The EMBO Journal vol.11 no.5 pp.1931 - 1939, 1992

Double helix conformation, groove dimensions and ligand binding potential of a G/C stretch in B-DNA

Udo Heinemann, Claudia Alings and Manju Bansal[1]

Institut für Kristallographie, Freie Universität Berlin, Takustraße 6, D-1000 Berlin 33, FRG

[1]On leave from the Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560 012, India

Communicated by W.Saenger

The self-complementary DNA fragment CCGGCGCCGG crystallizes in the rhombohedral space group R3 with unit cell parameters a = 54.07 Å and c = 44.59 Å. The structure has been determined by X-ray diffraction methods at 2.2 Å resolution and refined to an R value of 16.7%. In the crystal, the decamer forms B-DNA double helices with characteristic groove dimensions: compared with B-DNA of random sequence, the minor groove is wide and deep and the major groove is rather shallow. Local base pair geometries and stacking patterns are within the range commonly observed in B-DNA crystal structures. The duplex bears no resemblance to A-form DNA as might have been expected for a sequence with only GC base pairs. The shallow major groove permits an unusual crystal packing pattern with several direct intermolecular hydrogen bonds between phosphate oxygens and cytosine amino groups. In addition, decameric duplexes form quasi-infinite double helices in the crystal by end-to-end stacking. The groove geometries and accessibilities of this molecule as observed in the crystal may be important for the mode of binding of both proteins and drug molecules to G/C stretches in DNA.

Key words: B-DNA/conformation/G/C-rich/protein–DNA interaction/X-ray crystallography

Introduction

It is often assumed (e.g. Saenger, 1984; Palecek, 1991) that DNA molecules with non-random sequences adopt conformations which deviate significantly from canonical B-DNA as known from fibre diffraction analysis (Chandrasekaran and Arnott, 1989). Under the influence of superhelical tension, appropriate sequence signals may cause the formation of left-handed Z-DNA (Rich *et al.*, 1984), of cruciforms (Lilley, 1980) or of triple-stranded structures (Mirkin *et al.*, 1987; Htun and Dahlberg, 1989). As the superhelical tension varies in cells as a consequence of DNA transcription (Liu and Wang, 1987), such unusual structures may appear transiently and therefore may have functional significance.

Even in the absence of superhelical tension, the geometry of the DNA double helix appears to be modulated by the underlying nucleotide sequence. Proteins that are not base pair specific such as deoxyribonuclease (DNase) I or chemical nucleases, cleave DNA sequences at different rates depending on groove dimensions (Lomonossoff *et al.*, 1981; Drew and Travers, 1984). Drug molecules such as netropsin prefer narrow minor grooves for binding (Kopka *et al.*, 1985) which also show a characteristic hydration pattern (Drew and Dickerson, 1981; Chuprina *et al.*, 1991).

Studies by X-ray crystallography have shown that runs of A/T bases lead to a narrow minor groove often accompanied by high propeller twisting and bifurcated hydrogen bonding (Dickerson and Drew, 1981; Coll *et al.*, 1987; Nelson *et al.*, 1987; DiGabriele *et al.*, 1989). In addition, such sequences may be associated with a deflection of the double helix axis (Nadeau and Crothers, 1989). In contrast, little crystallographic evidence is available for long G/C stretches in B-DNA. The longest pure G/C oligonucleotide investigated was a hexamer (Cruse *et al.*, 1986), and in longer fragments crystallized, AT base pairs were interspersed in G/C sequences (Heinemann and Alings, 1989, 1991). In the past, G/C-rich DNA fragments showed a tendency to crystallize in the A-form (Shakked and Rabinovich, 1986; Kennard and Hunter, 1989). However, the solution structure of these oligonucleotides was proposed to be B-like based on NMR (Reid *et al.*, 1983; Rinkel *et al.*, 1986; Keniry *et al.*, 1987; Wolk *et al.*, 1989) and Raman (Benevides *et al.*, 1986; Peticolas *et al.*, 1988) spectroscopy or to be intermediate between the A- and B-forms based on circular dichroism measurements (Fairall *et al.*, 1989; Galat, 1990). These contradictory results have conferred the notion that in biological systems, G/C stretches in DNA either may have A-DNA-like properties or may undergo a transition to the A-form easily (McCall *et al.*, 1986; Rhodes and Klug, 1986; Drew and Travers, 1984).

The self-complementary DNA decamer CCGGCGCCGG (hereafter named the G/C decamer) for which a crystallographic analysis is reported here crystallizes as B-DNA double helix with usual Watson–Crick base pairing. Comparison of this duplex with the decamer CCAACGTTCC (Privé *et al.*, 1991) shows a significant change in both groove dimensions and base pair stacking geometry upon transition of AT base pairs to GC. The dodecamer duplex ACCGGCGCACA (Timsit *et al.*, 1989, 1991) has an octanucleotide sequence in common with the G/C decamer. A comparison of these two molecules demonstrates the influence of environment and flanking sequences on double helix conformation.

Results and Discussion

Refinement results

The structure refinement converged at an R factor of 16.7% with root-mean-square (r.m.s.) positional and thermal shifts in the last refinement cycle of 0.007 Å and 0.13 Å^2, respectively. The final structural model comprised 404 DNA atoms and 47 solvent molecules treated as water oxygens with unit occupancies. A difference Fourier map calculated

with this model had r.m.s. and maximum densities of 0.066 and 0.235 e/Å^3 respectively. Details concerning the crystallographic data and refinement parameters may be obtained from Table I. A representative portion of the electron density map is displayed in Figure 1. The structure amplitudes and atomic coordinates of the G/C decamer have been submitted to the Brookhaven Crystallographic Database.

The two strands of the G/C decamer form a B-type antiparallel double helix with Watson–Crick base pairing which is enforced by restraints during refinement. This is common practice in nucleic acid refinement (Westhof *et al.*, 1985) and has been applied in all structure analyses performed in our laboratory (Heinemann *et al.*, 1987, 1991; Lauble *et al.*, 1988; Heinemann and Alings, 1989, 1991).

