the duplex that show the overall A-type shape.

The average values quoted above, if taken by themselves, offer no more information than the averaged models derived from fiber diffraction studies. The quality of the single-crystal diffraction measurements, however, enables us to obtain a structure sufficiently detailed to observe structural features that may be dependent upon the sequence. These structural details will be discussed below. The crystal packing in d(CTCTAGAG) is similar to that encountered in all other A-type structures so far analyzed with the terminal base pairs of one duplex lying in the minor groove of adjacent duplexes leading to extensive van der Waals interactions. There are also two interduplex hydrogen bonds between O5′(C1) and O2(C1) and the symmetry-related atoms N3(A5) and N2(G6), respectively.

Table I: Geometrical Properties of Base-Pair Steps and Base Pairs

base pair	step	roll (deg)	slide (Å)	twist (deg)	rise (Å)	base-pair tilt (deg)	propeller twist (deg)
C1·G16						11	10
	1	6	1.4	33	3.3		
T2·A15						10	13
	2	11	1.6	32	3.1		
C3·G14						8	7
	3	5	1.4	41	2.9		
T4·A13						10	12
	4	6	1.7	21	3.2		
A5·T12							
av		7	1.5	32	3.1	10	11

Base Stacking Interactions. In the crystal structure of d(CTCTAGAG) there are four base-pair steps, two CpT (=ApG), one TpC (=GpA), and one TpA (=TpA). Only the TpA (=TpA) step has been observed in previous crystallographic studies of A-type DNA (Jain et al., 1987; Shakked et al., 1983; Wang et al., 1982b). Each base-pair step is portrayed in Figure 3. These diagrams allow a qualitative assessment of the stacking interactions. Points to consider are the degree of overlap from one base pair to the other and whether the overlap involves interstrand, intrastrand, or both types of interaction.

Steps 1–3 are pyrimidine–pyrimidine (=purine–purine) steps. The adjacent purines all interact with the five-membered ring of one base lying over the six-membered ring of the other. The overlap of adjacent pyrimidines is generally reduced, although the methyl groups of the thymines do appear to interact strongly with the cytosines in the CpT (=ApG) steps (steps 1 and 3). Steps 1 and 3 also show a small overlap involving atoms of bases on opposing strands. This cross-strand interaction occurs on the minor-groove side of the duplex and involves the amino N2 of the quanines and the keto O2 of the thymines. This interaction cannot occur at step 2, as the N2 group is absent due to replacement of guanine by adenine.

The central TpA (=TpA) step displays a marked degree of cross-strand purine–purine overlap involving the six-membered rings of each adenine. The overlap of adjacent bases is poor. This pyrimidine–purine step looks very similar to the TpA (=TpA) steps observed in d(GGTATACC) (Shakked et al., 1983) and r(GCG)d(TATACGC) (Wang et al., 1982b) and also to the central CpG (=CpG) steps found in d(CCCCGGGG) (Haran et al., 1987), d(GCCCGGGC) (Heinemann et al., 1987), d(GGGCGCCC) (Rabinovich et al., 1988), and d(GGCCGGCC) (Wang et al., 1982a).

Further analysis of base-pair interactions can be carried out in a quantitative manner using parameters calculated with the programs HELIX, BROLL, and TORANG (Dickerson, 1983; Dickerson & Drew, 1984; Calladine & Drew, 1984). In Table I we present values for roll, slide, twist, and rise per base pair for the four-pair steps and the propeller twist for the four unique base pairs.

The angle by which adjacent base pairs open up to the minor groove is the roll. In d(CTCTAGAG) steps 1, 3, and 4 display

