THE HIGH RESOLUTION STRUCTURE OF DNA BY SINGLE-CRYSTAL X-RAY METHODS

Thesis by

Horace Rainsford Drew III

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(Submitted April 18, 1981)

ACKNOWLEDGEMENTS

This thesis would not have been possible without the assistance and cooperation of a great many people. In roughly chronological order, I would like to thank my parents, Horace and Shelley Drew, Jr., who raised me as they saw fit; Rodger Nutt, who suggested that I major in chemistry; Merlyn Schuh, who stimulated my interest in biochemistry; Dick Dickerson and Tsune Takano, who taught me crystallography; Mick Becker, Ben Conner, Mary Kopka, and Dick Wing, for their encouragement; Keiichi Itakura, for a productive collaboration; Shoji Tanaka and Piotr Dembek, without whom I would have never been able to solve a DNA crystal structure; Shoji and Piotr again, for their fish parties; and finally, Dee Barr, who never complained.

ABSTRACT

The crystal structures of two synthetic DNA oligomers have been solved and refined. The first, a tetramer of base sequence d(CpGpCpG) or CGCG, crystallizes as four-base-pair fragment of left-handed Z' double helix. It differs from the Z helix adopted by the related hexamer sequence CGCGCG in that the pucker of guanine sugar rings is C1'-exo rather than C3'-endo. Crystals of the tetramer, grown from an exceptionally high-salt solution, incorporate excess magnesium chloride. The magnesium cations do not appear to play any structural role, but chloride anions, bound to guanine 2-amino groups in the minor groove, may induce the C3'-endo/C1'-exo guanine sugar pucker change via electrostatic repulsion of a nearby phosphate oxygen.

A second compound, the dodecamer CGCGAATTCGCG, crystallizes as slightly more than a full turn of right-handed B-DNA. The fact that both CGCG ends of this molecule remain right-handed, without a trace of any conformational instability, implies that left-right interfaces and left-handed helical regions might be rare or nonexistent in a DNA polymer of mixed sequence.

Although the dodecamer double helix is smoothly deformed into a radius of curvature of 110 Å, bending has little effect on local helix parameters. Instead, the sequence of bases along the chain plays a more important role in helix variability. Purine sugars exhibit the C2'-endo pucker of classical B-DNA, but pyrimidine sugars are distorted into a conformation intermediate between O1'-endo and C1'-exo. Because sugar pucker and glycosyl angle are closely related parameters, this

small difference in sugar pucker leads to subtle sequence-specific variations in base pair position and helix twist. In general, pyrimidine-purine steps open to the minor groove, while purine-pyrimidine steps open to the major.

Ordered water molecules are an integral part of the CGCGAATTCGCG structure, and they are found more in the grooves of the helix than in the first coordination shell of phosphate groups. Major groove waters are predominantly monodentate, but those in the GAATTC minor groove are organized into a single cooperative network or "spine" that extends from the surface of the base pairs to beyond the helical radius of phosphate oxygens. The inner two shells of this network exhibit hydrogen-bonded geometries which are likely to be specific for an uninterrupted stretch of A/T base pairs.

TABLE OF CONTENTS

CHAPTER 1

CGCG: THE PINK SWAN

"Now it is far from obvious, from a logical point of view, that we are justified in inferring universal statements from singular ones, no matter how numerous; for any conclusion drawn in this way may always turn out to be false: no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white."

---Karl R. Popper from

The Logic of Scientific Discovery

1.1 INTRODUCTION

For many years, DNA was thought to be an exclusively right-handed molecule. The structures of right-handed A and B-DNA (1) had been well-substantiated by fiber x-ray diffraction analysis (2), and the only proposal for a left-handed helix (in the alternating copolymer poly(dI-dC), Reference 3) had been thoroughly rebuked. Struther Arnott, one of the more prominent DNA structural chemists, wrote in 1974 (4):

> We describe here details of the structure of a peculiar DNA double helix that occurs when purine and pyrimidine nucleotides alternate along each polynucleotide chain. This structure has been obtained by analyzing X-ray diffraction patterns from oriented, crystalline fibers of poly(dA-dT)·poly(dA-dT), but since we obtain similar patterns from poly(dG-dC)·poly(dG-dC), it appears that there is no requirement for a particular purine or pyrimidine base.

He continues:

> Mitsui et al (1970) obtained from poly(dI-dC)·poly(dI-dC) diffraction patterns very similar to that shown in Plate I. However, the model they proposed for the molecular structure is a bizarre left-handed double helix with unusual furanose ring shapes. Our model for all these complementary poly(dPu-dPy)·poly(dPu-pPy) molecules, while distinct from previously described DNA structures, is a conformationally unexceptional, right-handed double helix with standard furanose rings.

Perhaps if people believe in an idea fervently enough, it cannot possibly come true. Although the structures of poly(dI-dC) and poly(dA-dT) may indeed be right-handed, that of poly(dG-dC) has since been shown to be left-handed double helix (5) which can tolerate unusual furanose ring shapes (6). Arnott, in a reinvestigation of the poly(dG-dC) structure, found that "after prolonged annealing"

this polymer would adopt a left-handed structure (7). In this case, "prolonged" must mean six years, from 1974 to 1980.

Actually, the left-right salt-induced transition of poly(dG-dC) had been seen in solution as early as 1972, but it was not then possible to deduce the structure of the high-salt form (8). The investigators must have suspected the truth, however, for they named the low-salt form "R" and the high-salt form "L".

There is probably no great story of biological significance to be told about left-handed DNA, but there may still be a lesson to be learned, about swans, double helices, and the process of scientific discovery. The remainder of this chapter deals with the crystal structure analysis of CGCG, a left-handed Z' double helix.

REFERENCES

1. Franklin, R. E. and Gosling, R. G. (1953) Acta Cryst. 6, 673-677.
2. Arnott, S., Dover, S. D. and Wonacott, A. J. (1969) Acta Cryst. B25, 2192-2206.
3. Mitsui, Y., Langridge, R., Shortle, B. E., Cantor, C. R., Grant, R. C., Kodama, M. and Wells, R. D. (1970) Nature 228, 1166-1169.
4. Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C. and Watts, L. (1974) J. Mol. Biol. 88, 523-533.
5. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. (1979) Nature 282, 680-686.
6. Drew, H. R., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 286, 567-573.
7. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. and Ratliff, R. L. (1980) Nature 283, 743-745.
8. Pohl, F. M. and Jovin, T. M. (1972) J. Mol. Biol. 67, 375-396.

1.2 A SALT-INDUCED TRANSITION

J. Mol. Biol. (1978) **125**, 535–543

LETTER TO THE EDITOR

A Salt-induced Conformational Change in Crystals of the Synthetic DNA Tetramer d(CpGpCpG)

From a low-salt solution, d(CpGpCpG) crystallizes in space group $P6_122$ with $a = b = 30{\cdot}8$ Å and $c = 43{\cdot}6$ Å. From a high-salt solution, d(CpGpCpG) crystallizes in space group $C222_1$ with $a = 19{\cdot}5$ Å, $b = 31{\cdot}3$ Å, and $c = 64{\cdot}7$ Å. The transition between these two forms appears to be fully reversible in the crystalline state. Native data sets have been collected and Patterson maps calculated from both low and high-salt crystals in preparation for structure analysis.

Solution studies by Pohl & Jovin (1972) have shown that the alternating copolymer poly(dG-dC) undergoes a reversible conformational change at high ionic strength, while related sequences such as the alternating RNA copolymer poly(G-C) and the DNA homopolymer poly(dG)·poly(dC) do not. The activation energy for this salt induced transition has a rather large value of 22 kcal/mole, and is practically independent of chain length. Ethidium bromide, an antibiotic that intercalates into the narrow groove of the double helix, exhibits a much higher affinity for the low-salt form of poly(dG-dC) than for the high-salt form, so that addition of ethidium bromide to a high-salt solution of poly(dG-dC) results in a co-operative reversal of the DNA conformation to its low-salt form (Pohl *et al.*, 1972).

No satisfactory structural model has been proposed to account for the salt-induced conformational change of poly(dG-dC). Sobell *et al.* (1977) have suggested that high-salt poly(dG-dC) may be β-kinked B-DNA, but present evidence does not support this proposal. If a kink in the double helix precedes intercalation, then the high-salt conformation of poly(dG-dC), if continuously kinked, should exhibit a higher affinity for ethidium bromide than should the low-salt conformation, in which a kink is a transient dynamic phenomenon. However, as discussed above, high-salt poly(dG-dC) actually binds ethidium much less tightly than does the low-salt form.

This paper is the first in a crystallographic study of the salt-induced transition of poly(dG-dC). We have obtained single crystals of the synthetic DNA tetramer d(CpGpCpG), to be referred to hereafter simply as CGCG, at both low and high ionic strength, and observed a phase transition from high to low salt within the crystal. Diffractometer data have been collected for each of the two crystal forms, three-dimensional Patterson vector maps have been calculated, and a search for heavy-atom derivatives is under way.

From a low-salt solution, crystals may be grown by diffusion of ethanol into 0·5 to 2·0 mg/ml aqueous solutions of DNA at neutral pH and room temperature. Precession photographs of this form are shown in Figure 1. The space group is $P6_122$ or $P6_522$, with cell dimensions as given in Table 1. The measured density of crystals grown from 5 mM-magnesium acetate indicates that there are 32,580 daltons of DNA and solvent per unit cell, or 2720 per asymmetric unit. Since the molecular weight of a single-stranded CGCG tetramer is 1170, the asymmetric unit may contain one

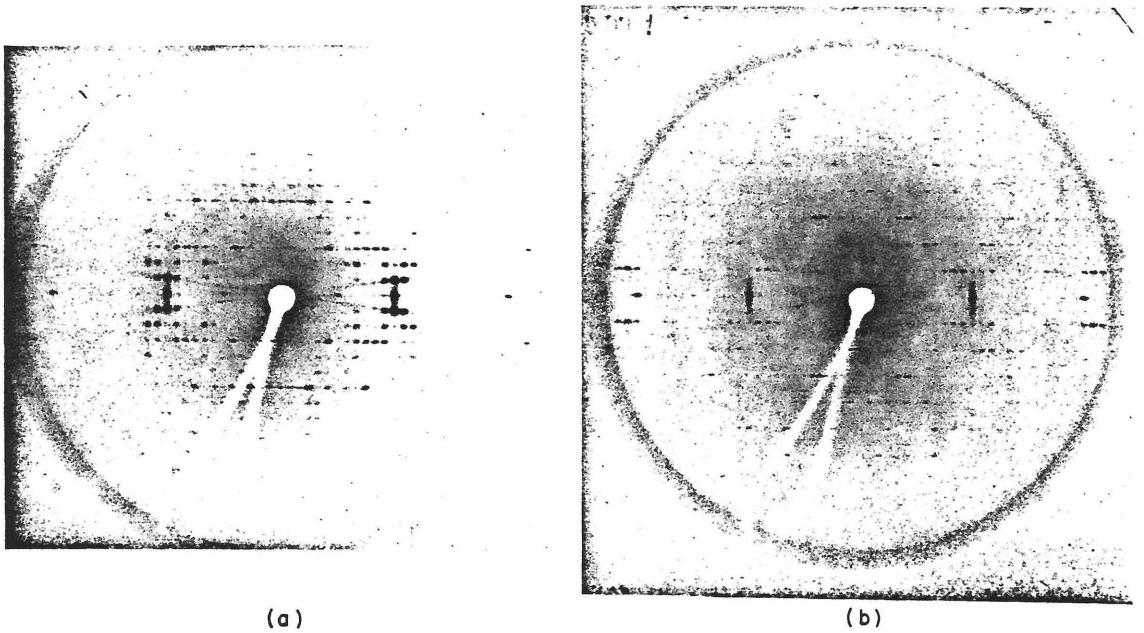


FIG. 1. Precession photographs of low-salt CGCG. $\mu = 30°$, $F = 50$ mm. (a) The $h0l$ zone; (b) the hhl zone. Space group $P6_122$, c^* axis horizontal.

tetramer in a crystal 43% DNA by weight, or 1·5 tetramers in a crystal 65% DNA. For reasons given below we favor the second choice.