Table I. Refinement statistics

Resolution range (Å)	8.0–2.2
Number of observations (1σ on F_o)	2019
(3σ on F_o)	1349
R value (1σ data) (%)	16.7
(3σ data) (%)	14.4
F_o/F_c correlation coefficient	0.957
$\Sigma F_o/\Sigma F_c$	1.009
$<\|F_o - F_c\|>$	80.0
Sugar–base bond distances (Å)	0.013/0.025
Sugar–base bond angle distances (Å)	0.031/0.040
Phosphate bond distances (Å)	0.043/0.040
Phosphate bond angle distances and H-bond distances (Å)	0.053/0.060
Planar groups (Å)	0.019/0.030
Chiral volumes (Å^3)	0.069/0.100
Single torsion non-bonded contacts (Å)	0.144/0.250
Multiple torsion non-bonded contacts (Å)	0.225/0.250
Isotropic thermal factors (Å^2)	
Sugar–base bonds	2.195/3.00
Sugar–base bond angles	3.268/5.00
Phosphate bonds	3.393/5.00
Phosphate bond angles, H-bonds	3.131/5.00
Structure amplitudes	56 – 120(sin θ/λ – 1/6)

F_o and F_c are observed and calculated structure amplitudes, respectively. The R value is $\Sigma|F_o - F_c|/\Sigma F_o$, and the correlation coefficient is $\Sigma[F_o - <F_o>)(F_c - <F_c>)] / [\Sigma(F_o - <F_o>)^2\Sigma(F_c - <F_c>)^2]^{1/2}$. For stereochemical parameters (bottom part), the left number gives the r.m.s. deviations from ideal values in the final model, and the right number is the target variance used in refinement. The weight applied to the corresponding restraint is the inverse square of the target variance.

Although the duplex molecule has a symmetric sequence, the two strands are not related by crystallographic symmetry. They can be superimposed onto each other with a r.m.s. distance between equivalent atoms of 0.76 Å. Taking into account the mean error in the atomic coordinates of the G/C decamer of 0.2 Å (according to Luzzati, 1952), this indicates a moderate but clearly recognizable influence of crystal packing on the DNA conformation.

In the following, the plus strand is numbered 1–10 and the minus strand is numbered 11–20. Conformational analyses have used the programs NEWHEL91 (Fratini *et al.*, 1982; Dickerson, 1985) and NUPARM (Bhattacharyya and Bansal, 1989). Structural parameters given conform to the 1988 Cambridge conventions (Dickerson *et al.*, 1989).

Global DNA structure

In the R3 lattice, G/C decamers pack end-to-end forming long B-type double helices that penetrate the crystal completely (see also below). This is accompanied by an exact 10 bp repeat and a straight helix. The base pair step in between decamer duplexes has normal geometry despite the missing phosphodiester link (Figure 2).

In the G/C decamer helix, characteristic groove dimensions are observed. The minor groove is uniformly wide compared with the fibre model of B-DNA (Chandrasekaran and Arnott, 1989) and with most B-DNA structures determined by X-ray crystallography (Dickerson, 1990; Privé *et al.*, 1991). There is, however, clear sequence-dependent variation in groove widths (Figure 3). The wide minor groove is caused predominantly by a sliding motion of the base pairs along their long axes. The major groove is on average narrower than that of canonical B-DNA.

The minor groove of the G/C decamer is not only unusually wide but also unusually deep (Table II). A convenient measure of groove depth is the displacement of base pairs from the best straight helix axis. The GC base pairs of the decamer are displaced from the axis which is pushed towards the minor groove edge of the base pairs. As a consequence, the minor groove appears deeper and the major groove appears shallower. This behaviour is characteristic of major grooves devoid of methyl groups. AT-rich sequences have deeper major grooves, possibly to accommodate the thymidine methyl groups. The influence of major groove methylation on groove depth has been demonstrated directly with the B-DNA decamer CCAG-

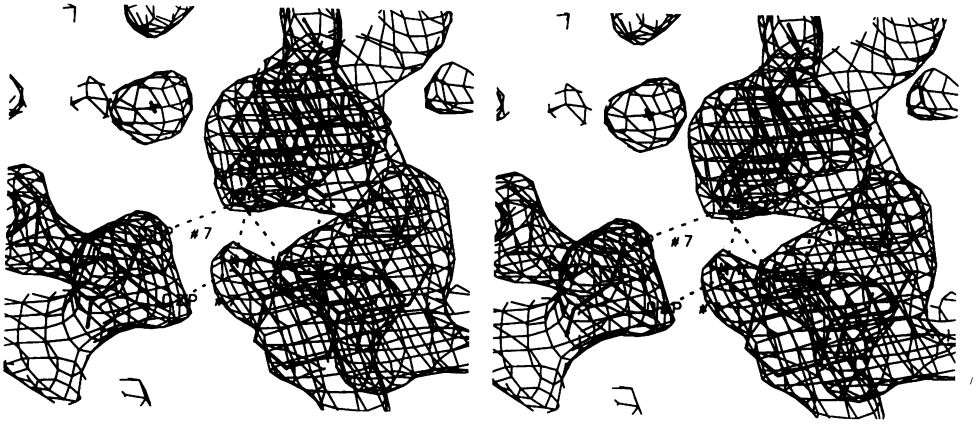

Fig. 1. Portion of the $(2F_o-F_c)$ electron density map around residues C(5) and C(15). Note the possible intermolecular hydrogen bonds between $C(5)N_4$ and $C(7)O_{1P}$ and between $C(15)N_4$ and $C(7)O_{2P}$ indicated by dashed lines.

GCCTGG and a single-site methylated variant (Heinemann and Hahn, 1992).

By X-ray fibre diffraction methods, Arnott *et al.* (1986) have shown that the hybrid duplex poly(dA)·poly(rU) has a narrow, deep major groove. This appears to contradict the above observation relating to groove dimensions in AT-rich sequences since uridine differs from thymine by the absence of the methyl group at position 5. The hybrid duplex, however, is a variant of A-DNA for which the influence of methylation on groove dimensions may be absent or different.

In the A-form of DNA, the base pairs are displaced by 4–5 Å from the helix axis towards the minor groove (Shakked and Rabinovich, 1986; Chandrasekaran and Arnott, 1989; Verdaguer *et al.*, 1991). Consequently, the major groove becomes very deep and the minor groove is shallow. The displacement observed in the G/C decamer is in the opposite direction to that found in A-DNA. Thus, the G/C stretch of DNA is not intermediate between A-DNA and B-DNA. In fact, intermediate forms have never been observed in crystal structures of isolated DNA (Heinemann, 1991). The only occurrence of an intermediate species, i.e. a DNA molecule with groove dimensions between A-DNA and B-DNA, has been in a sequence-specific complex with protein (Pavletich and Pabo, 1991).