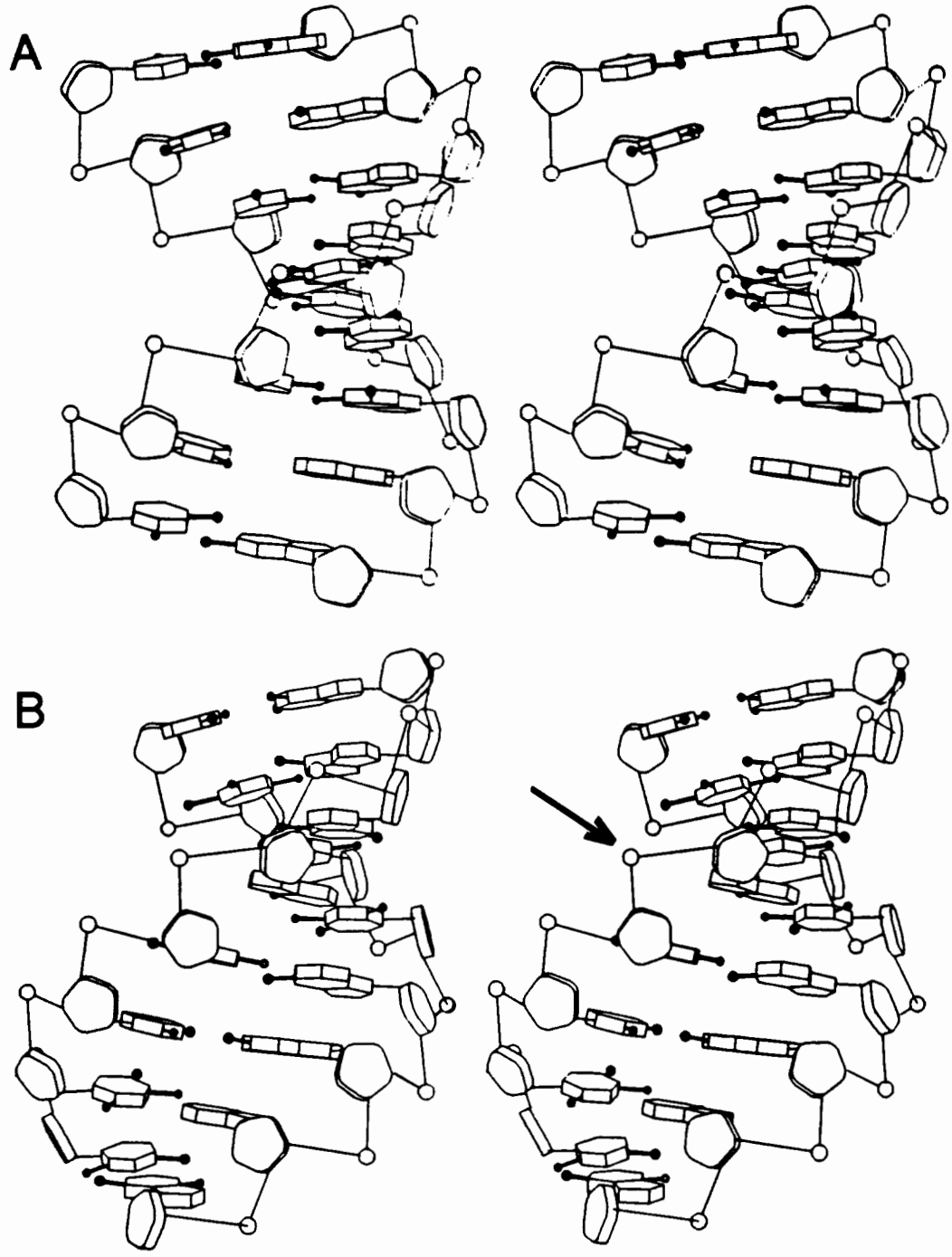

FIGURE 1: Schematic representation of the crystal structure of d(CTCTAGAG). Two stereoviews are given: (A) Side view showing both the major and minor groove; (B) a view showing most of the minor groove. Bases are shown as flat blocks, sugars as pentagons with rounded edges. The phosphate backbone is represented by a sphere for the phosphorus atoms and thin lines joining this atom to the furanose C3′ and C4′ positions. The phosphate group at A5, where an unusual backbone conformation is observed, is arrowed. These figures were produced with the plotting program of A. Lesk.

low roll values which indicate that the relevant base pairs are nearly parallel to each other. Step 2 displays the highest roll value in this structure, 11°. The distance by which adjacent base pairs are displaced along their long axis with respect to each other is termed longitudinal slide. Slide values observed in this present study are fairly constant, with an average value of 1.5 Å. The highest value is 1.7 Å at the TpA (=TpA) step. The longitudinal slide values for py·py(=pu·pu) steps appear to be similar in many of the structures done to date.

Twist represents the rotation per base-pair step. The average value of 32° results in 11.3 base pairs/turn of helix. These values correspond closely to those of the fiber diffraction models. The range in d(CTCTAGAG) is large, varying from 21 to 41°. The smallest value is observed at the central region, the pyrimidine–purine step.

The rise per base pair shows a range of 2.9–3.3 Å, with an average of 3.1 Å. The tilt of base pairs from a plane perpendicular to the helix axis averages 10°. In classical A-DNA this value is about 20°, while for B-DNA a value of −8° is observed. The mean distance of base pairs from the helix axis is 3.6 Å. This is almost 1 Å less than found for classical A-DNA. Propeller twist is larger for the A·T base pairs (average 12°) than for the G·C base pairs (average 9°).