The (0,0,12), (1,0,12), and (1,0,13) reflections in Figure 1 are exceptionally strong, with disorder streaks to either side of the c^* axis. The (0,0,24) reflection also is strong,

TABLE 1

Crystal data for low- and high-salt CGCG

Form	Low-salt	High-salt	High-to-low salt transition
a	30·8 Å	19·5 Å	31·2 Å
b	30·8	31·3	31·2
c	43·6	64·7	43·9
γ	120°	90°	120°
V	35,820 Å^3	39,490 Å^3	37,010 Å^3
Space group	$P6_122$	$C222_1$	$P3_121$
ρ_{obs} (g/cm^3)	1·51 ± 0·01	1·60 ± 0·05	
Daltons/cell (incl. solvent)	32,580	38,060	
Daltons/asymmetric unit	2720	4760	
CGCG monomers per asymmetric unit	1·5	2	
Per cent DNA by weight	65%	49%	

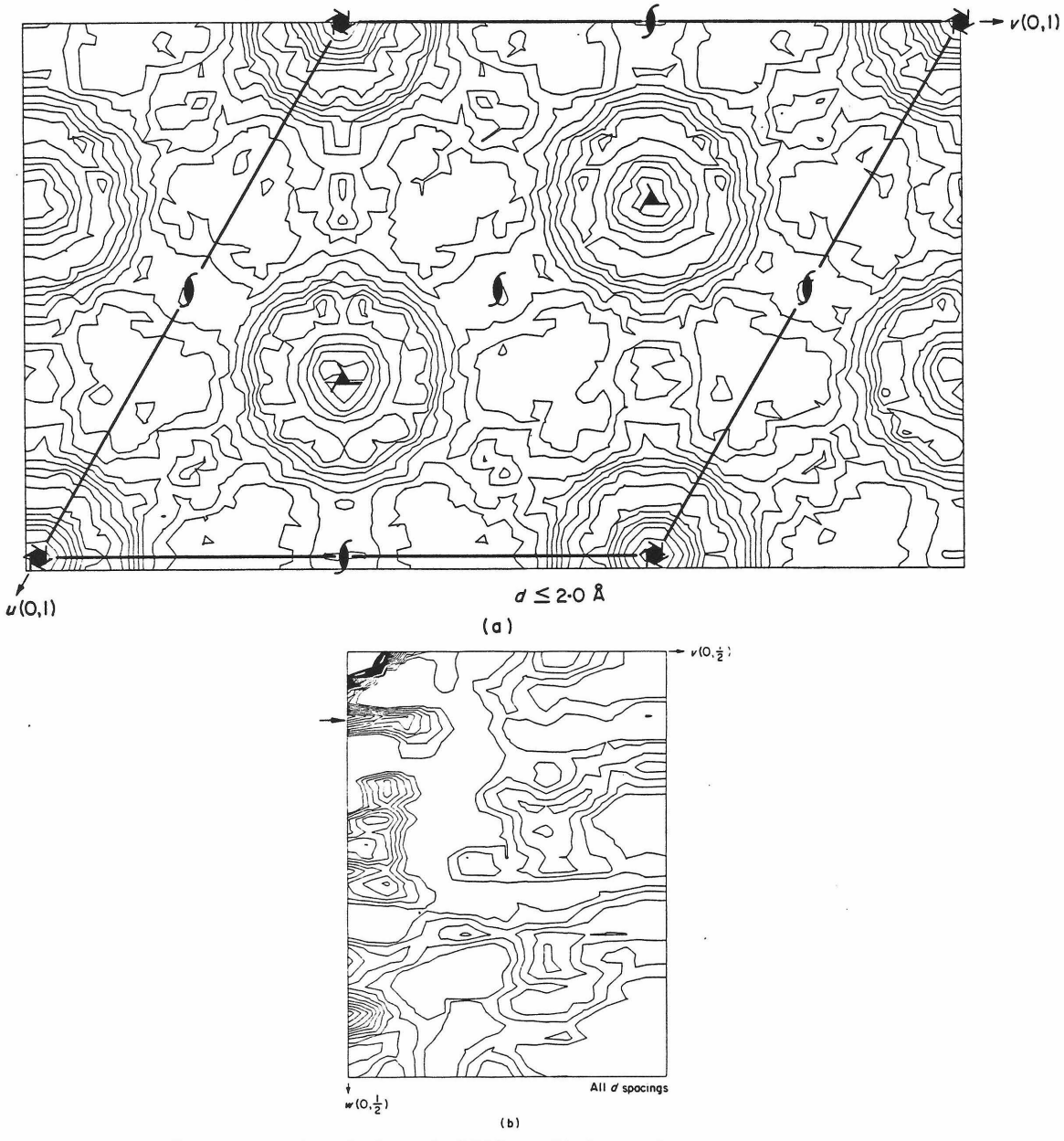


FIG. 2. Patterson sections for low-salt CGCG. (a) Harker section at $w = 1/6$, calculated using only high-angle reflections to enhance contributions from phosphorus atoms. (b) $u = 0$ section showing maximum at $w = 3{\cdot}6$ Å (arrow). Calculated using all reflections.

while the (0,0,6) and (0,0,18) reflections are weak. This suggests that the crystallographic 6_1 screw axis approximates a 12_1 axis, and that the DNA tetramers stack as a 12-fold double helix in which the base planes, separated by $c/12 = 43{\cdot}6/12 = 3{\cdot}6$ Å, are nearly normal to the c axis. One may regard such a helix as a continuous structure in which every fourth phosphate group has been deleted from each strand. The miss-

ing phosphates in the two chains may either be aligned (flush-end stacking of four-base-pair "drums"), or staggered by two bases (continuous helix with overlapping ends).

The a and b dimensions of 30·8 Å suggest that there are three such 12-fold helices in each unit cell, parallel to c and centered at $(x,y) = (0,0)$, (1/3,2/3), and (2/3,1/3), with a 17·8 Å spacing between helix axes. (Langridge *et al.* (1960) have discussed why DNA crystallizes in this way.) This packing scheme is confirmed by the $w = 1/6$ section through the full three-dimensional Patterson map in Figure 2(a). This is not only a Harker section with vectors between symmetry-related atoms; in space group $P6_122$ the $w = 1/6$ Patterson section is in fact a squared Fourier projection of the crystal structure (Buerger, 1959). Figure 2(a) was calculated using only high-angle reflections between 2·0 Å and 1·5 Å in the hope of enhancing the contribution of scattering from phosphorus relative to the lighter atoms (Rosenberg *et al.*, 1973). The three helices are clearly visible, each with a depression at its center. The helix at (0,0) sits on a crystallographic 6_1 screw axis, and the other two occupy 3_1 axes. A continuous 12-fold helix with every fourth phosphate group missing along each strand will have 3_1 symmetry. To accommodate 6_1 symmetry at the origin, one must assume that the helix at (0,0) is disordered, with half phosphates at every other position along the chain. This kind of disordering also produces optimal packing among the three helices in the cell if they are arranged so that the idealized helices (i.e. with all phosphates present) are translationally equivalent.

To appreciate this point, consider the hexagonal cell in Figure 3. If helices A, B, and C were translationally equivalent, then reflections with $h - k \neq 3n$ would be extinguished, and a reduced cell with $a = b = 17{\cdot}8$ Å would result. Although these reflections are not absent, they are systematically weaker than those reflections for which $h - k = 3n$. (Every 3rd horizontal row in Fig.1(a) is especially strong.) Hence helices A, B, and C are not far from being equivalent. The proposed model for low-salt CGCG requires 18 tetramers per cell, or 1·5 per asymmetric unit, and 65% DNA by weight.

Crystals of high-salt CGCG may be grown by evaporation of water from a solution initially containing 0·5 mg DNA/ml, 10% 2-methyl-2,4-pentanediol, and 0·2 M-magnesium chloride. In the early stages of evaporation while the salt concentration is relatively low, some of the DNA may crystallize in its low-salt form. As the salt concentration increases to 1 M-magnesium chloride with continued evaporation, these

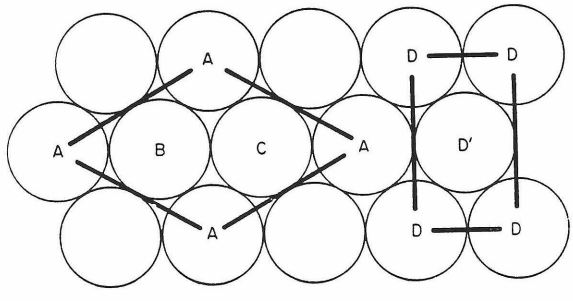


FIG. 3. Inferred packing of helices parallel to the c axis in the hexagonal low-salt form of CGCG (helices A, B, C), and the orthorhombic high-salt form (helices D and D′). Each circle represents a helix normal to the plane of the page. Helices A, B, and C are crystallographically non-equivalent. Helices D and D′ are identical except for the centering operation.

crystals shatter, and crystals of the high-salt form begin to appear. The X-ray photographs (Fig. 4) indicate space group $C222_1$ and the cell dimensions given in Table 1. The observed density leads to 38,060 daltons of DNA and solvent per unit cell, or 4760 daltons per asymmetric unit. Although there are no individual reflections in Figure 4 comparable in intensity to the strong 3·6 Å reflections in Figure 1, a clustering of moderately strong reflections to either side of the c^* axis with spacings around 3·6 Å in Figure 4(a) and (b) suggests that some helicity is retained, although obscured by crystal packing. A strong peak is observed 3·6 Å down the w axis in both the low-salt (Fig. 2(b)) and high-salt (Fig. 5(b)) $u = 0$ Patterson sections, suggesting that the bases stack roughly perpendicular to the c axis in both forms.

Interpretation of the high-salt form is aided by the unexpected finding that a solid-state rearrangement in the crystal to a somewhat disordered low-salt form may be induced if the mother liquor surrounding the crystals is diluted. X-ray photographs of crystals that have undergone this crystalline phase transition are shown in Figure 6. The general pattern of intensities resembles that in Figure 1, but with disordering beyond 7 Å resolution. The cell dimensions (Table 1) are close to those of the original low-salt form. The presence of an (0,0,3) reflection and the Laue symmetry of upper-level precession photographs show that the space group now is $P3_121$ rather than $P6_122$. The cluster of strong reflections around (0,0,12) that has been ascribed to helical base stacking reappears, but is quite diffuse. Whatever the structure of the high-salt form, it must be derivable from the low-salt structure by a relatively small movement of helices or tilting of bases, since the transition can be accommodated moderately well in the crystal.

The most probable relationship between helix packing modes in the low- and high-salt crystals is shown in Figure 3. In the high-salt crystals, helices D and D′ are identical except for the centering operation, and the structure can be indexed on a primitive pseudo-hexagonal lattice with dimensions $a = b = 18{\cdot}4$ Å, $\gamma = 116°$. In

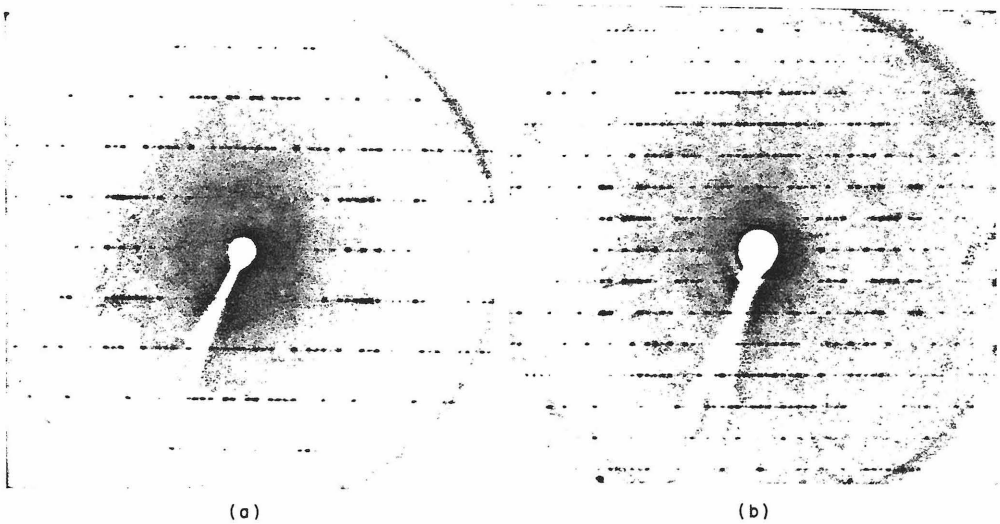
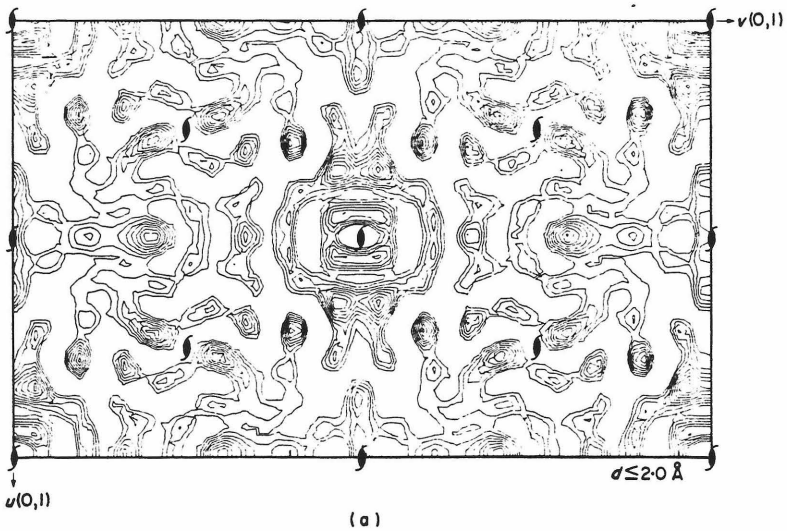


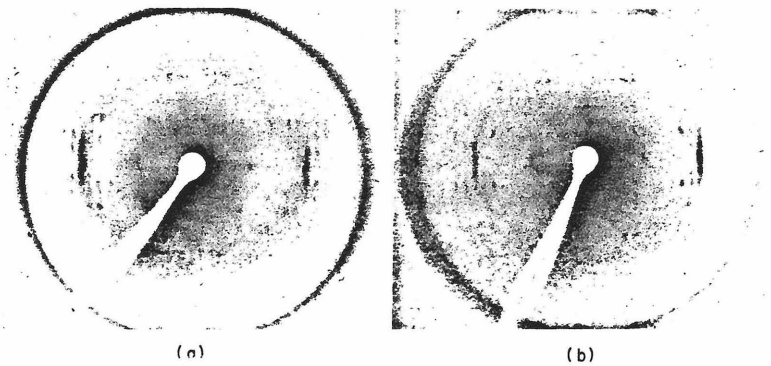
FIG. 4. Precession photographs of high-salt CGCG. $\mu = 25°$, $F = 70$ mm. (a) The $h0l$ zone; (b) the hhl zone. Space group $C222_1$, c^* axis horizontal. Note the pseudo 4_1 intensity distribution of (0, 0, l) reflections at low angles.