Crystal packing

Unusually intimate contacts between DNA helices are present in the rhombohedral space group R3. The end-to-end stacking of duplexes as one type of intermolecular interaction has already been described (see Figures 2 and 3). Equally important are contacts between the sugar–phosphate backbone and major groove atoms of the DNA. The 3_2 screw axes passing through the unit cell promote the crossing of helices over each other at an angle of 120° (Figure 4). The length of the *c*-axis (44.59 Å) determines the close

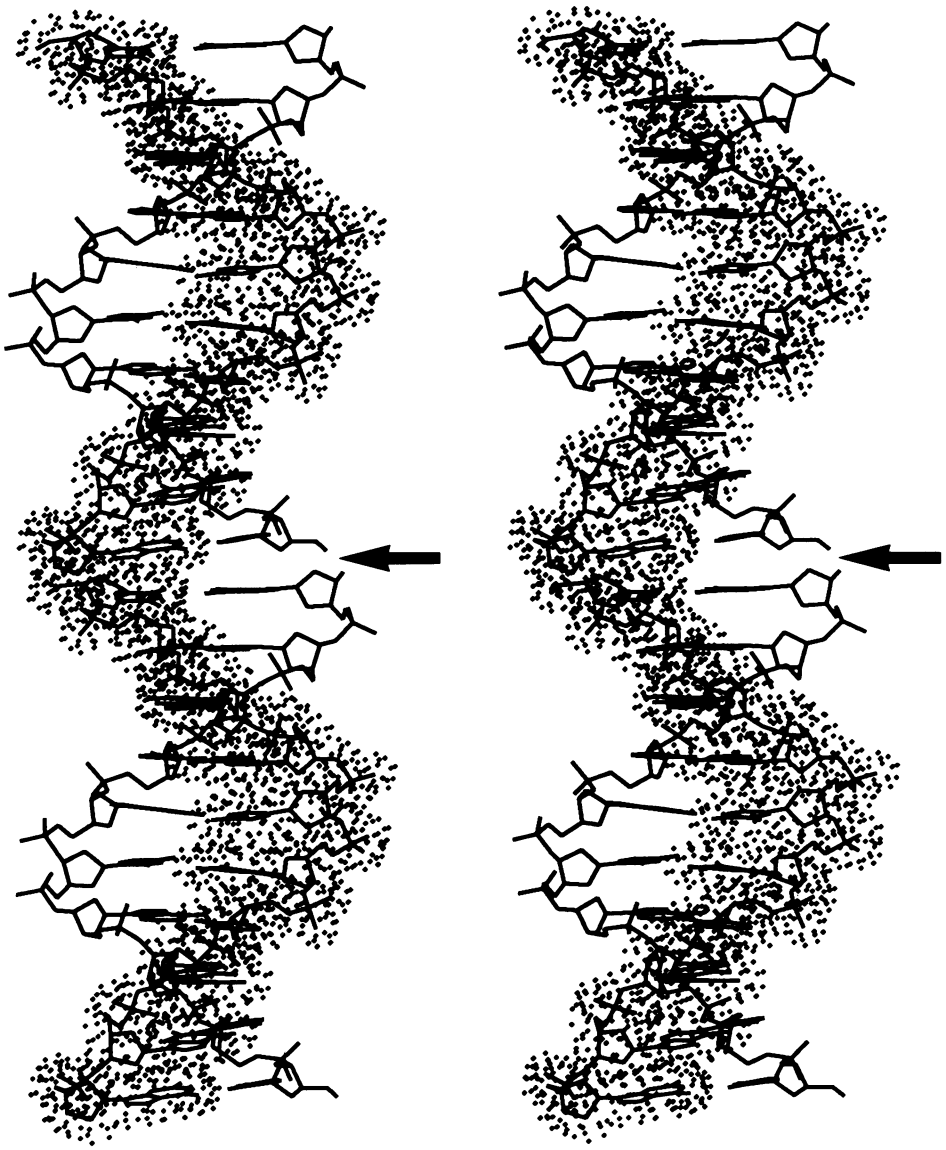

Fig. 2. Stereo view of two G/C decamer duplexes stacked end-to-end as in the crystal. To give an impression of the molecular surface, the van der Waals radii of the atoms in one strand are indicated by dots. From top to bottom there is an alternation of minor and major grooves. Note the nearly equivalent groove widths and the perfect stacking of 10 bp duplexes in the centre. The arrow points to the break in the sugar–phosphate backbone between decamer helices.

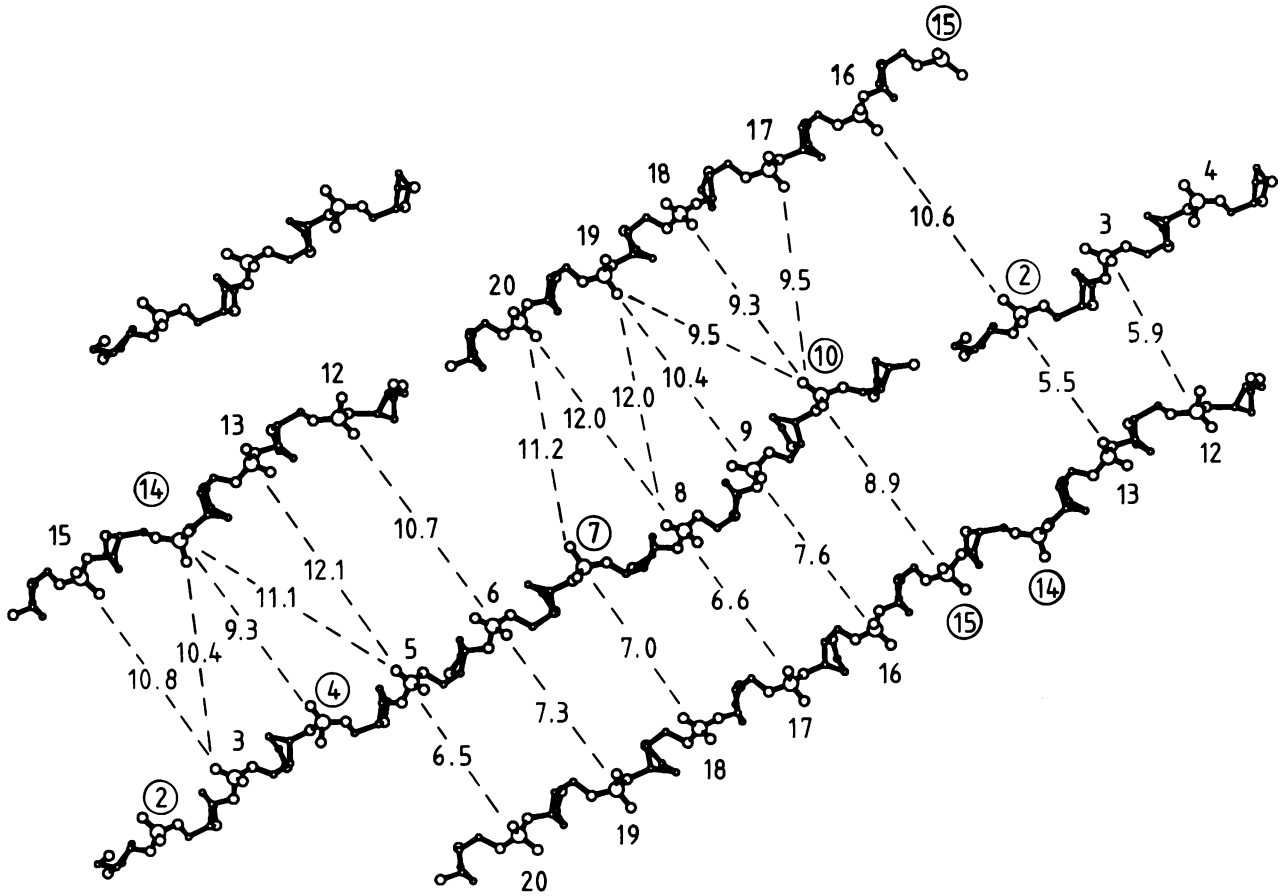

Fig. 3. Groove widths in the B-DNA decamer CCGGCGCCGG. The sugar–phosphate backbone is shown in an unrolled helix diagram, and groove widths are indicated as shortest P–P separation across the grooves minus 5.8 Å to account for the van der Waals radii of the phosphate groups. Since the decamer duplexes stack end-to-end in the crystal, the grooves continue across the gap as shown. The sequence numbers of nucleotides showing backbone torsion angles outside the normally observed ranges (Saenger, 1984) are circled.