As mentioned above, the central TpA (=TpA) step looks very similar to the central CpG (=CpG) steps observed previously. We note, however, that the propeller twist value at

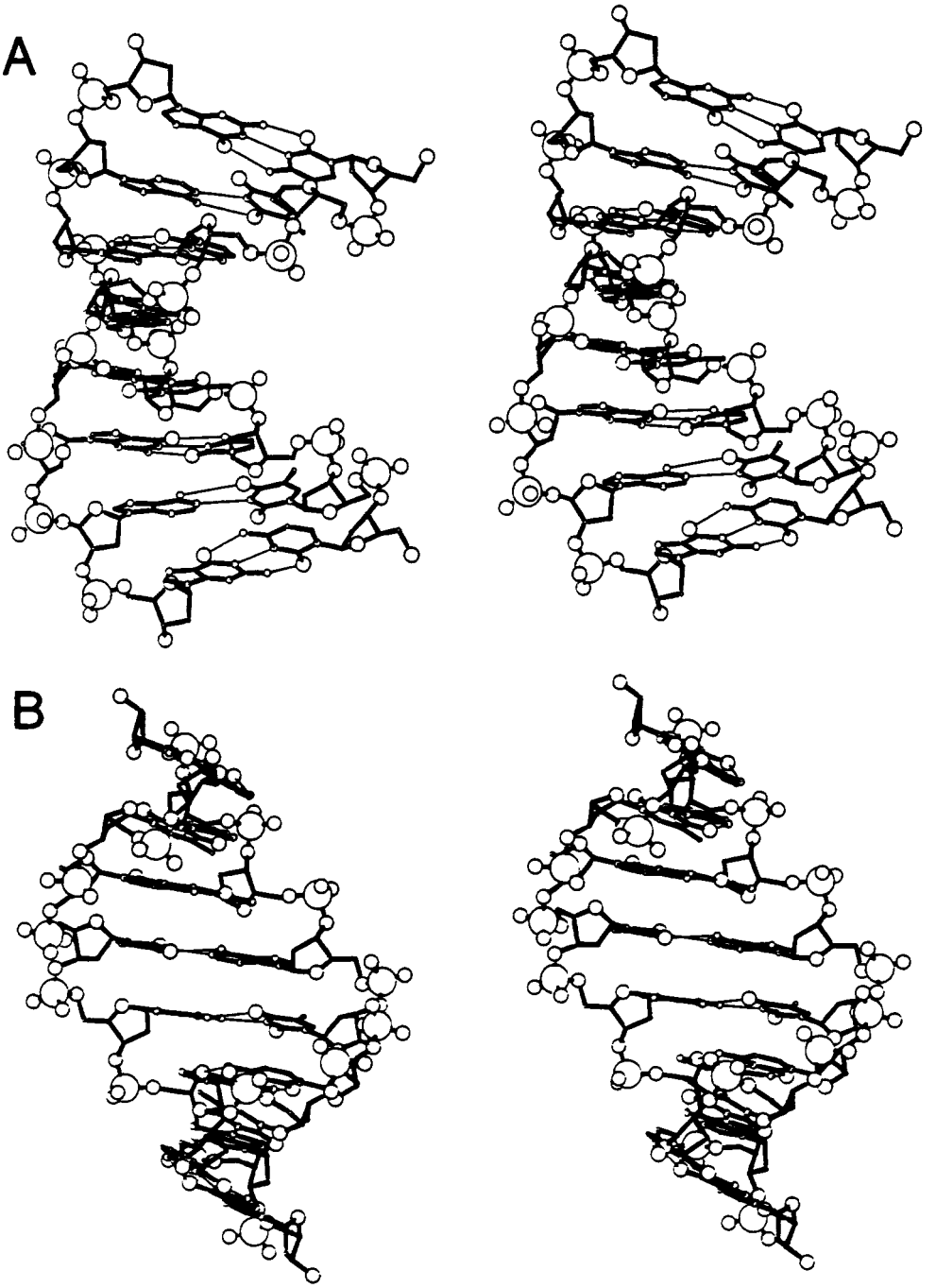

FIGURE 2: Stereoviews of the octamer: (A) Side view showing both the major and minor grooves; (B) a view into the deep major groove. Atoms are represented as spheres of decreasing radii in the order P, O, N, C. Covalent bonds are shown as thick lines, Watson–Crick hydrogen bonds as thin lines.

the TpA (=TpA) step increases by almost 9°. This twisting may enhance the stacking interactions in d(CTCTAGAG) and may be due to the absence of the amino group in the minor groove that would lead to steric clash. The central TpA (=TpA) step shows lower twist and higher slide values than other steps in the structure.

The Sugar–Phosphate Backbone. The sugar–base glycosidic angles (χ) and individual sugar–phosphate torsion angles are presented in Table II. In general, the average values correspond closely with those observed in other A-type oligonucleotide structures (Rabinovich & Shakked, 1986). The values for α and γ of A5 are, however, unusual and will be discussed below.

Values of χ are −anticlinal/−antiperiplanar; α falls in the −synclinal/−anticlinal range if α A5 is ignored. The β and γ angles are typically in the range ±antiperiplanar and +synclinal, respectively, while ϵ and ζ torsion angles fall in the ±antiperiplanar and −synclinal ranges, respectively.