(a)

(b)

FIG. 5. Patterson sections for high-salt CGCG. (a) Harker section at $w = 1/2$, calculated using only high-angle reflections. (b) $u = 0$ section showing maximum at $w = 3{\cdot}6$ Å (arrow). Calculated using all reflections. Figs. 2 and 5 are on the same scale.



(a) (b)

FIG. 6. Precession photographs of high-salt crystals of CGCG after re-introducing them to a low-salt environment. $\mu = 20°$, $F = 70$ mm. (a) The $h0l$ zone; (b) the hhl zone. Space group $P3_121$, c^* axis horizontal. Compare these photos with Fig. 1.

the low-salt crystals, helices A, B, and C are not identical. If they were, then a primitive hexagonal lattice would result, with dimensions $a = b = 17{\cdot}8$ Å, $\gamma = 120°$.

Simple division of the c axis length by 3·6 Å in the high-salt form suggests that there are 18 base-pairs up the c axis, but this number is incompatible with the 2_1 screw axis along c and the 2-fold axes perpendicular to c that are demanded by the space group. A better model with 16 base-pairs up the c axis is shown in Figure 7. If a four-base-pair double helix built from two CGCG monomers is considered to be a stubby cylinder or drum of diameter 17·8 Å and height $4 \times 3{\cdot}6$ Å $= 14{\cdot}4$ Å, then only a 9° angle between drum and helix axes is required for four drums tilted as in Figure 7 to take up 64·7 Å along the c direction. This model is compatible with the symmetry of space group $C222_1$, and yields two CGCG monomers or one double-helical drum per asymmetric unit in a crystal 49% DNA by weight. It also accounts for a pseudo 4_1 distribution of intensities along the c^* axis (strong (0,0,4) and (0,0,8) reflections), since the helix of tilted drums in Figure 7 has a local 4_1 screw axis that is reduced to a 2_1 axis in the crystal in the interest of lateral close packing of helices.

Is the salt-induced transition in crystalline CGCG the same as that of poly(dG-dC)? The answer to this question depends to a large extent on whether the conformation of CGCG in the high-salt crystal is continuous in the helical direction. If the drums shown in Figure 7 simply stack on one another without being interconnected, then the change from a low-salt to a high-salt domain in the crystal might represent a mere rearrangement to a more favorable packing scheme. This would necessarily be a non-specific effect of the strongly ionic environment, since we have observed that CGCG crystallizes isomorphously from a low-salt solution in the presence of five different counterions (Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}), and that the low-salt crystals remain ordered when trivalent ions (Sm^{3+} and Er^{3+}) are diffused into the water channels between helices.

On the other hand, if high-salt CGCG turns out to be a continuous 16-fold structure in which the tetramers are conformationally linked in the helical direction (either by base stacking where there is no phosphodiester connection, or by staggering of single strands), then the salt-induced conformational change of crystalline CGCG might be a good model for that of poly(dG-dC). From the approximate packing shown in Figure 7, we would predict for the salt transition of poly(dG-dC) an overall unwinding

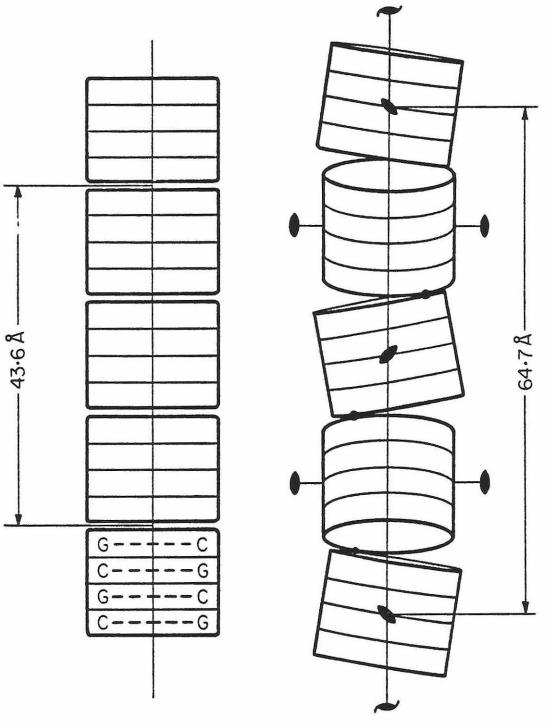


FIG. 7. Possible stacking models for double-helical dimers of CGCG that would account for the observed *c* axis lengths and on- or off-axis 3·6 Å strong intensities in the low-salt form (left) and high-salt form (right).

of 7·5° per base-pair (in going from a 12-fold to a 16-fold helix) and an overall extension of the helix by 11%.

While the preliminary crystallographic observations presented here adequately account for the pseudo-symmetry observed in the X-ray diffraction patterns of both low-salt and high-salt CGCG, the actual nature of this conformational change awaits solution and comparison of the two structures. Hopefully, the structural basis of the phase transition in crystalline CGCG will explain the unusual solution properties of high-salt poly(dG-dC) as well as the specificity of the transition for this DNA sequence.

This work is contribution no. 5788 from the Norman W. Church Laboratory of Chemical Biology. The authors thank Tsunehiro Takano and John Rosenberg for their helpful assistance, and Neil and Gretchen Mandel for their early support of this project. This work has been carried out with the support of grants from the National Institutes of Health and the National Science Foundation.

California Institute of Technology
Pasadena, Calif. 91125, U.S.A.

HORACE R. DREW
RICHARD E. DICKERSON

City of Hope National Medical Center
Duarte, Calif. 91010, U.S.A.

KEIICHI ITAKURA

Received 25 May 1978

REFERENCES

Buerger, A. J. (1959). *Vector Space*, chap. 7, John Wiley and Sons, New York.
Langridge, R., Wilson, H., Hooper, C., Wilkins, M. & Hamilton, L. (1960). *J. Mol. Biol.* **2**, 19–37.
Pohl, F. & Jovin, T. (1972). *J. Mol. Biol.* **67**, 375–396.
Pohl, F., Jovin, T., Baehr, W. & Holbrook, J. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3805–3809.
Rosenberg, J., Seeman, N., Kim, J., Suddath, F., Nicholas, H. & Rich, A. (1973). *Nature (London)*, **243**, 150–154.
Sobell, H., Tsai, C., Jain, S. & Gilbert, S. (1977). *J. Mol. Biol.* **114**, 333–365.

1.3 STRUCTURE OF THE HIGH-SALT FORM

Reprinted from Nature, Vol. 286, No. 5773, pp. 567-573, August 7 1980

High-salt d(CpGpCpG), a left-handed Z′ DNA double helix

Horace Drew, Tsunehiro Takano, Shoji Tanaka, Keiichi Itakura* & Richard E. Dickerson

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125
*City of Hope National Medical Center, Duarte, California 91010

*The DNA tetramer d(CpGpCpG) or CGCG crystallizes from high-salt solution as a left-handed double helix, the Z′ helix. Its structure differs from that of the other known left-handed helix, Z-DNA, by a C1′-*exo *sugar pucker at deoxyguanosines rather than C3′-*endo*, and these represent two alternative solutions to the same steric constraint arising from the* syn *glycosyl bond orientation. The apparent molecular basis for the Z to Z′ transition in going from intermediate to high salt is substitution of a bound anion for water at guanine amino groups, and consequent charge repulsion of anions and backbone phosphates.*

THE predominance of right-handed DNA helices of the A and B families was called into question recently by the discovery[1] that the double-helical DNA hexamer d(CpGpCpGpCpG) is a left-handed helix, and one in which the repeating unit is two base pairs rather than a single step. This 'Z helix' has since been found to explain the X-ray patterns from certain fibres with alternating purine–pyrimidine sequences[2]. We now report that the high-salt form of the tetramer d(CpGpCpG) (hereafter to be termed CGCG) also adopts a left-handed conformation, similar to but not identical with the Z helix.

We, like Wang[1], chose alternating CG oligomers of DNA for study because it has been known for some time[3–6] that alternating poly (dC-dG) undergoes a reversible, salt-induced conformation change with a midpoint at about 0.7 M $MgCl_2$ or 2.5 M NaCl (essentially the same ionic strengths); the exact nature of this change has been a matter of interest. We have reported that such a transition with a change in ionic strength of the surrounding medium can also be observed with the tetramer CGCG in the crystalline state itself[7]. Hexagonal crystals of a low-salt form can be grown by diffusion of ethanol into aqueous solutions of CGCG (for experimental details as well as crystallographic survey data, see ref. 7). High-salt crystals are grown by vapour diffusion from solutions containing DNA, $MgCl_2$ and 2-methyl-2,4-pentanediol (MPD). The hexagonal low-salt crystals that form initially in this latter method crack into smaller blocks as the $MgCl_2$ concentration approaches 1 M. These blocks and the crystals that nucleate subsequently show the orthorhombic X-ray pattern of the high-salt form. Conversely,

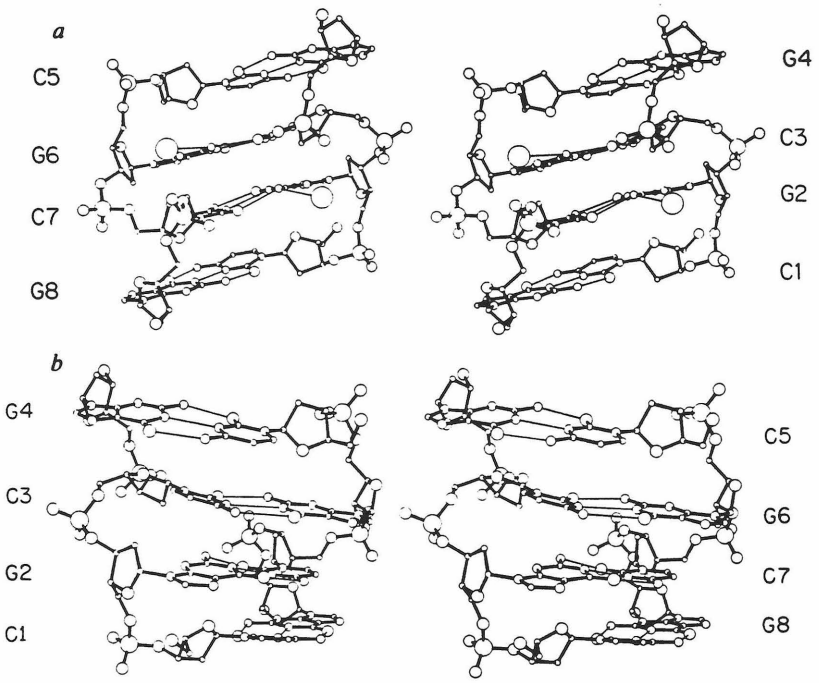


Fig. 1 Stereo pair drawings of the high-salt CGCG double helix. Cytosine and guanine bases are labelled C1 to G8 to the left and right of the two stereo images. Two chloride ions are shown in *a* as large spheres coordinated to the N2 amides of guanosines G2 and G6. Other spheres in order of decreasing radius represent P, O, N and C. The phosphorus atoms along the backbone are referred to as P2, P3 and P4 in one strand and P6, P7 and P8 in the other. *a*, View into the minor groove, down the approximate 2-fold axis. *b*, Oblique view from the outside, in the direction that would correspond to the major groove in the right-handed helices.

high-salt crystals that are placed in a more dilute solution convert immediately from the orthorhombic to a slightly disordered version of the hexagonal X-ray pattern[7]. Although the diffraction intensity distribution is quite different in the two forms, the transition is apparently small enough to be accommodated within the crystal lattice packing. Unit cell dimensions suggest that the transition involves a shifting of the double helices relative to one another along the *c* axis, with the lateral packing of helices being relatively unaffected, as in Fig. 3 of ref. 7. The structure of the high-salt form is now known and is the subject of the present article.