packing: a side-by-side packing of three helical cylinders with diameters of 18.5 Å (see Table II) would require 55.5 Å. Due to the observed close contacts between helices which bring the phosphate oxygens close to functional groups in the major groove, this characteristic packing parameter reduces by nearly 11 Å.

The type of packing described above for the G/C decamer has also been observed with sequence-related dodecamer helices (Timsit *et al.*, 1989, 1991). It has been suggested (Timsit *et al.*, 1989) that the interactions between helices seen in crystals may be a model for DNA contacts occurring during recombination. We note that the topology mimicked by crystal packing is that of a right-handed cross between helix segments as proposed for the four-way junction (Murchie *et al.*, 1989; Cooper and Hagerman, 1989; Duckett *et al.*, 1990). Preliminary model building (M.Bansal, unpublished) shows that strand exchange may in fact be possible without much change in the arrangement of molecules.

The cross-helix packing in combination with the shallow major groove permits a number of direct hydrogen bonds to be formed between phosphate oxygens and cytosine amino groups of the G/C decamer (Table III). The stabilization of B-DNA crystals by direct hydrogen bonding between phosphates and functional groups in the grooves is quite exceptional. It testifies to an unusual accessibility of the bases in the major groove which might be of functional importance for similar G/C stretches of DNA.

Table II. Groove dimensions in B-DNA

Sequence	Parameters			Minor groove		Major groove	
	Slide (Å)	*Disp.* (Å)	Radius (Å)	Width (Å)	Depth (Å)	Width (Å)	Depth (Å)
GGCGCC	0.66	0.76	9.26	6.8	8.7	10.7	8.2
AGGCCT	0.68	0.89	9.24	5.2	8.8	11.6	8.0
AGGCm^{5}CT	0.13	−0.32	9.54	4.2	7.9	12.5	9.5
AACGTT	0.35	−0.08	9.50	4.4	8.1	12.9	9.3
GAATTC	−0.28	−0.57	9.56	4.1	7.7	11.6	9.8
B-DNA fibre	0.56	0.57	9.22	5.9	8.5	11.4	8.3
A-DNA fibre	−1.22	−4.38	8.59	11.2	2.9	2.2	12.7

To avoid possible end effects, the parameters are based on the central six base pairs of high resolution oligonucleotide crystal structures; GGCGCC, this work; AGGCCT, Heinemann and Alings (1989), Protein Data Bank entry 1BD1; AGGCm^{5}CT, Heinemann and Alings (1991), 1D25; AACGTT, Privé *et al.* (1991), 5DNB; GAATTC, Dickerson and Drew (1981), 1BNA; the fibre A- and B-DNA models are from Chandrasekaran and Arnott (1989). Average helical parameters for the six base pairs were calculated with NUPARM (Bhattacharyya and Bansal, 1989) after fitting a best axis through the 10 or 12 bp helix using the C1′ atoms. Groove widths are based on phosphorus–phosphorus separations as shown in Figure 3. The groove depths are defined as Depth (minor) = *Radius* + *Disp.* −1.3; and Depth (major) = *Radius* − *Disp.* − 0.3, where *Radius* is the mean distance of the phosphorus atoms with respect to the helix axis and the constant terms depend on the differences between the van der Waals radii of the phosphate group (2.9 Å) and a dummy group of 1.5 Å radius placed at the mean positions of the functional groups in the two grooves.