Table II: Torsion Angles (deg) and Distances between Adjacent Phosphorus Atoms (Å)[a]

residue	χ	α	β	γ	δ	ϵ	ζ	P–P + 1
C1	−162			209	54	−171	−60	
T2	−156	−78	−190	57	87	−154	−78	6.1
C3	−147	−58	−181	42	83	−203	−30	6.2
T4	−145	−66	−176	35	66	−192	−43	5.6
A5	−180	−204	−151	166	66	−165	−53	7.1
G6	−164	−94	−169	56	83	−156	−58	5.7
A7	−156	−64	−169	40	90	−183	−55	6.1
G8	−140	−98	−151	83	101			

[a] Main-chain torsion angles defined by $P^{\alpha}O5'^{\beta}C5'^{\gamma}C4'^{\delta}C3'^{\epsilon}O3'^{\zeta}P$. Glycosyl torsion angle (χ) defined by O4′–C1′–N1–C2 for pyrimidines and O4′–C1′–N9–C4 for purines.

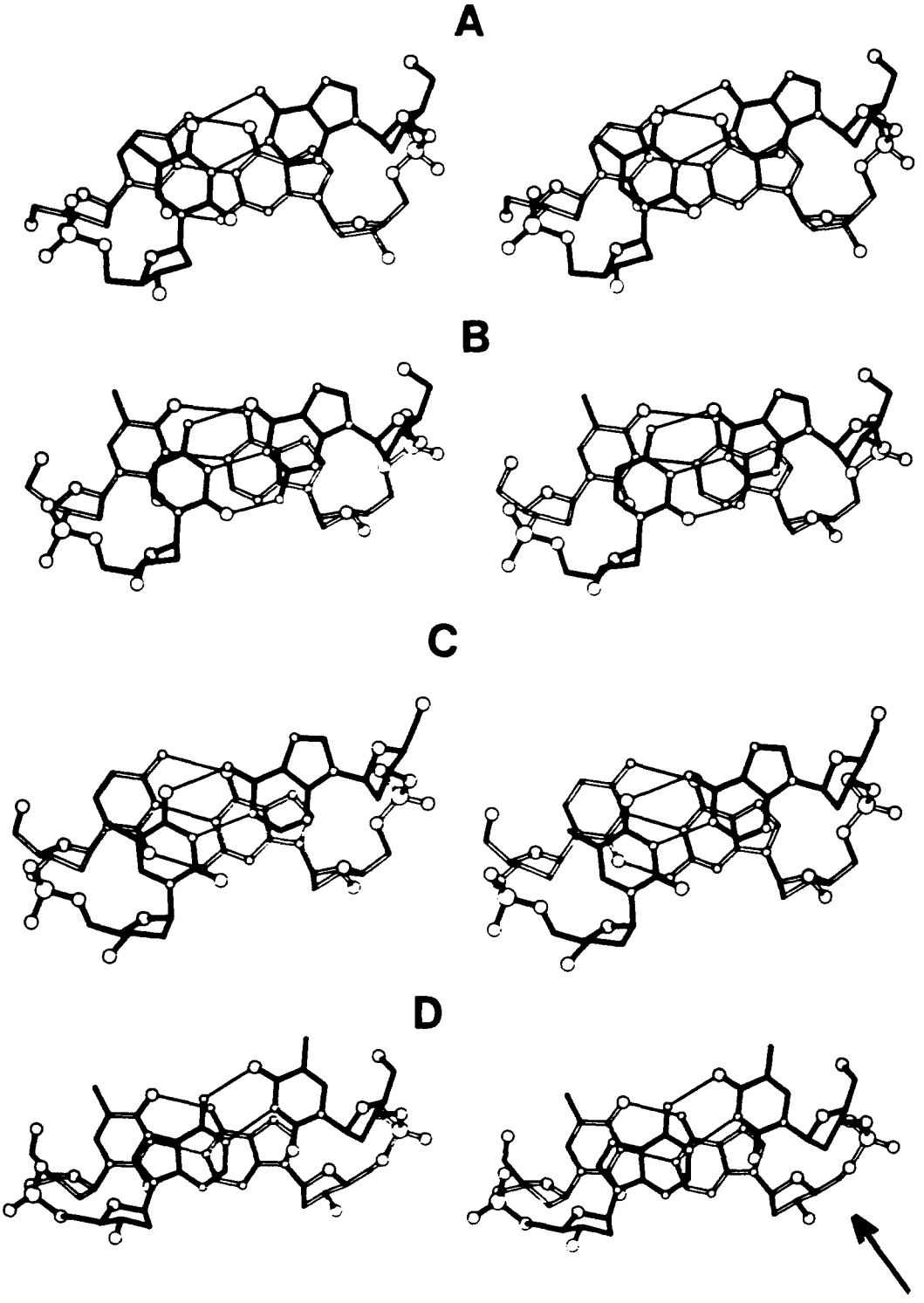

FIGURE 3: Stereoviews of the four base-pair steps in the crystal structure of d(CTCTAGAG). The view selected is perpendicuar to the best plane through the lower base pair. The major groove is to the top of each diagram. To highlight the upper base pair, the covalent bonds are filled in. Hydrogen bonds are shown as thin lines. Atoms are depicted as spheres of decreasing radii in the order P, O, N, C. (A) Shows step 1, C1·G16:T2·A15; (B) step 2, T2·A15:C3·G14; (C) step 3, C3·G14:T4·A13; (D) step 4, T4·A13:A5·T12. The unusual backbone conformation is arrowed.