Structure analysis

Crystallization data and preliminary X-ray survey results have already been reported[7]. High-salt crystals are orthorhombic, space group $C222_1$, with cell dimensions $a = 19.50$ Å, $b = 31.27$ Å, $c = 64.67$ Å, and with one CGCG double helix (two CGCG monomer strands) per crystallographic asymmetric unit. The empirical formula of $C_{38}N_{16}O_{22}H_{46}P_3$ per strand leads to a molecular weight of 1,172, and the measured crystal density of 1.6 g cm^{-3} indicates that the crystals are only 49% DNA by weight. The structure was solved by the method of single isomorphous replacement. Diffusion of various ions and complexes into pre-grown CGCG crystals failed to produce a usable isomorphous derivative. Instead, two covalent heavy atom derivatives were prepared by *de novo* synthesis of the tetramer, with 5-bromocytosine in place of one of the cytosines at the first or third position, using the triester synthesis[8]. The derivative with bromine at the third position (3Br-CGCG) crystallized isomorphously with the native compound, whereas that with bromine at the first position (1Br-CGCG) did not. In the light of the final structure, we now can see that insertion of bromine at the beginning of the molecule interferes with stacking of adjacent double helices up the *c* axis, whereas 3Br-CGCG has no such impediment to stacking. Single isomorphous replacement phases[9] calculated from the 3Br derivative yielded a mean figure of merit of 0.42 to a resolution of 2.7 Å. These phases were used to calculate a 'best' or least r.m.s. error electron density map[10]. A Kendrew wire model served as the starting point for Jack–Levitt refinement[11]. The completed structure consists of 158 non-hydrogen DNA atoms and 86 ordered solvent atoms per asymmetric unit. The current residual error or *R* factor for 1,900 reflections ($>2\sigma$) to a resolution of 1.5 Å (beyond which the pattern fades rapidly) is 19.9%.

The CGCG double helix

As shown in Fig. 1, two CGCG strands in the high-salt crystals form a short left-handed antiparallel double helix with Watson–Crick base pairing. As in the Z helix reported for the CGCGCG hexamer[1], cytosines retain their customary *anti* orientation about the C–N glycosyl bond, with the sugar ring turned away from the O2 carbonyl oxygen of the cytosine ring, and guanines adopt an unusual *syn* conformation with the sugar ring bent towards the guanine N3 ring nitrogen. The sugar–phosphate backbone therefore has a zigzag path, with two steps along the chain as the true helical repeat. The helical rotation between repeating units or pairs of base pair steps is −60°. The minor groove containing the guanine amino and cytosine carbonyl groups (Fig. 1*a*) is the only true groove in the helix, because what corresponds formally to the major groove with guanine carbonyl and cytosine amino groups (Fig. 1*b*) is pushed out to the surface of the helix so that it is not a groove at all. The *syn–anti* alternation of glycosyl bonds may limit all such zigzag left-handed helices to alternating purine–pyrimidine sequences, because placing a pyrimidine base in the *syn* conformation brings its O2 carbonyl oxygen close enough to the O1′ oxygen of deoxyribose to disturb their first hydration shells and permit dipole–dipole repulsion.

The low-salt CGCGCG hexamer[1] and the present high-salt CGCG tetramer differ slightly in the tilt of base planes to the helix axis and base repeat spacing along the helix axis. In the hexamer the bases are tilted 7° away from perpendicularity to the helix axis and the average rise per residue along the axis is 3.7 Å. In the tetramer the base plane tilt is 9° and the average rise per residue is 3.8 Å. These differences are associated with a more fundamental change in sugar conformation at deoxyguanosine. In the hexamer the sugar pucker is C2′-*endo* at cytosines and C3′-*endo* at guanines. The tetramer also has C2′-*endo* sugars at cytosines, but at guanines the conformation is C1′-*exo*. This latter is a conformational variant of C2′-*endo*, and quite distinct from C3′-*endo* (Fig. 2). Hence, to a crude first approximation, the sugar pucker is repetitive along the CGCG chain rather than alternating as in CGCGCG.

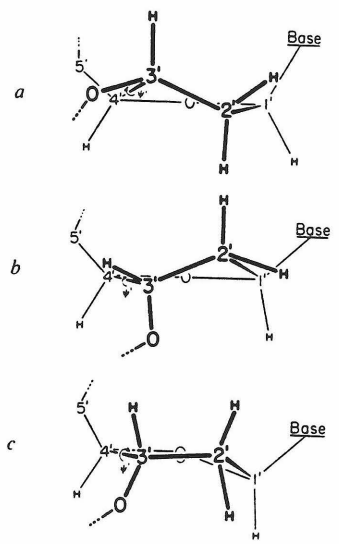


Fig. 2 Deoxyribose ring conformations: *a*, C3′-*endo*; *b*, C2′-*endo*; *c*, C1′-*exo*. The value of the C4′–C3′ torsion angle ψ' is clearly defined by the positions of atoms O3′ (attached to C3′) and C5′, and is 82° (*a*), 144° (*b*), and 120° (*c*) for the conformations drawn here.

The evidence for sugar pucker comes less from direct observation of the rings themselves than from the geometry of the backbone chain. As the X-ray data for the tetramer do not

Table 1 Comparison of left-handed DNA helices

Helix	Z′	Z
Where found	High-salt CGCG	Low-salt CGCGCG
Repeating units per turn	6.07 ± 0.31	6
Lateral crystal packing diameter*	18.4 Å	18.1 Å
Rise per repeat	7.61 ± 0.11 Å	7.43 Å†
Base pair tilt to helix axis	9°	7°
Rotation per repeating unit	−59.4 ± 3.1°	−60°
Glycosyl torsion angle (χ)		
Deoxycytidine	*anti* (23°)	*anti* (21°, 32°)‡
Deoxyguanosine	*syn* (−106°)	*syn* (−112°, −118°)‡
Sugar pucker (ψ')		
Deoxycytidine	C2′-*endo* (141°)	C2′-*endo* (138°, 147°)‡
Deoxyguanosine	C1′-*exo* (122°)	C3′-*endo*§ (99°, 94°)‡
Distance of P from axis		
d(CpG)	6.0 ± 0.3 Å	6.9 Å
d(GpC)	8.2 ± 0.4 Å	8.0 Å
P–P distance across helix		
d(CpG)	11.4 and 12.1 Å	12.5 Å
d(GpC)	16.5 Å	15 Å
Ref.	This work	1

As the repeating unit in each helix is two consecutive bases, C and G, all helix parameters given have been expressed in terms of this repeat, rather than individual bases. Idealized parameters for the Z′ helix have been obtained using a helix-building program provided by John Rosenberg.

*$d = 0.5\,(a^2 + b^2)^{\frac{1}{2}}$, where *a* and *b* are crystal cell dimensions.

†$c/6$, where *c* is the cell dimension along the helix axis.

‡Variants Z_I and Z_{II}, respectively[27].

§Reported as C3′-*endo*, but with torsion angles closer to C2′-*exo* (ref. 12).

extend to atomic resolution, every atom in the deoxyribose ring cannot be resolved and positioned accurately from electron density alone. A better guide to sugar puckering is the torsional angle about the C4′–C3′ bond in the five-membered ring (ψ' in Fig. 2). The X-ray results allow this angle to be specified with considerable accuracy, as the C5′-C4′-C3′-O3′ linkage is an extended chain with well resolved atoms. This torsional angle is directly related to the sugar pucker as shown in Fig. 2: ~82° for the C3′-*endo* configuration, 144° for C2′-*endo* and an intermediate 120° for C1′-*exo* (ref. 12). In the CGCG structure it averages 141 ± 4° for the four cytosine positions, and 122 ± 1° for three of the four guanines; at one of the 3′-terminal guanines, G8, the final O3′ oxygen is disordered, so its ψ' value of 141° is not representative of the ideal helical structure.

The C1′-*exo* conformation can be regarded as a slightly 'bent' C2′-*endo*, and this deformation can be seen from the tetramer structure to arise from a close contact between the N3 nitrogen atom of a guanine ring and the C2′ atom in the sugar ring that is in *syn* orientation to it. This distance averages 3.1 ± 0.1 Å in the crystal, with a C1′-*exo* pucker. Forcing the ring back into a C2′-*endo* conformation would push these atoms 0.4 Å closer together, leading to an unacceptably short non-bonded contact. The observed tetramer backbone, therefore, represents the best that can be done using a repetitive C2′-*endo*-like sugar pucker, in the presence of alternating *syn* and *anti* glycosyl bonds. The alternation of 2′-*endo* and 3′-*endo* present in the CGCGCG hexamer[1] represents a different solution to the same steric problem.

We have termed our structure the Z′ helix because it differs from the Z helix to about the same degree, if in a different way, as the A′ helix with 12 base pairs per turn [13,14] does from the A helix with 11 (ref. 15). Transitions between both pairs, A to A′ and Z to Z′, can be induced by changes in salt concentration. A similar salt-induced transition is observed between the two principal members of the third family of helices, B with 10 base pairs per turn and C with $9\frac{1}{3}$ (refs 16–18). From this viewpoint it might have been more consistent for the C helix to be termed the B′, but consistency in an evolving scientific nomenclature is a goal scarcely to be hoped for.

Table 2 Solvent binding to base amino groups in CGCG

Base no.	N-to-solvent distance (Å)	Refined total electrons	Probable identity
Minor groove:	N2 of guanines		
G2	3.6	16	Cl^-
G4	3.5	8	Cl^-
G6	3.2	11	Cl^-
G8	2.7	5	H_2O
Major groove:	N4 of cytosines		
C1	—	—	—
C3	3.1	4	H_2O or Cl^-?
C5	3.2	7	Cl^-
C7	3.6	8	Cl^-

Unlike CGCGCG, the molecular helix axis in CGCG does not coincide with a crystallographic axis, and the molecules are not stacked in such a way as to simulate a continuous helix. In Fig. 3, the two molecules are related by a 2-fold axis normal to the plane of the paper through the centre of the diagram. Horizontal 2-fold axes above and below these molecules also relate them to other molecules not shown at the top and bottom of Fig. 3. An elongated solvent peak of eight electrons occupancy extends between positions 3.5 Å away from the N2 amino groups of guanines G4 on the two molecules shown, and the distance from the amino group suggests that this is a partial-occupancy, disordered chloride site rather than water. At the other end of each molecule, the amino group of guanine G8 is coordinated to a water molecule at 2.7 Å distance rather than to Cl^-, and a network of solvent molecules crosses each horizontal 2-fold axis to connect to another molecule not shown. Each molecular helix axis is tilted 14.3° away from the crystallographic *c* axis, which is vertical in Fig. 3.

In addition to this mismatch of molecular helix axes, the two molecules shown are not stacked in a way that continues the helical twist. For this to be true, the lower two base pairs of the top molecule in Fig. 3 should be rotated by −60° relative to the upper two base pairs of the bottom molecule. The observed rotation is only −30°, and the reason for this smaller rotation is that it maximizes base pair overlap between molecules. The

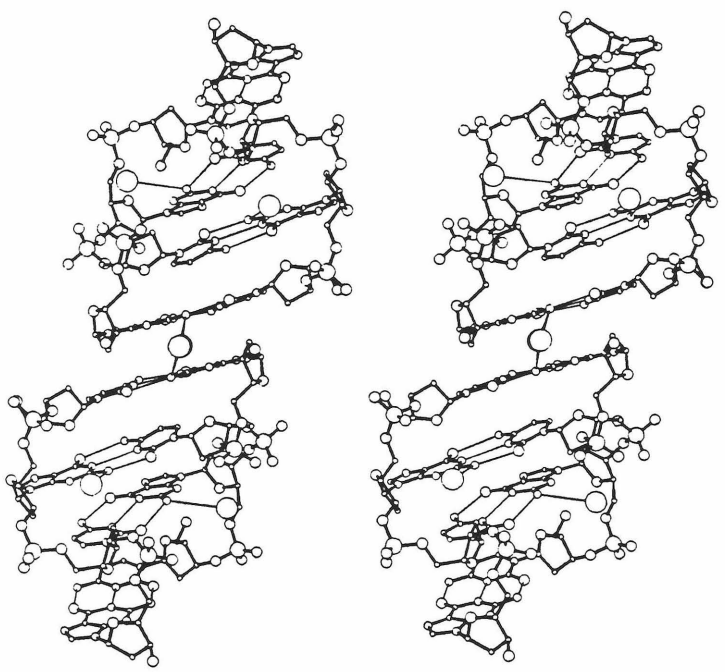


Fig. 3 Stereo drawing of stacking of neighbouring molecules along the *c* axis (vertical). These two molecules are related by a crystallographic 2-fold axis normal to the plane of the drawing through the central bridging chloride ion (large circle). This ion is actually a cigar-shaped peak elongated vertically from one guanine plane to the other, and could represent statistical occupancy of two sterically exclusive coordination positions to the two G4 guanine rings.

absence of linking phosphate groups permits this more favourable stacking with base pairs almost eclipsing one another, whereas within a helix the GpC base stacking involves a 32° twist. The combination of a 60° helical twist within each molecule and a 30° non-helical twist between molecules turns the c axis into a pseudo-4_3 screw axis, leading to a strengthening of $(0, 0, l)$ reflections with $l = 4n$ and weakening of the others, as was noted from the first survey X-ray photographs[7]. The molecular packing is exactly as was predicted in Fig. 7 of ref. 7, except that the lateral 2-fold axes run between molecules rather than through the molecular centres. The 2-fold symmetry of the two halves of any one molecule is only approximate.