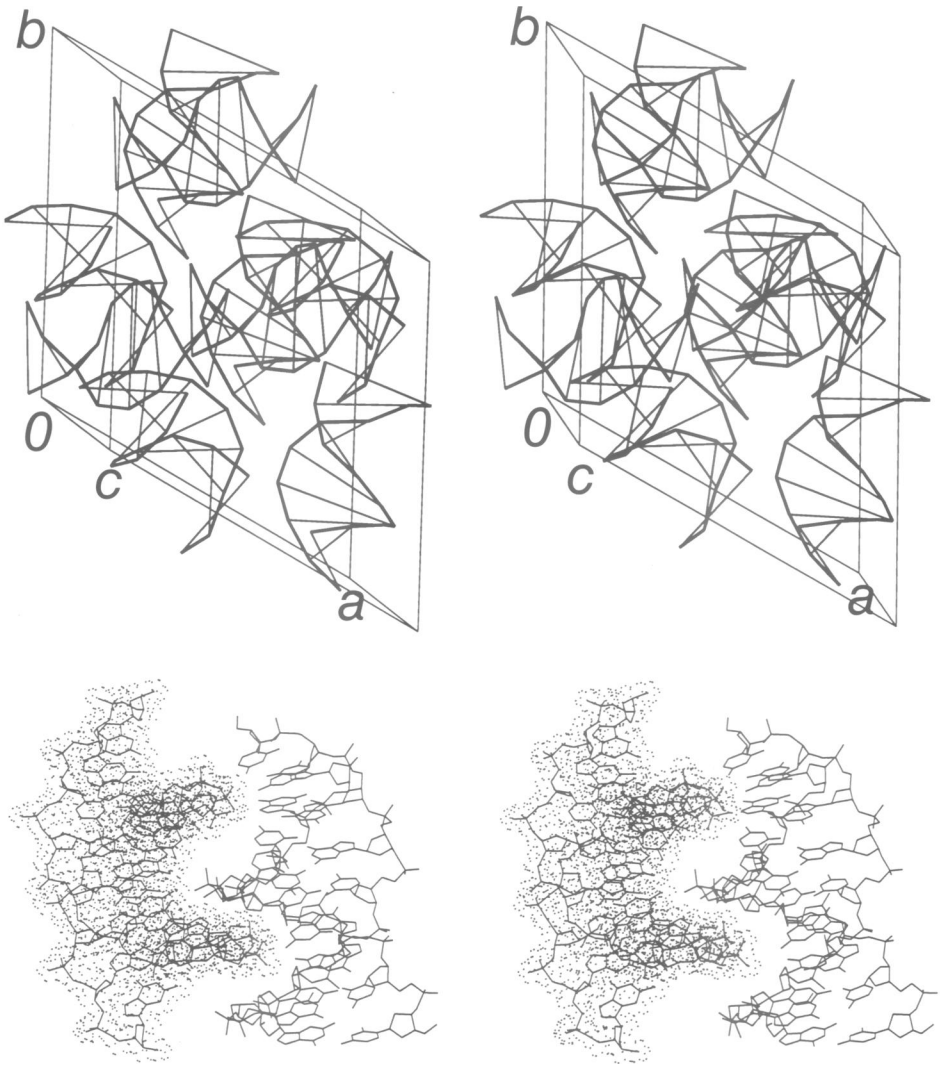

Fig. 4. Crystal packing in the R3 unit cell of CCGGCGCCGG. (**a**) Stereo diagram of the unit cell looking approximately along the *c*-axis. G/C decamer duplexes are depicted schematically. Their phosphorus positions (nine for each strand) are connected by bold lines along the strands and by thinner lines across the strands. Note the close packing of molecules related by 3-fold screw axes. (**b**) Stereo diagram showing the interactions of two decamers in detail. The molecule on the left has been given a van der Waals surface in a dot representation.

It might be asked whether the shallow major groove of the G/C decamer is an inherent property of the duplex or whether it might not result from crystal packing constraints. Since similar groove dimensions have been observed for a dodecameric duplex of related sequence and since a very different geometry is present in a decamer with identical purine–pyrimidine pattern but different sequence (see below), it appears that it is in fact the sequence and not external factors that exerts the prime influence on the groove dimensions in B-DNA. Clearly, this claim needs to be substantiated by further work on related DNA fragments.

Sequence-related structure

Despite its unusual groove dimensions, the G/C decamer shows base pair geometries and stacking patterns that are within the range normally observed in B-DNA crystal structures (Kennard and Hunter, 1989; Dickerson, 1990). Some important helical parameters are plotted in Figure 5 and compared with those characterizing the sequence-related decamer CCAACGTTGG (Privé *et al.*, 1991). In both cases, the stereochemistry of the base pair step in between individual decamer helices does not deviate strongly from intrahelical stacks. Although both decamers are typical variants of B-DNA, their local stacking patterns may deviate significantly from those of canonical B-DNA. This is exemplified by the wide range of values adopted by the local helical parameters *twist* and *slide*.

The considerably asymmetric distribution of parameter values of the G/C decamer suggests a non-negligible influence of crystal packing on double helix conformation. In contrast, CCAACGTTGG crystallizes in space group C2 and shows identical conformation of both strands (or of both ends of the double helix).

Effect of base pair transitions: comparison with CCAACGTTGG

The DNA fragment CCAACGTTGG differs in sequence from the G/C decamer only by the exchange of its four AT base pairs with GC base pairs. Its three-dimensional structure is known at high resolution from the X-ray diffraction analysis of non-isomorphous crystals (Privé *et al.*, 1991). A comparison of both structures may help to understand the effects of transition mutations on DNA conformation.

The two decamer helices may be superimposed onto one

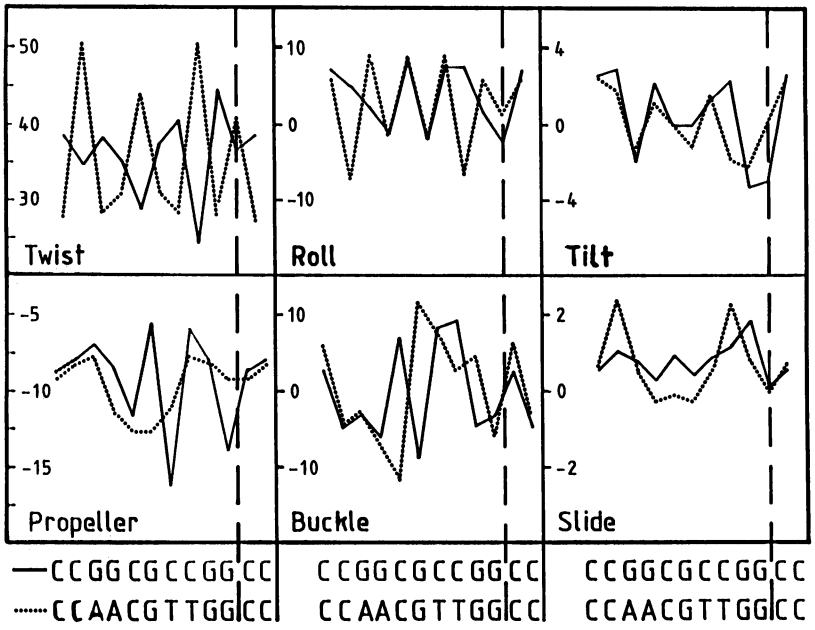

Fig. 5. Helical parameters for the G/C decamer and for the related DNA duplex CCAACGTTGG (Privé *et al.*, 1991) calculated with NUPARM (Bhattacharyya and Bansal, 1989). In both crystal structures, the base pair step in between decamer helices has normal geometry. Note the striking difference in base pair stacking between the C–A/T–G and the equivalent C–G/C–G steps.

Fig. 6. Least-squares superimposition of the G/C decamer and the CCAACGTTGG duplex (Privé *et al.*, 1991). The view is into the minor groove of the double helices which is clearly wider in the G/C decamer (heavy lines).

another with a Δ r.m.s. between their C1′ positions of 1.19 Å. From Figure 6 it is evident that the global structures of both molecules are quite different: the minor groove of the G/C decamer is considerably wider than that of the helix containing AT base pairs. Even more significant differences are present in local base pair stacking (see Figure 5). The C–A and T–G base pair steps in CCAACGTTGG have a very unusual geometry with very high *twist* and *slide* coupled with negative *roll* which is also found in related sequences (Privé *et al.*, 1987; Heinemann and Alings, 1989, 1991), but which is not the only conformation possible for this dinucleotide step (Yamagi *et al.*, 1991). The C–G steps at the same sequence positions in the G/C decamer have just the opposite characteristics of *twist* and *roll*. Thus, base pair transitions may have profound effects on both global and local structural features of DNA helices although purine–pyrimidine patterns remain unchanged. Consequently, the analysis and prediction of nucleic acid conformation has to