Torsion angle δ is a useful guide to the conformation of the furanose ring (Levitt & Warshel, 1978) and in conjunction with the pseudorotation parameters (Altona & Sundaralingam, 1972) allows an assignment of the sugar puckering mode. The furanose conformations of residues 1–7 are C3′-endo as evidenced by δ values (range 54–90°, average 76°), the amplitude of pucker (range 40–59, average 49), and the pseudorotation phase angle (range 5–40°, average 21°). Residue G8 has a furanose ring with C1′-endo conformation. This unusual puckering may simply be due to the extra freedom associated with the sugar as there are no restraints on conformation imposed by a 3′ residue.

The α and γ A5 angles are +antiperiplanar and produce an all-trans conformation about the P–O5′–C5′–C4′ bonds. A similar observation was reported at the central pyrimidine–purine step in d(CCCCGGGG) and d(GCCCGGGC). In d(GGCCGGCC) the α angle was observed to be trans; the γ angle can be considered to lie between the gauche and trans conformations. In these sequences there is a marked degree of interstrand overlap of the guanines. In the present study

similar stacking interactions are observed despite replacement of the G·C base pairs by A·T pairs. This central step has the smallest helical twist (21°) due to the strong cross-strand stacking. Cross-strand interactions are a dominant feature in these fragments. They force the pliable sugar–phosphate backbone to adjust with a resulting extension and compression in the backbone and unusual conformations. This unusual structural feature is arrowed in Figures 1 and 2. A close inspection of the central CTAG region and all symmetry-related atoms within 20 Å leads us to conclude that this unusual backbone conformation does not arise as a consequence of crystal packing forces. Although full details for the structure of d(GTGTACAC) are not yet available, inspection of the published stereoviews (Jain et al., 1987) suggests that very similar stacking interactions and backbone conformations occur, being observed at that central TpA (=TpA) step. The intrastrand adjacent phosphorus atom distances given in Table II highlight this unusual structural feature. The normal value is about 6.0 Å in A-DNA and 7.0 Å in B-DNA (Saenger et al., 1986). At A5 the distorted backbone results in one value of 7.1 Å.

The P···P distances across the helix from one strand to the other provide an approximation of groove width. Groove width is calculated as the P···P separation minus 5.4 Å to account for the van der Waals radii of two phosphate groups. In d(CTCTAGAG) the minor-groove width is quite variable. Distances between phosphorus atom 4:16, 5:15, 6:14, 7:13, and 8:12 range from 6.8 to 9.7 Å with an average of 8.7 Å. The narrowest point of the minor groove involves the phosphate group A5, where the backbone is extended. This value of 6.8 Å is comparable to the width of the minor groove at the CGCG ends of the B-form dodecamer structure (Fratini et al., 1982). The fiber models for A- and B-DNA have minor-groove widths of about 11 and 6 Å, respectively (Arnott & Hukins, 1972; Saenger, 1984).

Due to the shortness of the sequence only a very approximate guide to major-groove width can be obtained. This is found from the P2···P10 distance of 8.0 Å. This value is much greater than predicted from fiber diffraction studies, where values for A- and B-DNA are approximately 4 and 11–12 Å, respectively. An even wider major groove was observed in the nonamer d(GGATGGGAG) (McCall et al., 1986).

Thermal Parameters and Hydration. Individual thermal parameters were refined with weak restraints between linked atoms. The average thermal parameters (Å^2) for phosphate, furanose, and base groups are approximately 15, 13, and 7, respectively. This decrease from phosphates to sugars to bases has been observed in all oligonucleotide structures published to date and shows the increased flexibility of the phosphate–sugar backbone relative to the bases. The average thermal parameter for the solvent molecules is 35 Å^2, the range from 15 to 55 Å^2. Crystallography can only locate the most ordered solvent positions; hence, we obtain an idea of where the most stable hydration points are in the structure (Westhof, 1987). Given that DNA structure is dependent upon the degree of hydration, it is important that care is taken in the location of the hydration sites. In the crystal structure of d(CTCTAGAG) we have located 35 solvent positions per strand. Every solvent position has one or more contacts to a hydrogen-bonding partner in the range 2.4–3.5 Å. About 60% of the solvent positions identified are in the primary hydration shell. The phosphate groups are the main hydration sites. As previously noted in the structure of d(GCCCGGGC) (Heinemann et al., 1987), the phosphates are individually hydrated. In d(CTCTAGAG) there is only one inter-phosphate bridging solvent position. Given the close proximity of many of the phosphate groups, this is surprising and not in agreement with some previous observations of A-type structures (Saenger et al., 1986). Several of the potential hydration sites in the minor groove are protected from interacting with solvent due to the presence of symmetry-related duplexes.