The idealized Z′ helix

To establish the helix axis and other helical parameters as shown in Table 1, the rotational and translational parameters that best mapped the first CG half of the molecule onto the second half had to be found analytically. As with CGCGCG, there are six repeating units per turn of ideal continuous helix. The rise per repeat and the tilt of base planes to the helix axis are both slightly greater in CGCG than in CGCGCG, but the most conspicuous differences are the deoxyguanosine sugar puckering, mentioned earlier, and the phosphate separation distances listed in Table 1.

Figure 4 shows three CGCG molecules stacked on top of one another with their helix axes coincident and with −60° rotation between them to simulate one complete turn of Z′ helix. Although each molecule is correct as observed, this idealized arrangement of them has no relationship to the actual molecular packing in the crystals. Zigzag lines connect the phosphorus atoms, including missing phosphates necessary to turn the three molecules into a continuous helix. As was noted for the Z helix, each cytosine deoxyribose ring has its oxygen atom packed against a guanine ring one step farther along the chain. Each base plane in Fig. 4 is puckered or twisted, and the observed deformations imply a steric repulsion by this close-packed sugar oxygen. A careful examination of Fig. 2 of ref. 1 reveals a similar apparent repulsion between guanine rings and deoxyribose oxygens in the Z helix, and this may be a characteristic feature of these left-handed zigzag structures.

Ion binding and helix conformational change

CGCGCG was crystallized from low-salt conditions and CGCG from high salt. The explanation for their observed difference in backbone conformation, and also for the salt dependence of the transition, may lie in the presence of bound chloride ions in the CGCG structure. DNA itself is a polyanion; if one assumes a charge of −3 on each CGCG strand and the observed crystal packing, then the crystals are 2.0 M in negative charges which must be neutralized by cations. By comparison, crystals of CGCGCG with an assumed −5 charge per strand and the published cell constants[1] are 2.6 M in negative charge, which sets a lower limit on the compensating cationic charge molarity. Although CGCGCG was crystallized from low-salt solution conditions, it is by no means in a low-salt or low-ionic strength environment within the crystals themselves, and this may be significant for the helix conformation in the crystal. CGCGCG might better be described as an intermediate salt structure.

As mentioned earlier, high-salt CGCG has additional chloride anions bound within the crystal structure, and these are listed in Table 2. In the minor groove, two sites of occupancy 16

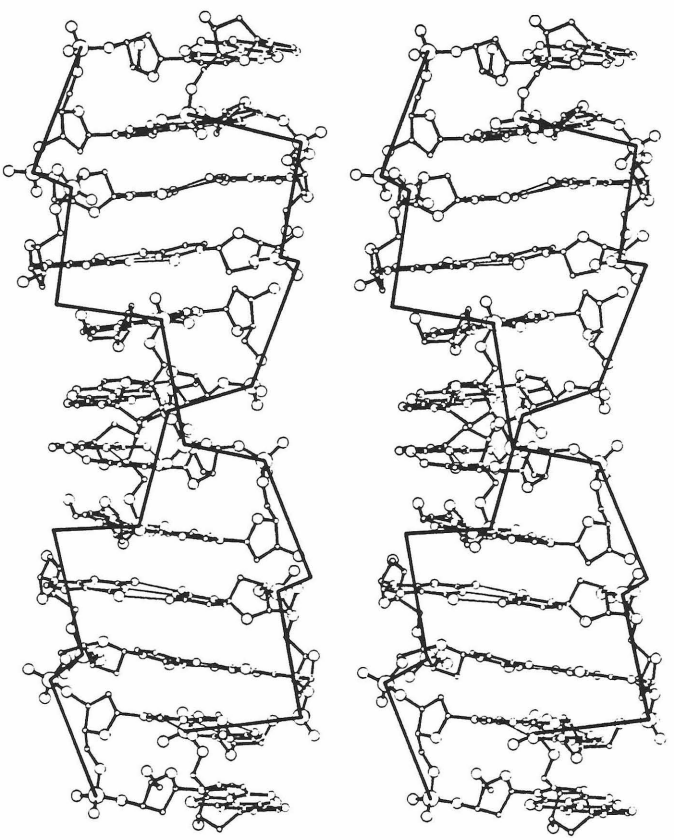


Fig. 4 Stereo view of stacking of three CGCG molecules on top of one another in such a way as to simulate a continuous 6-fold Z′ helix. Zigzag lines connect the phosphate backbone of an ideal helix, including the absent phosphates between molecules. The base plane normal is tilted 9° to the helix axis. The nearest phosphates on opposite strands are only 8.1 Å apart across the minor groove. The 'major groove' is pushed to the outside of the helix and is not a groove at all. Base plane puckering is such as to suggest steric repulsion by adjacent cytosine deoxyribose oxygen atoms. This is an isometric drawing with the viewing point at infinity; previous stereo figures have been perspective drawings with a finite viewing distance.

Table 3 Nucleic acid double-helix geometry

Family	A		B		Z	
Helix sense	Right		Right		Left	
Glycosyl bonds	*anti*		*anti*		*syn/anti*	
Sugar pucker	C3'-*endo*		C3'-*exo* (or *endo*)	C2'-	Alternating	
Helix	A	A'	B	C	Z	Z'
Bases per turn	11	12	10.4–10*	$9\frac{1}{3}$	12	12
Rotation per base	32.7°	30°	34.6°–36°	38.6°	−60°/2	−60°/2
Pitch (Å)	30.9	36	33.8	31	44.6	45.7
Rise per base (Å)	2.8	3.0	3.38	3.31	7.43/2	7.61/2
Lateral packing diameter (Å)	23	23	19.3	19.2	18.1	18.4
Base tilt	16°–19°	10°	−6°	−8°	−7°	−9°
Refs						
Conformation	15	13, 14	15	16, 18	1	This work
Transformations within family	13, 14, 17		16–18, 29, 30		7, this work	
Transformation from B family	16, 21, 28		—		3–7, 20	

*Fibre diffraction experiments indicate 10 base pairs per turn, but solution measurements[31] and theoretical calculations[32] favour a slightly larger non-integral value. The entire B–C family probably represents a continuum of rotation angles in the range 34°–39°.

and 11 electrons are found 3.6 Å and 3.2 Å away from the N2 amino groups of guanines G2 and G6 (Fig. 1*a*). An elongated 8-electron peak bridges N2 amino groups of guanines G4 on adjacent molecules as depicted in Fig. 3. Two more partial-occupancy chloride sites are coordinated to amino group N4 of cytosines C5 and C7 on the 'major groove' surface of the molecule. Hence, allowing for partial occupancy and other possible minor sites, we estimate that there are approximately four bound chlorides per double-stranded helix or two per monomer. This leads to a total negative charge molarity from DNA plus chloride of 3.7 M, which again places a lower limit on compensating cation molarity.

The phosphate group connecting bases G6 and C7 in CGCG is rotated out and away from the narrow groove relative to its position in CGCGCG, in a manner which suggests electrostatic repulsion between it and the chloride bound to G6 (Fig. 1*a*). The phosphate connecting G2 and C3 is rotated away from the chloride bound to G2 in a similar manner, and together these account for the 1.5 Å increase in GpC phosphate separation in the Z' helix shown in Table 1. By contrast, in CGCGCG the phosphate and the guanine N2 amino groups are actually bridged by a hydrogen-bonding water molecule, as shown schematically in the left half of Fig. 5. This difference in phosphate position is sufficient to account for the increase in torsion angle ψ' from 82° in CGCGCG to 122° in CGCG, and the change in sugar conformation at guanine from C3'-*endo* to C1'-*exo*. It is significant that Pohl and Jovin[3] did not see a salt-induced conformational change in either poly(dI-dC) or poly(dA-dT), both of which lack the N2 amino group in the minor groove. Patel's observation that poly(dI-dC) shows only a single phosphorus NMR resonance[19] rather than the two found with poly(dG-dC)[20] indicates that the (dI-dC) polymer is not already in a zigzag left-handed helix in the conditions examined. It may be that the loss of the solvent interaction in Fig. 5 tips a delicate free energy balance in favour of a right-handed helix.

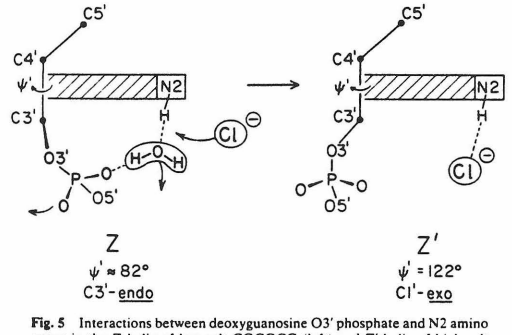


Fig. 5 Interactions between deoxyguanosine O3' phosphate and N2 amino group in the Z helix of low-salt CGCGCG (left) and Z' helix of high-salt CGCG (right). In CGCGCG the two groups are bridged by a water molecule, whereas in CGCG the phosphate is repelled by a bound chloride. This can account for the observed increase in ψ' (the C4'–C3' torsion angle) from 82° in CGCGCG to 122° in CGCG, and the change in sugar pucker from C3'-*endo* to C1'-*exo*.

This is the first single-crystal study of an alteration in double-helix geometry produced by binding of ions, but not the first time that such mechanisms have been invoked to explain transitions monitored by circular dichroism or other methods. Both the A to A' transition[13,14] and the B to C[16–19] may involve specific ion binding, and there may be a certain amount of parallelism in the behaviour of the three principal families of nucleic acid double helices: A, B and Z.

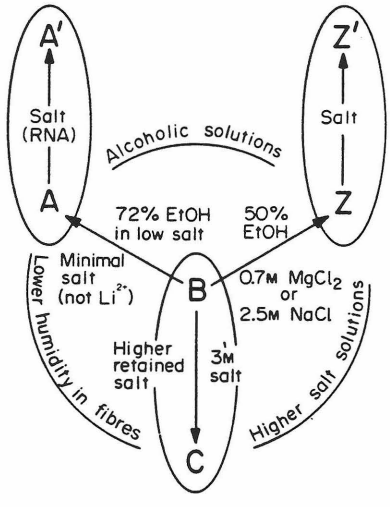


Fig. 6 Families of double helices and transitions between them. Transitions between members of one family, within oval outlines, are influenced by salt concentration. Transitions between families can be brought about by changing relative humidity and salt content in fibres or films, or changing ionic strength or solvent polarity in solution. Critical salt and ethanol concentrations, from refs. 3, 6, 17, are expressed as midpoints rather than endpoints of transitions. The B to Z transition, involving a change in helix sense, may be restricted only to certain alternating purine/pyrimidine sequences.

Why a left-handed helix at all?

To date, X-ray, circular dichroism, laser raman and/or NMR evidence has been found for a salt- or alcohol-induced transition that probably involves a right- to left-handed helix change, with poly(dG-dC) (ref. 3) and poly(dI-Br^{5}dC) (ref. 20), but not with poly(dI-dC) itself, poly(dA-dT), poly(dA-Br^{5}dU), poly(dA-I^{5}dU), the ribonucleic acid poly(rG-rC), or any sequence lacking a purine/pyrimidine alternation[3,20]. It is not difficult to see why zigzag left-handed helices might be limited to such alternating sequences: the steric repulsion mentioned earlier between cytosine and a *syn* deoxyribose ring may be sufficient. However, why should even poly(dG-dC) adopt a left-handed geometry in preference to a right-handed helix of the A or B families, and why the apparent discrimination between various possible alternating sequences?

Properties of the six principal helices in the three helix families are compared in Table 3, and their known interconversions are schematized in Fig. 6. The three sectors of the circle about the B helix represent the three ways of bringing about transitions: lowering the humidity of fibres or films, increasing

the alcohol concentration, or raising the ionic strength by increasing the salt concentration. Transitions within one family seem to be salt dependent and may even involve specific ion binding. This is certainly true for anion binding in the Z to Z′ transition, as the X-ray results have demonstrated. Ivanov[17] has also observed a correlation between the relative abilities of alkali metal cations to induce the B to C transition, the radii of the hydrated cations and the width of the minor groove as the helix winds down from 10 base pairs per turn to $9\frac{1}{3}$. Little information is available about the A to A′ transition in DNA, but in RNA the transition is salt induced. All these interconversions within families are probably non-cooperative as they occur gradually over a reasonably broad range of conditions.