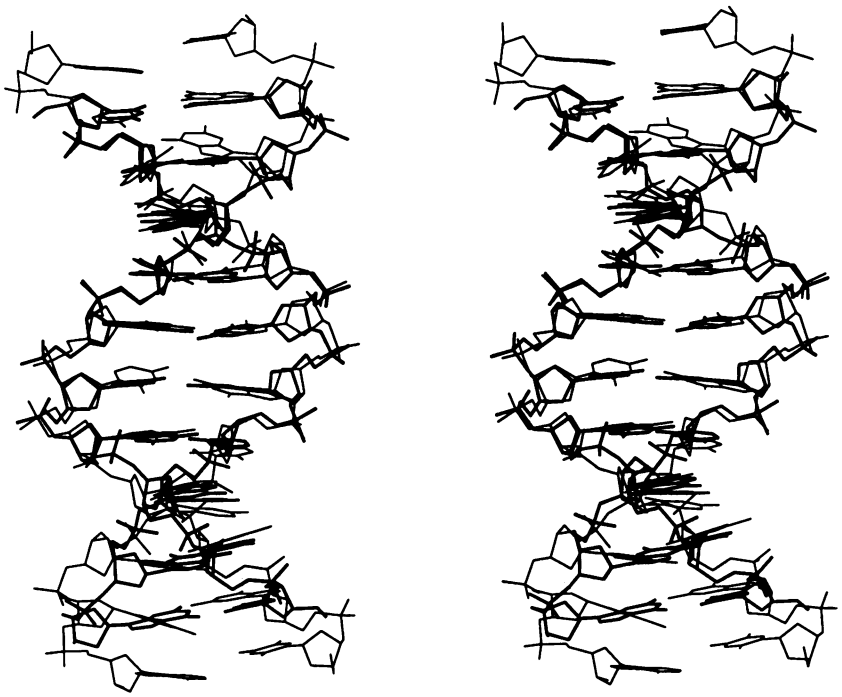

Fig. 7. Comparison of the crystal structures of CCGGCGCCGG and of ACCGGCGCCACA (Timsit *et al.*, 1989, 1991). The C1′ atoms of residues 1–8 and 13–20 of the G/C decamer (heavy lines) were superimposed in a least-squares procedure on to the equivalent C1′ positions of residues 2–9 and 16–23 of the dodecamer with an Δ r.m.s. of 0.73 Å. The view is into the minor groove of the helices which has similar dimensions in spite of different local conformations.

consider individual bases and must not rely on purine–pyrimidine patterns as initially thought possible (Calladine, 1982; Dickerson, 1983).

Effect of sequence context: comparison with ACCGGCGCCACA

The G/C decamer shares with the dodecamer ACCGGCGCCACA the underlined octamer sequence. The two molecules show similar crystal packing (see above) with the exception that the dodecamer has to have an inter-helical base pair step with negative *twist* value in order to form continuous helices (Timsit *et al.*, 1989, 1990). A superimposition of the two molecules (Figure 7) shows quite similar groove dimensions in the sequence region common to them. Base pair stacking geometries are also similar in the centre of the common region, but display considerable differences towards the ends. In the dodecamer, normal Watson–Crick base pairing is not uniformly present. Instead, through part of the duplex unusually large *propeller twist* and base pair *buckle* lead to the disruption of Watson–Crick hydrogen bonding and to the formation of new hydrogen bonds with neighbouring bases from the opposing strand (Timsit *et al.*, 1991). This trend continues from the CACA motif into the conserved octamer region.

Four factors may contribute to the phenomenon that an identical octamer sequence may adopt different local conformations in a different sequence context and in different (although nearly isomorphous) crystals. (i) The dodecamer structure was refined without the usual restraints that enforce normal Watson–Crick hydrogen bonding (Timsit *et al.*, 1991). Omitting part of the normally included stereochemical information from the refinement against limited resolution X-ray diffraction data is expected to yield more unconventional DNA models. (ii) The dodecamer helix may suffer from some conformational strain due to the crystal packing producing long helices that include base pair steps with negative *twist*. (iii) The CACA sequence element present in the dodecamer, but not in the G/C decamer, has a peculiar conformation (Timsit *et al.*, 1991) which might influence the neighbouring sequences. In this line of thinking, the sequence context would be very important for DNA conformation. (iv) As devil's advocate one might finally argue that the three-dimensional structure of DNA is so weakly determined that even octamer sequences may adopt many different conformations (or even any conformation) at the mercy of external factors such as crystal packing. This latter point of view is certainly at variance with the observed structural similarity in the central region of the common octamer element as well as with evidence from both further crystallographic work (e.g. Heinemann and Alings, 1991) and solutions studies. NMR spectroscopy, for instance, shows that sequence-dependent features are detectable in synthetic oligonucleotides (Metzler *et al.*, 1990). For the present problem it appears likely that a combination of the first three points may resolve the apparent discrepancy between the results of Timsit *et al.* (1991) and our crystal structure of the G/C decamer. Which way we have to weight these three explanations remains unclear, however.

Table III. Possible intermolecular hydrogen bonds

From	To	Distance (Å)	Type
$C(5)N_4$	$C(7)O_{1P}$	3.3	Major groove–phosphate
$G(10)O_{3'}$	$C(1)O_{4'}$	3.4	End-to-end
$G(10)O_{3'}$	$C(18)O_{1P}$	3.2	End–phosphate
$C(11)N_4$	$C(5)O_{1P}$	3.2	Major groove–phosphate
$C(12)N_4$	$G(4)O_{1P}$	3.3	Major groove–phosphate
$C(15)N_4$	$C(7)O_{2P}$	3.0	Major groove–phosphate

Implications for protein binding

Specific protein–DNA interactions occur predominantly in the major groove of B-DNA which is wider and carries more information in the form of sequence-dependent hydrogen-bonding patterns than the minor groove (Berg *et al.*, 1982; Steitz, 1990; Harrison, 1991). The unusually shallow major groove in the G/C stretch of DNA described above may thus be of functional significance for DNA binding. In the G/C decamer, the base pair edges in the major groove are exceptionally accessible for ligands as observed in the crystal in the form of phosphate–base hydrogen bonds. In much the same way, protein functional groups might approach the base pairs in the major groove. Conversely, the minor groove is deeper and considerably wider than in DNA of mixed sequence. The increased width of the minor groove might also permit base–ligand interactions that are precluded in narrower grooves.

The dodecamer of related sequence, ACCGGCGCCACA (Timsit *et al.*, 1989, 1991), has similar groove dimensions despite deviations in local base pair geometry. In general, it seems that the width and depth of the grooves in B-DNA is a function of the content of AT base pairs or, more generally, of the content of major groove methyl groups (see Table II). As their number increases, the major groove tends to become wider and deeper while the minor groove shows an opposite trend. It is possible that these structural features play a role in specific protein–DNA recognition.

DNase I cuts DNA with above average efficiency at G/C stretches (Drew and Travers, 1984). The crystal structure of a DNase I–oligonucleotide complex (Suck *et al.*, 1988) shows a widened minor groove of the protein bound DNA. It thus appears that G/C sequences are good substrates for the enzyme because they have a wide minor groove in solution as well as in crystals.

Implications for drug binding

Most groove-binding drug molecules are specific for the minor groove of B-DNA where they bind without strict sequence specificity but with preferences for either A/T or G/C-rich sequences (Zimmer and Wähnert, 1986). They have in common specific requirements for the width and curvature of the groove (Goodsell and Dickerson, 1986). Netropsin and distamycin, for instance, prefer the narrow minor groove of A/T-rich DNA for binding (Kopka *et al.*, 1985) whereas mithramycin and anthramycin bind to G/C-rich stretches. The characteristic width and depth of the minor groove of G/C-rich regions in B-DNA, which is clearly distinct from the groove geometry of A/T-rich DNA, may thus help to explain the binding behaviour of groove-specific drug molecules.