Discussion

We have determined the crystal structure of d(CTCTAGAG) and presented the results of our analysis. The right-handed duplex formed in the crystal displays an overall A-type structure, but on close inspection it becomes apparent that a number of structural features are very similar to those of B-type DNA. Even the hydration pattern is more like that of B-DNA with the phosphates individually hydrated (Saenger et al., 1986). Our crystal structure clearly shows oligonucleotide conformations intermediate between the monotonous extremes represented by fiber diffraction A- and B-DNA models.

Is this diffraction study indicative of what occurs in solution and hence, by implication, to biological processes? Consider the following points regarding the relevance of X-ray-derived structures. A single-crystal structure represents some minimum energy conformation attainable by the molecule. As such, it could contribute to an equilibrium distribution and thus be of biological significance. Oligonucleotide crystals contain about 50% by weight bulk solvent. The environment of large areas of these DNA fragments must closely resemble the environment in solution. The crystal lattice is of course stabilized by interactions between symmetry-related molecules, but it is important to realize that forces stabilizing a crystal of this type are precisely those involved in biological processes, namely, electrostatic and van der Waals forces.

Our crystal structure of d(CTCTAGAG) can be compared with the structure observed in solution by circular dichroism (CD). This technique gives an idea about the global conformation of DNA in solution and can differentiate between A-, B-, and Z-DNA (Cantor & Schimmel, 1980). Fairall et al. (unpublished results) have used CD to investigate the structures of several oligonucleotides including d(CTCTAGAG). This oligomer gave a CD spectrum that was attributed to a conformation intermediate between A- and B-DNA, thus suggesting that our structure is a representative model for this specific sequence. As such, it may offer insight into some sequence-dependent protein–DNA interactions involving CTAG sites.

The dominant feature in the structure occurs at the central TpA (=TpA) step. Pyrimidine–purine steps are highly bistable (Calladine & Drew, 1984) since the purines can have strong stacking interactions either with the base on the same strand or with the base of the opposing strand. In d(CTCTAGAG) there is cross-strand adenine–adenine stacking at the central step that is able to direct the phosphate backbone into an unusual trans-trans conformation. We note that not all pyrimidine–purine steps have this distortion of the backbone, but it is clearly observed in the context of a CTAG segment. The TpA (=TpA) step displays unusual thermodynamic properties (Travers & Klug, 1987; Gotoh & Tagashira, 1981), with TpA (=TpA) being less stable than ApT (=ApT). The relative stability may be explained by the different stacking interactions observed crystallographically (Klug et al., 1979; Calladine & Drew, 1984), where the overlap of bases is greater for the ApT (=ApT) step than for the TpA (=TpA) step. The reduced thermal stability of TpA (=TpA) may account for the pronounced rate of transient base-pair opening observed by NMR studies (Patel et al., 1982). The TpA (=TpA) step features

in a number of protein–DNA interactions involving nucleases and repressors.

As a prelude to discussing the details of these protein–DNA interactions, it is important to mention briefly some previous work. In this context the terminology of Helene and Lancelot (1983) is appropriate and will be used. They have defined "selective recognition" as the process whereby proteins bind specifically to a defined sequence. Original theories of selective recognition concentrated on hydrogen-bonding interactions between the proteins and the bases (von Hippel & McGhee, 1972; Seeman et al., 1976; von Hippel & Berg, 1986). Whereas this facet of protein–DNA recognition is clearly of great importance, it is not the sole determinant of selectivity.

Solution studies (Drew, 1984; Drew & Travers, 1984) suggest that the sequence-dependent nuclease cleavage of the *Escherichia coli tyr*T promoter is determined by the structural polymorphism of DNA. Crystallographic studies of oligonucleotide fragments provided details enabling their results to be rationalized in terms of sequence-dependent variations in the groove widths of the double helix and local variations in phosphate accessibility. The crystal structure of an octanucleotide complex with DNase I subsequently confirmed much of Drew and Travers conclusions (Suck et al., 1988). Micrococcal and S1 nucleases are able to recognize and cut TA, TATA, and CTAG sites preferentially (Dingwall et al., 1981; Drew & McCall, 1987; Flick et al., 1986). It is noteworthy that cleavage at the CTAG site is very prominent in this respect. Enhanced cleavage generally occurs in regions of low helix stability and where an exposed accessible phosphate backbone, able to fit into the enzyme active site, is available. The results of our study on d(CTCTAGAG) have shown that an unusual trans-trans configuration is induced in the phosphate backbone at the CTAG site. This structural feature may be an important recognition signal between the DNA and certain selective nucleases. The previously discussed thermodynamic properties of the TpA (=TpA) step could then make a contribution to the kinetic specificity of the enzyme mechanism.