The transitions between the B helix and A or Z families are sharper and are cooperative in nature[4,21]. This is not surprising, because a major change in backbone conformation at one point along the helix will make it easier for neighbouring regions to adopt the same changes, and for the transformation to propagate along the helix. Only the B to C conversion is appreciably temperature sensitive: with a $\Delta H°$ of -10 kcal, the C form is increasingly favoured as the temperature falls[22,23]. In contrast, the B to A and B to Z transitions are essentially isoenthalpic and independent of temperature[3,21].

Ions in high-salt conditions could be imagined to have any of three roles in effecting helix transformations: (1) specific binding to DNA, (2) charge-cloud screening of phosphate–phosphate backbone repulsions, or (3) a general lowering of activity of water molecules because of hydration of ions. Phosphate screening has been invoked to explain both the B to Z and the B to C transitions[1,3,17], but this seems unlikely for two reasons. Although phosphate groups on different helix strands are closer together in the Z or C helices than in the B, they are fully as close in A and Z. If the screening argument were valid, the A helix should be stabilized by high-salt conditions. In fact, it seems to be the form of choice only in minimal salt conditions, whether in fibres or in solution. Moreover, the B to Z conversion is favoured both by alcohol and by salt, and these have opposite effects on phosphate repulsions: salt diminishes the repulsion by charge-cloud screening, but alcohol increases the strength of negative charge repulsion by lowering the dielectric constant of the surrounding medium. Manning[24] has also pointed out that, because of the phenomenon of charge condensation along a polyelectrolyte, the local ion concentration around a polyelectrolyte charge and its degree of neutralization do not follow a simple law of mass action; the phosphate charges in DNA may be largely neutralized by even a small excess of cations, as recent ^{23}Na NMR experiments have shown[25].

The one effect that is shared by the drying of fibres, increasing salt concentration and increasing the amounts of such solvents a ethanol, isopropanol or dioxane, is a general lowering of the activity or effective concentration of water molecules. Each treatment, in a sense, is a kind of local dehydration of the polymer. If one regards the B form as most stabilized by interactions with water molecules, whether specific or nonspecific, and the A and Z forms as less involved with water, then all of the inter-family transitions in Fig. 6 become consistent. Although achieved by different means, they have a common result. This interpretation would also explain why both the 'low-salt' hexamer and the high-salt tetramer of CG adopt a Z conformation in the crystal, because the water activity is lowered by both the high DNA concentration and the high cation concentration required for charge neutrality. The act of crystallization is a process of concentration for both DNA and ions.

As Fig. 6 suggests, the left-handed Z helix family actually seems to be the conformation of choice when water activity is lowered, but only if the base sequence permits its adoption. At present, this seems to require an alternating purine/pyrimidine sequence, or even poly(dG-dC). Is this preference for the Z helix sufficiently strong to produce it even if adjacent regions of DNA are of sequences that cannot adopt the Z conformation? More information is needed from crystal structure analyses of related double helices. Data have been collected in this laboratory by Drew and Wing[26] on a self-complementary DNA dodecamer with the sequence CGCGAATTCGCG, having adjacent Z-compatible and Z-incompatible segments. Other dodecamers are being synthesized with sequences that, hopefully, will exhibit classical B or C helix structures. These, and other carefully chosen sequence variants, should greatly assist the unravelling of the importance of solvent and other factors in stabilizing different helical types, the significance of the new left-handed structures and their ability to influence the conformation of neighbouring sequences that cannot adopt the Z or Z′ configuration.

We thank Michael Becker and Peter Dervan for many productive discussions of helix structure. This is contribution no. 6198 from the Norman W. Church Laboratory of Chemical Biology. This work was carried out with the support of NIH Grants GM-12121 and GM-24393, and NSF Grant PCM79-13959. H.D. was the recipient of a NIH Predoctoral Traineeship.

Received 1 April; accepted 19 June 1980.

1. Wang, A. H.-J. *et al. Nature* **282**, 680–686 (1979).
2. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. & Ratliff, R. L. *Nature* **283**, 743–745 (1980).
3. Pohl, F. M. & Jovin, T. M. *J. Molec. Biol.* **67**, 375–396 (1972).
4. Pohl, F. M., Jovin, T. M., Baehr, W. & Holbrook, J. J. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3805–3809 (1972).
5. Pohl, F. M., Ranade, A. & Stockburger, M. *Biochim. biophys Acta* **335**, 85–92 (1973).
6. Pohl, F. M. *Nature* **260**, 365–366 (1976).
7. Drew, H. R., Dickerson, R. E. & Itakura, K. *J. molec. Biol.* **125**, 535–543 (1978).
8. Itakura, K., Katagiri, N., Bahl, C. P., Wightman, R. H. & Narang, S. A. *J. Am. chem. Soc.* **97**, 7327–7332 (1975).
9. Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. *Acta crystallogr.* **B24**, 997–1003 (1968).
10. Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. *Acta crystallogr.* **14**, 1188–1195 (1961).
11. Jack, A. & Levitt, M. *Acta crystallogr* **A34**, 931–935 (1978).
12. Levitt, M. & Warshel, A. *J. Am. chem. Soc.* **100**, 2607–2613 (1978).
13. Arnott, S., Fuller, W., Hodgson, A. & Prutton, I. *Nature* **220**, 561–564 (1968).
14. Arnott, S., Hukins, D. W. L., Dover, S. D., Fuller, W. & Hodgson, A. R. *J. molec. Biol.* **81**, 107–122 (1973).
15. Arnott, S. & Hukins, D. W. L. *Biochem. biophys. Res. Commun.* **47**, 1504–1509 (1972).
16. Marvin, D. A., Spencer, M., Wilkins, M. H. F. & Hamilton, L. D. *J. molec. Biol.* **3**, 547–565 (1961).
17. Ivanov, V. I., Minchenkova, L. E., Schyolkina, A. K. & Poletayev, A. I. *Biopolymers* **12**, 89–110 (1973).
18. Arnott, S. & Selsing, E. *J. molec. Biol.* **98**, 265–269 (1975).
19. Patel, D. in *Nucleic Acid Geometry and Dynamics* (ed. Sarma, R. H.) 185–231 (Pergamon, New York, 1980).
20. Patel, D. J., Canuel, L. L. & Pohl, R. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2508–2511 (1979).
21. Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D. & Schyolkina, A. K. *J. molec. Biol.* **87**, 817–833 (1974).
22. Brahms, S., Brahms, J. & Van Holde, K. E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3453–3457 (1976).
23. Chan, A., Kilkuskie, R. & Hanlon, S. *Biochemistry* **18**, 84–91 (1979).
24. Manning, G. S. *Q. Rev. Biophys.* **11**, 179–246 (1978).
25. Anderson, C. F., Record, M. T. Jr & Hart, P. A. *Biophys. Chem.* **7**, 301–316 (1978).
26. Wing, R. M., Drew, H., Tanaka, S., Itakura, K. & Dickerson, R. E., in preparation.
27. Wang, A. H.-J. *et al. Science* (submitted).
28. Tunis-Schneider, M. J. & Maestre, M. F. *J. molec. Biol.* **52**, 521–541 (1970).
29. Hanlon, S., Brudino, S., Wu, T. T. & Wolf, B. *Biochemistry* **14**, 1648–1660 (1975).
30. Wolf, B. & Hanlon, S. *Biochemistry* **14**, 1661–1670 (1975).
31. Wang, J. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 200–203 (1979).
32. Levitt, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 640–644 (1978).

1.4 CONFORMATION AND DYNAMICS IN Z-DNA

Conformation and Dynamics in a Z'-DNA Tetramer

HORACE R. DREW and RICHARD E. DICKERSON

Norman W. Church Laboratory of Chemical Biology,
California Institute of Technology,
Pasadena, California U.S.A. 91125

Submitted to the Journal of Molecular Biology

Running Title: Conformation and Dynamics in Z'-DNA

SUMMARY

Two kinds of conformational variability are known for left-handed Z-DNA: the Z-to-Z' transition, which involves a change in guanine sugar pucker from C3'-endo to C1'-exo, and the Z_I-to-Z_{II} transition, which corresponds to a simple three-atom phosphate group rotation. Neither of these motions substantially affects base stacking or helical twist, and this is because the degree of independent motion of phosphates, sugars and base pairs is greater in the left-handed Z helix than in right-handed B-DNA. Thermal vibrations as deduced from refined isotropic temperature factors are smaller for Z-DNA than for B-DNA, and are more evenly distributed throughout the helix framework, which also is indicative of reduced conformational coupling. These conclusions are based on single-crystal structure analyses of CGCG and CGCGCG helical fragments; a more complete analysis of Z-DNA flexibility will have to wait for comparable analyses of much longer oligomers.

1. Introduction

The discovery of left-handed Z-DNA in crystals of the hexamer CGCGCG (Wang et al, 1979) has prompted a resurgence of interest in DNA conformation and dynamics. No double helix has yet been shown to adopt a left-handed conformation in vivo, but certain alternating copolymers, notably poly(dG-dC), poly(dG-dT) and poly(dI-dBrC), can assume Z helix geometries in solution or in fibers under conditions of decreased water activity (Pohl and Jovin, 1972; Pohl, 1975; Patel et al, 1979; Arnott et al, 1980).

The probability of finding left-handed DNA in the midst of a right-handed genome may be considerably diminished, however, by the instability of a left-right interface. No one has yet been able to build a stereochemically acceptable model for a left-right interface without the flip-out of at least one base pair, and the surprisingly large activation energy of 21 kcal/mole for the left-right conversion (Pohl and Jovin, 1972) implies that this is exactly what is happening in solution. This would be consistent with the observation that the DNA dodecamer sequence CGCGAATTCGCG, having two Z-compatible CGCG ends and a Z-incompatible AATT center, remains fully right-handed under salt and hydration conditions in which CGCG and CGCGCG become left-handed (Wing et al, 1980).

Still, it may be of some use to study the conformational properties of Z-DNA in order to better understand, by comparison, properties

of the more biologically important right-handed B form. B-DNA has recently been characterized in several single-crystal x-ray structure analyses (Quigley et al, 1980; Drew et al, 1981), and a comparison of B and Z structures might prove particularly informative.

In this paper, we present a detailed analysis of Z-DNA conformation and dynamics as seen in the high-salt orthorhombic form of the DNA tetramer CGCG. This compound crystallizes from both low-salt and high-salt solution as a four-base-pair fragment of Z helix (Drew et al, 1978; Crawford et al, 1980; Drew et al, 1980), and a previous article dealt with the mechanism by which chloride anions, substituting for ordered water molecules, induce a change in guanine sugar pucker from C3'-endo to C1'-exo (Drew et al, 1980). To be accurate, we should refer to the C1'-exo, chloride-bound structure as Z'-DNA to distinguish it from the C3'-endo, water-bound Z form, but it is a thesis of this paper that Z and Z' are trivial variants of the same helix structure. The same will be shown to hold true for phosphate-rotation variants Z_I and Z_{II} (Wang et al, 1981).

2. Refinement of the Structure

CGCG crystallizes from high-salt solution (0.5-1.0 M $MgCl_2$) in space group $C222_1$, with cell constants a = 19.50, b = 31.27 and c = 64.67 Å. Low-salt hexagonal crystals, with Laue symmetry P6/mmm as determined by precession photography, appear at $MgCl_2$ concentrations

of less than about 500 mM (Drew et al, 1978). (The 0.2 mM salt concentration reported by Crawford et al (1980) for preparation of the high-salt form is incorrect.)

Two CGCG single strands, or one double helix, make up the asymmetric unit of the orthorhombic crystal. The structure was initially solved by single isomorphous replacement, and refined as described in Drew et al (1980) to a two-sigma R factor of 0.20. An additional six cycles of Jack-Levitt refinement (bringing the total number of cycles to 90) have led to some improvements in resolution and stereochemistry. These latter cycles included all 3200 zero-sigma data between the resolution limits of 5.5 and 1.5 Å, and not just the 1900 two-sigma data in the same range. Each reflection was measured once from a crystal of dimensions 0.6 x 0.3 x 0.2 mm. Final R factors are 0.21 for the two-sigma data set and 0.29 for the less accurate zero-sigma data set. Two poorly positioned low-occupancy solvent peaks were removed from the original phasing model, leaving 158 non-hydrogen DNA atoms and 84 ordered solvent in the asymmetric unit. Final atomic coordinates, individual isotropic temperature factors, observed and calculated structure factors, and calculated phases have been deposited with the Brookhaven Protein Data Bank.

Because of less-than-atomic resolution, we have used the C5'-C4'-C3'-O3' torsion angle as a guide to sugar conformation at guanine residues. All deoxyribose rings in the structure as originally reported (Drew et al, 1980) were somewhat flattened from their true

shapes, and we know now that this distortion was caused by the use of improper energy restraints, chiefly bond angles within the deoxyribose rings. Westhof and Sundaralingam (1980) have since published an authoritative work on the geometry of the ribose ring, as observed by single-crystal x-ray methods. Internal ribose bond angles remain approximately constant in the transition from C3'-endo to C2'-endo, no angle varying by more than 3°. Appropriate standard values are 109.5° at O1', 106° at C1' and C4', and 101.5° at C2' and C3'. Use of these new values has improved the shape of rings at both cytosines and guanines.