To what extent these drug molecules are directed by the groove geometry (and not by specific hydrogen bonding) can be assessed by measuring their binding to oligonucleotides with different major-groove methylation states. Methylating position 5 of cytosines in G/C-rich DNA should be expected to render the minor groove geometry more A/T-like, thereby changing the affinity of drug molecules without altering functional groups.

Materials and methods

Crystallization and diffraction data collection

2.6 μmol of CCGGCGCCGG with dimethyltrityl-protected 5′-end were purchased from TIBMolbiol (Berlin) and purified by reversed phase FPLC. After acid deprotection, final purification was via anion exchange FPLC. As determined by UV melting experiments, the duplex dissociates at 59°C in 20 mM sodium cacodylate pH 7.0.

Crystals were grown at 4°C by microdialysis of 2 mM DNA in 10 mM Tris–HCl pH 7.0, 150 mM magnesium chloride, against the same buffer supplemented with 2-methyl-2,4-pentanediol (MPD). Small colourless plates grew within 2 weeks at an MPD concentration of 24% (v/v). Prior to diffraction work at room temperature, crystals were stabilized by raising the MPD concentration to 40%. Crystal system and space group as determined from precession photographs and verified by diffractometry were rhombohedral, R3. The unit cell parameters, a = 54.07(2) Å and c = 44.59(1) Å (hexagonal axes) indicated that one decamer duplex was present in the asymmetric unit.

X-ray diffraction data were collected from a 0.4 × 0.3 × 0.1 mm^3 specimen on a Turbo-CAD4 diffractometer using Ni-filtered CuK_α radiation produced by an Enraf-Nonius F571 rotating anode generator running at 4.5 kW with 0.3 mm focus. An asymmetric unit of reciprocal space was measured to a nominal resolution of 2.2 Å with ω-scans, variable scan time depending on count rate, stationary background measurement of both sides of a reflection, and measurement of three controls after full hours of exposure time. Since the crystal had suffered little decay after completion, data collection was repeated in the same resolution range and continued until the total loss in intensity exceeded 10%. Data reduction included corrections for Lorentz and polarization factors for crystal decay and for absorption by the semi-empirical method of North *et al.* (1968).

A total of 3078 reflections was observed at the 1σ level and merged with $R_{sym} = \Sigma_{ij}|<I_i> - I_{i,j}| / \Sigma_{ij} I_{i,j}$ = 7.5% to give 2086 unique reflections, equivalent to 78% of the observations expected at 2.2 Å resolution. The outermost shell between 2.5 and 2.2 Å resolution contained 68% of the theoretically observable reflections. In the final cycles of structure refinement, 2019 structure amplitudes between 8 and 2.2 Å were included. Thus, there are ~200 observations per base pair of DNA.

Structure solution and refinement

The dodecamer ACCGGCGCCACA is related in sequence to CCGG-CGCCGG and crystallized as B-type double helix in the same space group with related cell parameters (Timsit *et al.*, 1989). This indicated that the decamer might also be in the B conformation in the crystals described here. Therefore, a canonical B-DNA double helix was constructed from coordinates derived from fibre diffraction studies (Chandrasekaran and Arnott, 1989) and used as a molecular search fragment in the direct space multi-dimensional search program ULTIMA (Rabinovich and Shakked, 1984).

Using 37 structure amplitudes between 25 and 10 Å, ULTIMA provided a set of trial solutions which were further evaluated by rigid body refinement at increasing resolution, first based on group scattering factors for molecular fragments and later on atomic scatterers. This refinement converged at R = 44% and a correlation coefficient of 0.589 for 10 to 3 Å data. The 10 best solutions had in common crystal packing by end-to-end stacking of 10 bp helices.

Refinement of the molecular model was performed with CORELS (Sussman *et al.*, 1977) and later with NUCLSQ (Westhof *et al.*, 1985) gradually increasing the resolution to include all observations to the limit of 2.2 Å. Inspection of difference electron density maps with FRODO (Jones, 1985) and an analysis of crystal contacts indicated an error in position and orientation of the duplex in the unit cell, corresponding to a 1 bp misset of the decamer sequence relative to the true setting on the quasi-continuous helix in the crystal. After this was corrected, refinement started again from the fibre diffraction derived starting model.

With the same methods as before the structure refinement proceeded smoothly and without manual interventions to the final model of CCGGCGCCGG. In the later stages of refinement, possible water molecule positions were selected by stereochemical criteria from the top 20 peaks of difference Fourier maps and incorporated into the model. Due to the limited resolution of the data, water molecules were refined with unit occupancies and no attempt was made to identify bound cations. Structure refinement was terminated when the parameter shifts became negligible and difference electron density maps were without interpretable features.

Acknowledgements

We are grateful to Drs Y.Timsit and D.Moras for the communication of unpublished DNA coordinates. Dr W.Saenger is thanked for continued support and advice, and H.Schindelin and M.Hahn for helpful comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 344/D3, by the Fonds der Chemischen Industrie and by the Alexander von Humboldt-Stiftung through a fellowship to M.B.

References

Arnott,S., Chandrasekaran,R., Millane,R.P. and Park,H.-S. (1986) *J. Mol. Biol.*, **188**, 631–640.

Benevides,J.M., Wang,A.H.-J., Rich,A., Kyogoku,Y., van der Marel,G.A., van Boom,J.H. and Thomas,G.J.,Jr (1986) *Biochemistry*, **25**, 41–50.

Berg,O.G., Winter,R.B. and Von Hippel,P.H. (1982) *Trends Biochem. Sci.*, **7**, 52–55.

Bhattacharyya,D. and Bansal,M. (1989) *J. Biomol. Struct. Dyn.*, **6**, 635–653.

Calladine,C.R. (1982) *J. Mol. Biol.*, **161**, 343–352.

Chandrasekaran,R. and Arnott,S. (1989) In Saenger,W. (ed.), *Landolt-Börnstein, New Series, Group VII.* Springer, Berlin, Vol. 1b, pp. 31–170.

Chuprina,V.P., Heinemann,U., Nurislamov,A.A., Zielenkiewicz,P., Dickerson,R.E. and Saenger,W. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 593–597.

Coll,M., Frederick,C.A., Wang,A.H.-J. and Rich,A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8385–8389.

Cooper,J.P. and Hagerman,P.J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7336–7340.

Cruse,W.B.T., Salisbury,S.A., Brown,T., Cosstick,R., Eckstein,F. and Kennard,O. (1986) *J. Mol. Biol.*, **192**, 891–905.