A similar backbone conformation can also be induced by CpG steps. This type of pyrimidine–purine step is also involved in specific interactions with certain nucleases, DNase I and *Msp*I. This conformation may represent a similar recognition signal for these other enzymes. The difference in functional groups and thermodynamic properties may then be responsible for discrimination between CpG- and TpA-containing sequences.

Repressor proteins interact with DNA by binding to the operator sequences in a highly selective manner. Studies on mutant repressors show that certain contacts between the protein and operator phosphate backbone are able to contribute energy and specificity of the interaction (Nelson & Sauer, 1986). Crystallographic and solution studies of the 434 repressor/operator complex have indicated that protein–phosphate interactions are essential for repressor activity (Anderson et al., 1985; Bushman et al., 1985). Of particular interest is the recently reported structure of the *E. coli trp* repressor/operator complex fully refined to a resolution of 2.4 Å (Otwinowski et al., 1988). The operator sequence is TGTAC-TAGTTAACTAGTAC. The two CTAG segments constitute an important part of the operator in the recognition by *trp* repressor (Bass et al., 1987). The CTAG sequence is also the rarest tetranucleotide sequence in *E. coli* (Otwinowski, personal communication). The accurately determined complex structure shows that there are no direct hydrogen-bonding or nonpolar interactions to the bases, although the repressor recognizes and binds at this site. Otwinowski et al. (1988) have postulated that selective recognition is accomplished by the sequence-directed influence on the phosphate backbone. It is interesting to note that the important protein–phosphate contacts involve the CTAG segments of the operator and that, as noted by the authors, the conformation of this section has partial A-form character. The minor groove of the operator is widest at the section opposite the helix–turn–helix section of the protein that is directed into the major groove. Similar minor-groove width is seen in the CTAG section of our octamer structure. The CpT (=ApG) and TpA (=TpA) steps in the operator show low twist and high slide values and the TpA (=TpA) steps in the protein–DNA complex and in the octamer structure have similar purine–purine cross-strand overlap. This cross-strand overlap may induce a locally stable conformation of the CTAG site having the unusual phosphate backbone observed in the crystal structure of d(CTCTAGAG) and may represent a specific shape that the *trp* repressor is best able to recognize and bind. In the protein–DNA complex we observe the results after binding. Although the extended backbone conformation is no longer present in the complex, this can be explained by the structural rearrangements that have to occur to form a stable interface between the protein and the DNA. The thermodynamic properties of the TpA (=TpA) step could then make a contribution to the production of a stable, rigid protein–DNA complex by providing little resistance to the necessary structural rearrangements.

Although full structural details are not yet available, it is tempting to conclude that the TpA (=TpA) step may be a more general recognition feature. In the 434 repressor/operator complex, for example, where the DNA sequence is TACAATATATAATTGTA, the critical phosphate group described above is 5′ to the TpA sequence underlined. A stronger argument can be put forward for the *met* repressor from *E. coli*. This protein selectively recognizes operators that consist of between two and five copies of an 8-base-pair repeating sequence (Belfaiza et al., 1986). On examining this sequence, we have noticed that such repeats result in a number of the rare CTAG sites separated by four nucleotides very like that observed with the *trp* repressor. For example, the operator sequence AGACGTCTAGACGTCTAGACGTCT with two of the rare CTAG sites can be constructed from three repeats of the *met* repressor consensus sequence. The similarity extends beyond the DNA sequence itself. In the *trp* repressor/operator complex there are direct interactions between threonine 81 and a phosphate at the CTAG site. According to the sequence alignment by Rafferty et al. (1988), the *met* repressor would also have a threonine in the same position on the second helix of the helix–turn–helix motif. The results of a crystallographic analysis in progress (Rafferty et al., 1988) will provide more information about protein–DNA interactions and a test of selective recognition involving induced backbone changes discussed in this paper.

ACKNOWLEDGMENTS

We thank Prof. R. E. Dickerson and Dr. E. Westhof for providing indispensible software; Dr. A. Wang for coordinates; Dr. A. Lesk for figures; Prof. P. Sigler, Z. Otwinowski, and Dr. B. Luisi for details of the *trp*/operator complex prior to publication; L. Fairall and Dr. D. Rhodes for communicating the results of their CD study also in advance of publication; Drs. M. Moore, H. Nelson, H. Drew, W. Cruse, and especially M. McCall for comments and encouragement. We also thank S. Anelli for her patience in the preparation of the manuscript and the referees for well-informed and useful criticisms.

Registry No. d(CTCTAGAG), 118419-00-4.

REFERENCES

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205–8212.

Anderson, J. E., Ptashne, M., & Harrison, S. E. (1985) *Nature 316*, 596–601.

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504–1509.

Bass, S., Sugiono, P., Arvidson, D. N., Gunsalus, R. P., & Yonderain, P. (1987) *Genes Dev. 1*, 565–572.

Belfaiza, J., Parsot, C., Martel, A., Bonthier de la Tour, C., Margarita, D., Cohen, G. N., & Saint-Girons, I. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 867–871.