The two-sigma R factor actually rose slightly, from 0.20 to 0.21, during the final six cycles of refinement. This was not because of changes in the sugar geometries, but because of a change in the data set against which the model was refined. The quantity being minimized in these last cycles was not the two-sigma, but the zero-sigma R factor, and this fell from 0.32 to 0.29 as the two-sigma value increased from 0.20 to 0.21.

3. Conformational Flexibility in a Z Helix

Two research groups have presented discussions on the conformational flexibility of Z-DNA. We originally showed that guanine sugars in the high-salt orthorhombic form of CGCG adopt a C1'-exo pucker (Drew et al, 1980), as compared to a C3'-endo pucker for these

rings in low-salt CGCG and CGCGCG (Wang et al, 1979; Crawford et al, 1980). The mechanism of sugar pucker change also involves a slight GpC phosphate rotation, and is related to ion binding in the high-salt crystal (see Figure 5 of Drew et al (1980)). The C1'-exo guanine form was called Z'-DNA in order to distinguish it from the C3'-endo guanine Z form.

Wang et al (1981) later pointed out that GpC phosphates in a Z helix can take either of two orientations, rotated into (I) or out of (II) the minor groove. The numerals I and II in each case indicate how many water molecules are involved in a hydrogen-bonded bridging interaction between the guanine 2-amino group and a GpC phosphate oxygen. This Z_I-Z_{II} motion is similar both in direction and effect to the Z-Z' motion, but can easily be distinguished by careful structural analysis.

a. The Z_I-to-Z_{II} Transition

As part of a comparison of variations in the Z helix, published coordinates for Z_I and Z_{II}-DNA (Wang et al, 1981) were expanded via helical symmetry operations to generate a CGCG double helix. In Figure 1, the Z_{II} variant of CGCG (dark bonds) is shown superimposed on the Z_I variant (light bonds). Significant differences between the two conformations are localized entirely in the GpC phosphate linkage, where three atoms, the phosphorus and its two

negatively charged oxygens, are either rotated into (Z_I) or out of (Z_{II}) the minor groove.

Which torsion angles are involved in the Z_I-to-Z_{II} motion?* As shown in Table 1, guanine angles ε and ζ change significantly, by -75° and +124°, respectively. At cytosines, it is the α and β angles that shift, by -77° and -57°. All of these changes take place as a result of the three-atom PO_2 group rotation. Other differences, such as the 45° change in guanine α values, are not structurally related to the phosphate swiveling.

b. The Z-to-Z' Transition

Using a least-squares superposition program, coordinates for orthorhombic CGCG (Z'-DNA) were fitted to those of Z_I-DNA. Superposition was based mainly on atoms in the four base pairs and the CpG portions of the sugar-phosphate backbone. The free 5'-terminal oxygen, and C2', C3', O3' atoms of the 3'-terminal sugars were given zero weight during fitting, because as end groups they are free to adopt conformations not typical of an ideal infinite helix. Atoms in the GpC phosphate groups were also given zero weight, because they belonged to a region of the molecule whose relative motion we wished to observe.

As shown in Figure 2, the only significant differences between Z and Z' result from a change in sugar conformation at central guanine

residues. In Z-DNA, the sugar pucker lies intermediate between C3'-endo and C2'-exo, while in Z'-DNA it is close to C1'-exo. This sugar pucker alteration also induces a slight shift in the position of GpC phosphates. There are two GpC phosphates in Figure 2, at the left and right center of the diagram. Both of the Z-DNA phosphates (light bonds) are in a I orientation, but the swivel of Z' phosphates (dark bonds) is mixed: the one on the right (between G2 and C3) happens to be like Z_I, while the one on the left (between G6 and C7) is Z_{II}. Thus, differences in position between light and dark GpC phosphates in this figure describe, on the right, a simple Z_I to Z_I' motion and, on the left, a combination of Z-to-Z' and Z_I-to-Z_{II} motions (more precisely, Z_I to Z_{II}').

In terms of torsion angles, Z and Z' differ only in the guanine δ linkage. As shown in Table 2, Z_I and Z_{II} guanines have an average $\langle\delta\rangle = 97°$, while for Z' guanines $\langle\delta\rangle = 122°$. A particularly informative way to study these structural variations is via a (δ,χ) plot, in which the sugar ring torsion angle δ is plotted against glycosyl angle χ. A plot of this sort for orthorhombic CGCG is shown in Figure 3. Cytosine residues 1, 3, 5 and 7 cluster about Z_I and Z_{II} positions, while central guanine residues 2 and 6 lie slightly to the right of Z_I and Z_{II} values. Sugar rings at both 3'-terminal guanines 4 and 8 are disordered, in one case toward C2'-exo and in the other toward C2'-endo. (X-ray evidence for disorder in these unconstrained sugars will be presented below in Section 5.)

c. Sugar Pucker and Glycosyl Angle

The (δ,χ) plot of Figure 3 also illustrates several principles of Z helix structure. The first of these is that cytosines and guanines exhibit different kinds of conformational flexibility, and this is probably because they adopt very different orientations about the glycosyl bond. Cytosine glycosyl angles χ are free to change between -145° and -180°, close to their anti value of -154° in right-handed A-DNA (Arnott et al, 1980), but cytosine sugar puckers are relatively restricted to a narrow range around C2'-endo. This does not mean that Z-DNA cytosine sugars must always be C2'-endo, merely that no other sugar conformation has been observed to date.

At guanines, χ angles are relatively fixed in a syn orientation of about +70°, but sugar puckers are more variable, able to change from C3'-endo/C2'-exo to C1'-exo. In neither extreme does the sugar attain its lowest energy conformation of C3'-endo or C2'-endo. Why should this be? A plausible reason might be that close nonbonded contacts between the sugar and the base limit the range of flexibility. In the C1'-exo limit, the 2'-carbon of the sugar ring comes within 2.9 Å of guanine N3; to reach the C2'-endo conformation would require an even shorter and stereochemically unacceptable separation of approximately 2.5 Å. In the C2'-exo limit, it is the 3'-carbon of the sugar that closely approaches the guanine base.

These differences between cytosine and guanine flexibilities are informative, but there is a more fundamental similarity which should not be overlooked: at both cytosines and guanines, glycosyl angles χ show little dependence on the conformation of deoxyribose sugars. Left-handed Z-DNA is very unlike right-handed B-DNA in this respect, since one of the strongest torsional correlations in a B helix is between δ and χ (Drew et al, 1981).

d. Base Stacking and Helix Twist

In B-DNA, where the δ-χ correlation is quite strong, a slight distortion of the sugar pucker from C2'-endo to C1'-exo induces large structural perturbations via changes in the χ angle (Dickerson and Drew, 1981). A sugar pucker change of similar magnitude in Z-DNA (the Z-Z' motion) does not substantially affect the χ angle, and hence would be expected to have little influence on base stacking or helix twist. Actual base overlap geometries for orthorhombic CGCG, as presented in Figure 4, demonstrate that this is indeed the case. In these dinucleotide steps, sugar conformations average C1'-exo at guanines, yet ring overlaps are almost exactly equivalent to those reported for CGCGCG and hexagonal CGCG, structures in which the guanine puckers are C3'-endo/C2'-exo (Wang et al, 1979; Crawford et al, 1980).

Variations in the twist of a Z helix must also be limited since, as shown in Table 3, local twist angles for orthorhombic high-salt CGCG average -15° at CpG steps and -45° at GpC, producing in their sum the same -60° rotation per two-base-pair step that is characteristic of Z-DNA (Wang et al, 1979). Local rotation angles at CpG and GpC steps for low-salt CGCG and CGCGCG have not yet been reported, but judging from the near-identity of base stacking geometries they cannot be too different from the -15° and -45° seen here. Propellor twists within G/C base pairs are only on the order of 6°-9°, and this low value may be due to interstrand base overlap (Dickerson and Drew, 1981).

4. Distribution of Thermal Motion in a Z Helix

There is yet another difference between the Z and B forms of DNA, one that may also be related to changes in the δ-χ relationship. In a B helix, phosphates exhibit considerably more thermal and/or statistical motion than do sugars, and sugars more than base pairs (Drew et al, 1981). A possible structural interpretation of this effect is that small perturbations in base stacking or helix twist are amplified as they propagate radially outward toward the phosphates. Since phosphates and sugars can move more independently of one another in Z-DNA than in B, the prediction of this model would be that thermal motion should be more evenly distributed among groups in Z-DNA.

Isotropic temperature factors for high-salt CGCG, as determined from least-squares x-ray refinement, are consistent with this hypothesis. They average 5.4 at base pairs, 7.9 at sugars and 7.8 at phosphates, to give a base/phosphate ratio of 5.4/7.8 = 0.69. Comparable values for a B-DNA dodecamer are 28, 42 and 51, and produce a ratio of only 28/51 = 0.55 (Drew et al, 1981). Direct comparison of these numbers is not necessarily justified, however, since the overall isotropic temperature factor of the B helix crystal is much larger than that of the Z (Wilson plot values of 32 and 8, respectively).

a. Vibration Diagrams for Orthorhombic CGCG

The temperature factors listed above are group averages, and do not convey any information about the three-dimensional distribution of thermal motion in the crystal. One of the simplest ways to illustrate this latter feature is by means of a "vibration diagram," such as the one shown in Figure 5a. The size of each atom in this drawing has been set proportional to its r.m.s. displacement, which in turn is related to the square root of its individual isotropic thermal factor B. Atomic radii have been placed on the same relative scale as the equivalent dodecamer stereos in Drew et al (1981).

Individual B values are much nosier than group B values at this resolution (1.5 Å), and from the variance of values within cytosine rings an estimate of δ-B = 5 may be made for the one-sigma error in

any light-atom thermal parameter. There may be some signal present in the noise, however, since the R factor increases substantially from 0.21 to 0.25 if individual atomic B values are replaced by group-averaged parameters (i.e., a single averaged value for all eight atoms in a cytosine ring). Incidentally, the group value of 0.25 is not much less than a single-parameter Wilson plot value of 0.26, simply because different groups in the structure have roughly equivalent temperature factors (5-6 for bases, 8 for sugars and phosphates).

b. Twofold-Averaging of Temperature Factors

In order to enhance the signal-to-noise ratio, individual isotropic B values were averaged over the two chemically equivalent CGCG strands in the asymmetric unit, and the results are shown in Figures 5b and 5c. The distribution of twofold-averaged B factors is chemically reasonable: all 158 atoms, except those on the 3' end of each strand, display thermal and/or statistical motions of similar magnitudes. Only at 3'-terminal guanine residues 4 and 8 are larger thermal vibrations seen, involving the five-membered part of the guanine ring and the C2', C3', O3' atoms of the unconstrained sugar, but these motions are almost certainly due to end group effects.

5. Conclusions

Although Z-DNA has no known biological function, its physical properties may be so different from those of biologically active right-handed B-DNA that much can be learned from a study of its structure. Conformational variations within a Z helix appear to be limited by the syn orientation of guanine residues, and the few conformational motions which do occur (Z_I-to-Z_{II}, Z-to-Z') do not affect the helix twist. This is probably because the sugar ring torsion angle δ (C4'-C3') and the glycosyl torsion angle χ (C1'-N) are only weakly correlated with one another. Isotropic thermal motions in the orthorhombic CGCG Z' helix are more uniformly distributed among phosphates, sugars and base pairs than is the case for B-DNA, and this is also consistent with more independent kinds of motion for Z-DNA rigid groups.

The observations presented here have been somewhat tentative, because the tetramer sequence CGCG is too short to be used as a good guide to the behavior of an ideal infinite helix. The same problem will be faced, to a lesser extent, with the hexamer sequence CGCGCG (Wang et al, 1979). A crystal structure analysis of a full turn of left-handed Z helix, possibly a dodecamer of sequence CGCGTATACGCG or CGTACGCGTACG, would do much to improve our understanding of Z-DNA conformation and dynamics.

Acknowledgments

We would like to thank R. E. Marsh for his helpful criticism and advice. This work was carried out with the support of National Institutes of Health grants GM-12121 and GM-24393, and National Science Foundation grant PCM79-13959. H.D. was also the recipient of a National Institutes of Health Predoctoral Traineeship. This is contribution No. 6407 from the Division of Chemistry and Chemical Engineering.

References

Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. and Ratliff, R. L. (1980) . Nature 283, 743-745.

Crawford, J. L., Kolpak, F. J., Wang, A. H.-J., Quigley, G. J., van Boom, J. H., van der Marel, G. and Rich, A. (1980). Proc. Natl. Acad. Sci. USA 77, 4016-4020.