Dickerson,R.E. (1983) *J. Mol. Biol.*, **166**, 419–441.

Dickerson,R.E. (1985) In Jurnak,F. and McPherson,A. (eds), *Biomolecular Macromolecules and Assemblies.* Wiley, New York, Vol. 2, pp. 471–494.

Dickerson,R.E. (1990) In Sarma,R.H. and Sarma,M.H. (eds), *Structure and Methods*, Volume 3, *DNA and RNA*, Adenine Press, Schenectady, pp. 1–38.

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

Dickerson,R.E. *et al.* (1989) *EMBO J.*, **8**, 1–4.

DiGabriele,A.D., Sanderson,M.R. and Steitz,T.A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1816–1820.

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

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

Duckett,D.R., Murchie,A.I.H., Clegg,R.M., Zechel,A., von Kitzing,E., Diekmann,S. and Lilley,D.M.J. (1990) In Sarma,R.H. and Sarma,M.H. (eds), *Structure and Methods*, Vol. 1, *Human Genome Initiative and Recombination.* Adenine Press, Schenectady, pp. 157–181.

Fairall,L., Martin,S. and Rhodes,D. (1989) *EMBO J.*, **8**, 1809–1817.

Fratini,A.V., Kopka,M.L., Drew,H.R. and Dickerson,R.E. (1982) *J. Biol. Chem.*, **257**, 14686–14707.

Galat,A. (1990) *Eur. Biophys. J.*, **17**, 331–342.

Goodsell,D.S. and Dickerson,R.E. (1986) *J. Med. Chem.*, **29**, 727–733.

Harrison,S.C. (1991) *Nature*, **353**, 715–719.

Heinemann,U. (1991) *J. Biomol. Struct. Dyn.*, **8**, 801–811.

Heinemann,U. and Alings,C. (1989) *J. Mol. Biol.*, **210**, 369–381.

Heinemann,U. and Alings,C. (1991) *EMBO J.*, **10**, 35–43.

Heinemann,U. and Hahn,M. (1992) *J. Biol. Chem.*, in press.

Heinemann,U., Lauble,H., Frank,R. and Blöcker,H. (1987) *Nucleic Acids Res.*, **15**, 9531–9550.

Heinemann,U., Rudolph,L.-N., Alings,C., Morr,M., Heikens,W., Frank,R. and Blöcker,H. (1991) *Nucleic Acids Res.*, **19**, 427–433.

Htun,H. and Dahlberg,J.E. (1989) *Science*, **243**, 1571–1576.

Jones,T.A. (1985) *Methods Enzymol.*, **115**, 157–171.

Keniry,M.A., Levenson,C. and Shafer,R.H. (1987) *J. Biomol. Struct. Dyn.*, **4**, 745–756.

Kennard,O. and Hunter,W.N. (1989) *Q. Rev. Biophys.*, **22**, 327–379.

Kopka,M.L., Yoon,C., Goodsell,D., Pjura,P. and Dickerson,R.E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1376–1380.

Lauble,H., Frank,R., Blöcker,H. and Heinemann,U. (1988) *Nucleic Acids Res.*, **16**, 7799–7816.

Lilley,D.M.J. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6468–6472.

Liu,L.F. and Wang,J.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7024–7027.

Lomonossoff,G.P., Butler,P.J.G. and Klug,A. (1981) *J. Mol. Biol.*, **149**, 745–760.

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

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

Metzler,W.H., Wang,C., Kitchen,D., Levy,R.M. and Pardi,A. (1990) *J. Mol. Biol.*, **214**, 711–736.

Mirkin,S.M., Lyamichev,V.I., Drushlyak,K.N., Dobrynin,V.N., Filipov,S.A. and Frank-Kamenetskii,M.D. (1987) *Nature*, **330**, 495–497.

Murchie,A.I.H., Clegg,R.M., von Kitzing,E., Duckett,D.R., Diekmann,S. and Lilley,D.M.J. (1989) *Nature*, **341**, 763–766.

Nadeau,J.G. and Crothers,D.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2622–2626.

Nelson,H.C.M., Finch,J.T., Luisi,B.F. and Klug,A. (1987) *Nature*, **330**, 221–226.

North,A.C.T., Phillips,D.C. and Mathews,F.S. (1968) *Acta Crystallogr.*, **A24**, 351–359.

Palecek,E. (1991) *Crit. Rev. Biochem. Mol. Biol.*, **26**, 151–226.

Pavletich,N.P. and Pabo,C.O. (1991) *Science*, **252**, 809–817.

Peticolas,W.L., Wang,Y. and Thomas,G.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2579–2583.

Privé,G.G., Heinemann,U., Chandrasegaran,S., Kan,L.-S., Kopka,M.L. and Dickerson,R.E. (1987) *Science*, **238**, 498–504.

Privé,G.G., Yanagi,K. and Dickerson,R.E. (1991) *J. Mol. Biol.*, **217**, 177–199.

Rabinovich,D. and Shakked,Z. (1984) *Acta Crystallogr.*, **A40**, 195–200.

Reid,D.G., Salisbury,S.A., Bellard,S., Shakked,Z. and Williams,D.H. (1983) *Biochemistry*, **22**, 2019–2025.

Rhodes,D. and Klug,A. (1986) *Cell*, **46**, 123–132.

Rich,D., Nordheim,A. and Wang,A.H.-J. (1984) *Annu. Rev. Biochem.*, **53**, 791–846.

Rinkel,L.J., Sanderson,M.A., van der Marel,G.A., van Boom,J.H. and Altona,C. (1986) *Eur. J. Biochem.*, **159**, 85–93.

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

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

Steitz,T.A. (1990) *Q. Rev. Biophys.*, **23**, 205–280.

Suck,D., Lahm,A. and Oefner,C. (1988) *Nature*, **332**, 465–468.

Sussman,J.L., Holbrook,S.R., Church,G.M. and Kim,S.-H. (1977) *Acta Crystallogr.*, **A33**, 800–804.

Timsit,Y., Westhof,E., Fuchs,R.P.P. and Moras,D. (1989) *Nature*, **341**, 459–462.

Timsit,Y., Vilbois,E. and Moras,D. (1991) *Nature*, **354**, 167–170.

Verdaguer,N., Aymami,J., Fernandez-Forner,D., Fita,I., Coll,M., Huyn-Dinh,T., Igolen,J. and Subirana,J. (1991) *J. Mol. Biol.*, **221**, 623–635.

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

Wolk,S., Thurmes,W.N., Ross,W.S., Hardin,C.C. and Tinoco,I.,Jr (1989) *Biochemistry*, **28**, 2452–2459.

Yanagi,K., Privé,G.G. and Dickerson,R.E. (1991) *J. Mol. Biol.*, **217**, 201–214.

Zimmer,C. and Wähnert,U. (1986) *Prog. Biophys. Mol. Biol.*, **41**, 31–112.

Received on December 30, 1991; revised on February 3, 1992