Blundell, T. L., & Johnson, L. N. (1976) *Protein Crystallography*, Academic Press, London.

Bushman, F. D., Anderson, J. E., Harrison, S. C., & Ptashne, M. (1985) *Nature 316*, 651–653.

Calladine, C. R., & Drew, H. R. (1984) *J. Mol. Biol. 178*, 773–782.

Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry*, Freeman, San Francisco.

Conner, B. N., Yoon, C., Dickerson, J. L., & Dickerson, R. E. (1984) *J. Mol. Biol. 174*, 663–695.

Dickerson, R. E. (1983) *J. Mol. Biol. 174*, 419–441.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.

Dingwall, C., Lommonossof, G. P., & Laskey, R. (1981) *Nucleic Acids Res. 9*, 2659–2673.

Drew, H. R. (1984) *J. Mol. Biol. 176*, 535–557.

Drew, H. R., & Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535–556.

Drew, H. R., & Travers, A. A. (1984) *Cell 37*, 491–502.

Drew, H. R., & McCall, M. J. (1987) *J. Mol. Biol. 197*, 485–511.

Flick, J. T., Eissenberg, J. C., & Elgin, S. C. R. (1986) *J. Mol. Biol. 190*, 619–633.

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 686–707.

Gotoh, O., & Tagashira, Y. (1981) *Biopolymers 20*, 1033–1042.

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

Heinemann, U., Lauble, H., Frank, R., & Blocker, H. (1987) *Nucleic Acids Res. 22*, 9531–9550.

Helene, C., & Lancelot, G. (1983) *Prog. Biophys. Mol. Biol. 39*, 1–68.

Hendrickson, W. A., & Konnert, J. H. (1981) *Biomolecular Structure and Conformation, Function and Evolution* (Srinivasen, R., Ed.) pp 43–57, Pergamon Press, Oxford, U.K.

Jain, S., Zon, G., & Sundaralingam, M. (1987) *J. Mol. Biol. 197*, 141–145.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z., & Steitz, T. A. (1979) *J. Mol. Biol. 131*, 669–680.

Leslie, A. G. W., Arnott, S., Chandrasekaran, R., & Ratliff, R. L. (1980) *J. Mol. Biol. 143*, 49–72.

Levitt, M., & Warshel, A. (1978) *J. Am. Chem. Soc. 100*, 2607–2613.

Luzzati, U. (1952) *Acta Crystallogr. 5*, 802–810.

McCall, M., Brown, T., & Kennard, O. (1985) *J. Mol. Biol. 183*, 385–396.

McCall, M., Brown, T., Hunter, W. N., & Kennard, O. (1986) *Nature 322*, 661–664.

McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, Wiley, San Francisco.

Nelson, H. C. M., & Sauer, R. (1986) *J. Mol. Biol. 192*, 27–38.

Ollis, O. L., & White, S. W. (1987) *Chem. Rev. 87*, 981–995.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321–329.

Patel, D. J., Kozlowski, S. A., Ikuta, S., Itakura, K., Bhatt, R., & Hare, D. R. (1982) *Cold Spring Harbor Symp. Quant. Biol. 47*, 197–206.

Rabinovich, D., Haran, T., Eisenstein, M., & Shakked, Z. (1988) *J. Mol. Biol. 200*, 151–161.

Rafferty, J. B., Phillips, S. E. V., Rojas, C., Boulot, G., Saint-Girons, I., Guillon, Y., & Cohen, G. N. (1988) *J. Mol. Biol. 200*, 217–219.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Saenger, W., Hunter, W. N., & Kennard, O. (1986) *Nature 324*, 385–388.

Seeman, W. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804–808.

Shakked, Z., & Rabinovich, D. (1986) *Prog. Biophys. Mol. Biol. 47*, 159–195.

Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W. B. T., Salisbury, A., & Viswamitra, M. A. (1983) *J. Mol. Biol. 166*, 183–201.

Sheldrick, G. M. (1976) *SHELX* 76 System of Computing Programs, University of Cambridge, U.K.

Suck, D., Lahm, A., & Oefner, C. (1988) *Nature 332*, 464–468.

Travers, A. A., & Klug, A. (1987) *Philos. Trans. R. Soc. London, B 317*, 537–561.

von Hippel, P. H., & McGhee, J. D. (1972) *Annu. Rev. Biochem. 41*, 231–300.

von Hippel, P. H., & Berg, O. G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1608–1612.

Wang, A. H.-J., Quigly, G. J., Kolpak, F. J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1980) *Science 211*, 171–176.

Wang, A. H.-J., Fujii, S., van Boom, J. H., & Rich, A. (1982a) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3968–3972.

Wang, A. H.-J., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A. A., & Rich, A. (1982b) *Nature 299*, 601–604.

Westhof, E. (1987) *Int. J. Biol. Macromol. 9*, 185–192.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.