Dickerson, R. E. and Drew, H. R. (1981). J. Mol. Biol., submitted.

Drew, H. R., Dickerson, R. E. and Itakura, K. (1978). J. Mol. Biol. 125, 535-543.

Drew, H. R., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980). Nature 286, 567-573.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1981). Proc. Natl. Acad. Sci. USA, in press.

Levitt, M. and Warshel, A. (1978). J. Am. Chem. Soc. 100, 2607-2613.

Patel, D., Canuel, L. L. and Pohl, F. M. (1979). Proc. Natl. Acad. Sci. USA 76, 2508-2511.

Pohl, F. M. (1975). Nature 260, 365-366.

Pohl, F. M. and Jovin, T. M. (1972). J. Mol. Biol. 67, 375-396.

Quigley, G. J., Wang, A. H.-J., Ughetto, G., van der Marel, G., van Boom, J. H. and Rich, A. (1980). Proc. Natl. Acad. Sci: USA 77, 7204-7208.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. (1979). Nature 282, 680-686.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H. and Rich, A. (1981). Science 211, 171-176.

Westhof, E. and Sundaralingam, M. (1980). J. Am. Chem. Soc. 102, 1493-1500.

Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980). Nature 287, 755-758.

Footnotes

*For definitions of torsion angles, see Table 1.

Table 1. Torsion Angles for Z-DNA

Residue	χ	α	β	γ	δ	ε	ζ	Variant
C	-159	-137	-139	56	138	- 94	80	I
C	-148	146	164	66	147	-100	74	II
(II-I)	11	- 77	- 57	10	9	- 6	- 6	
G	68	47	179	-165	99	-104	- 69	I
G	62	92	-167	157	94	-179	55	II
(II-I)	- 6	45	14	- 38	- 5	- 75	124	

Notes to Table 1:

Adapted from Wang et al (1981).

Main chain conformation angles are defined as:

$$P\overset{\alpha}{\text{——}}O5'\overset{\beta}{\text{——}}C5'\overset{\gamma}{\text{——}}C4'\overset{\delta}{\text{——}}C3'\overset{\varepsilon}{\text{——}}O3'\overset{\zeta}{\text{——}}P$$

with zero at the fully eclipsed or cis position and positive clockwise rotation of the farther pair of atoms. The glycosyl angle is similarly defined in terms of atoms O1'——C1'$\overset{\chi}{\text{——}}$N1——C2 for pyrimidines (C,T) or O1'——C1'$\overset{\chi}{\text{——}}$N9——C4 for purines (G,A). (Adapted from Wang et al (1981).)

Table 2. Torsion Angles for Orthorhombic CGCG (Z'-DNA)

Residue	χ	α	β	γ	δ	ε	ζ	Variant
C1	-167	--	--	-122	137	- 78	78	-
G2	64	86	170	165	116	-120	- 57	I
C3	-178	178	-154	87	137	- 99	87	I
G4	64	41	-167	-161	103	--	--	-
C5	-146	--	--	27	140	- 92	72	-
G6	75	76	175	179	120	-179	24	II
C7	-150	175	-178	39	148	- 71	47	II
G8	77	96	163	-176	148	--	--	-
Z' Avgs:								
C	-160	177	-166	51*	141	- 85	71	I+II
G	70	75	175	-178	122	-150	- 17	I+II
Z Avgs:								
C	-154	185	-168	61	143	- 97	77	I+II
G	65	70	186	-184	97	-142	- 7	I+II
(Z'-Z):								
C	- 6	- 8	2	- 10	- 2	12	- 6	I+II
G	5	5	- 11	6	25	- 8	- 10	I+II

*Omitting C1 value.

Table 3. Local Helix Parameters for Orthorhombic CGCG (Z'-DNA)

Base Pairs	Propellor Twist (°)	Helix Twist Angle* (°)	Rise Per Base Pair (Å)	Type of Step
C1/G8	7.2 (2.3)			
		-14.1 (0.1)	5.33 (0.07)	CpG
G2/C7	7.7 (2.2)			
		-45.0 (0.3)	3.50 (0.11)	GpC
C3/G6	9.1 (2.3)			
		-14.8 (0.3)	4.71 (0.05)	CpG
G4/C5	5.6 (1.9)			
Averages:	7.4	-14.5	5.02	CpG
		-45.0	3.50	GpC

Notes to Table 3:

Propellor twist is defined as the dihedral angle between base planes. Helix twist and rise are based on vectors between C1' and N positions, as in Drew et al (1981). The overall helix twist and rise for the single dinucleotide step in CGCG are -59.0°(2.3) and 7.57 Å (0.09).

*Rotation per base pair.

Figure Captions

FIGURE 1. Molecular superposition of conformations Z_I (light bonds) and Z_{II} (dark bonds), as illustrated for a CGCG double helix. Positions of base pairs and sugars are not systematically affected by the Z_I-to-Z_{II} transition, because structural differences between the two forms reside primarily in the orientation of the central GpC phosphates.

FIGURE 2. Molecular superposition of conformations Z_I (light bonds and Z' (dark bonds). Coordinates for Z' are those of high-salt orthorhombic CGCG. Bases in the two crystallographically independent strands of CGCG are numbered 1-4 and 5-8 in their respective 5'-to-3' directions, and sugar rings of 1 and 5 are numbered here for identification.

FIGURE 3. Correlation plot between glycosyl torsion angle χ and C5'-C4'-C3'-O3' torsion angle δ. Each sugar is represented by a numbered circle, with angles taken from Table 2. Z_I and Z_{II} locate idealized variants of the Z helix as derived from other single-crystal x-ray structures (Wang et al, 1981). Guanines are distinguished by heavy circles, and cytosines by light. The δ values for ten standard endo and exo conformations, and their calculated energies in kcal/mole relative to C3'-endo, are given along the top

of the plot (Levitt and Warshel, 1978).

FIGURE 4. Base stacking in orthorhombic CGCG (Z'-DNA). The bases in each pair of adjacent dinucleotides are numbered at left and right. The viewing direction corresponds to a view down the top of the helix in Figure 2. (a) and (c) are CpG steps, and (b) is the GpC step whose phosphate positions differ in the different variants of the Z helix. Note the C1'-exo pucker of guanine sugar rings in (b).

FIGURE 5. "Vibration diagram" depiction of the CGCG double helix. (a) and (b) are oblique views into what formally corresponds to the major groove, and (c) is a perpendicular view into the minor groove. Identification numbers for the bases in the two strands are given in (a). The view in (c) is identical to that in Figure 2. In (a) the radius of each atom has been set to u/3, where u is the r.m.s. displacement as obtained from its individual isotropic temperature factor parameter, $B = 8\pi^2u^2$. Temperature factors which refined to negative values (as low as -7) were reset to +2 for purposes of the illustration. In (b) and (c), atom sizes correspond to values of u/3 that have been twofold-averaged over both CGCG single strands.

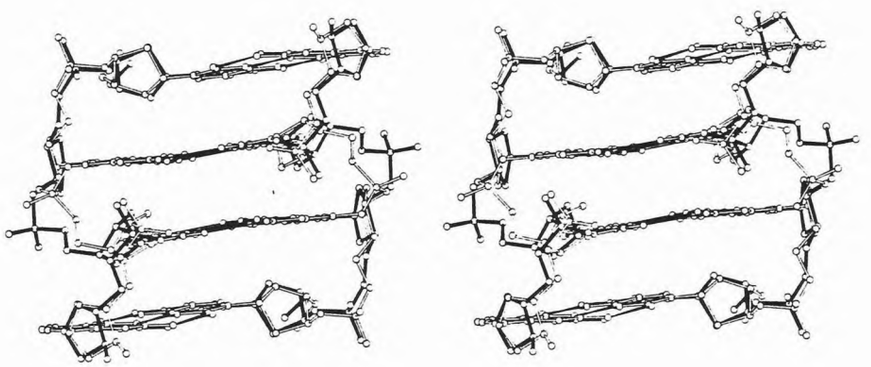


Figure 1

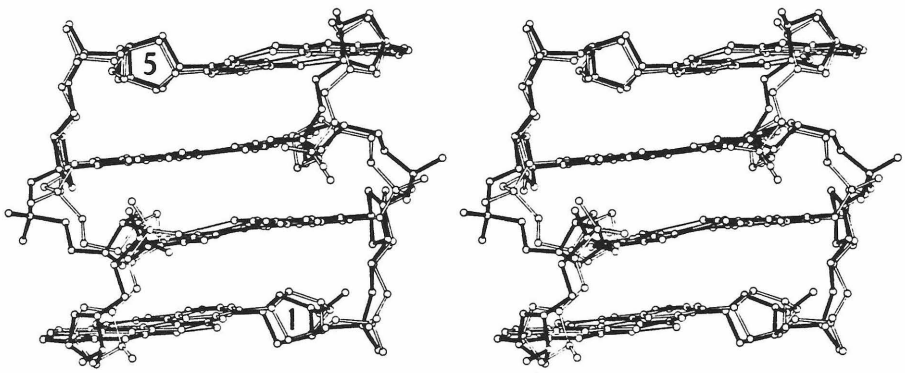


Figure 2

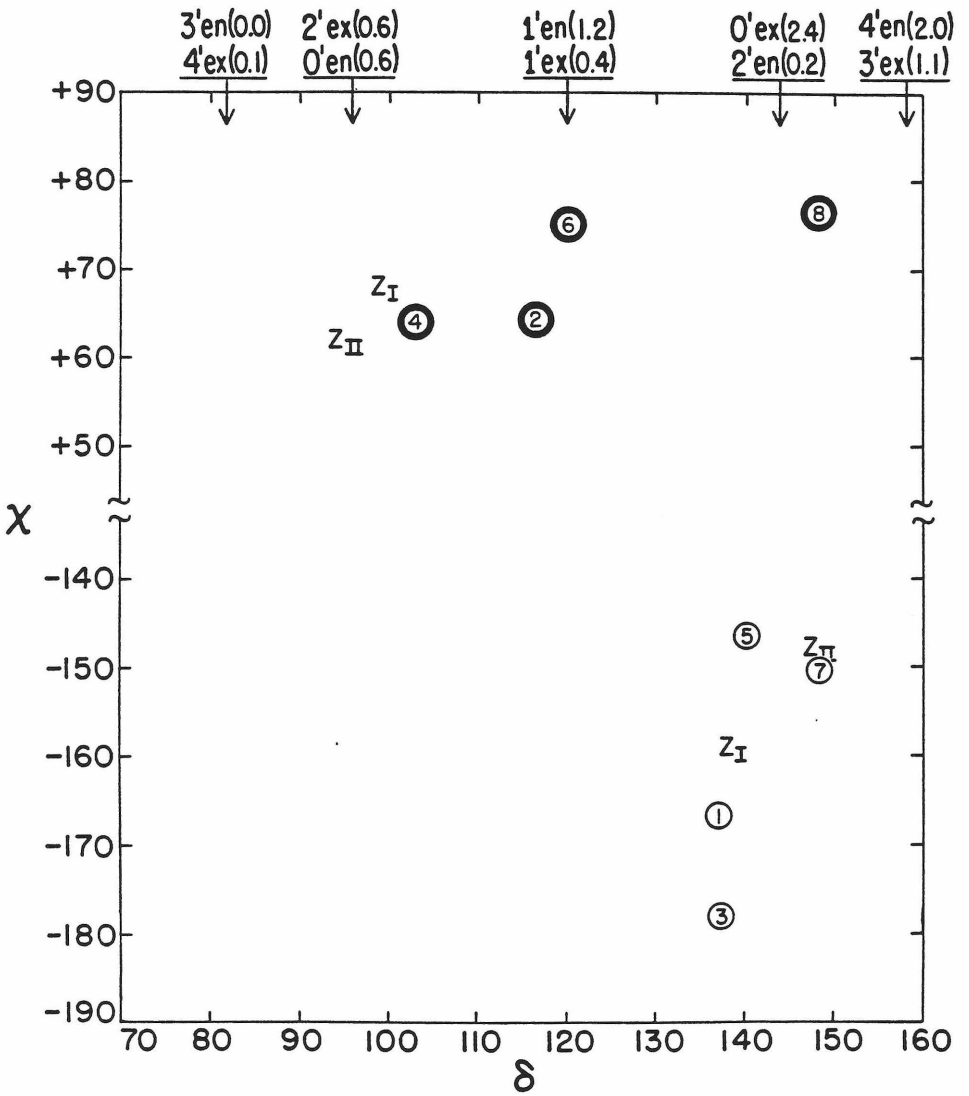

3'en(0.0)
4'ex(0.1)
2'ex(0.6)
0'en(0.6)
1'en(1.2)
1'ex(0.4)
0'ex(2.4)
2'en(0.2)
4'en(2.0)
3'ex(1.1)
+90
+80
+70
+60
+50
-140
-150
-160
-170
-180
-190
χ
70
80
90
100
110
120
130
140
150
160
δ
ZI
ZII
ZI
ZII
1
2
3
4
5
6
7
8

Figure 3

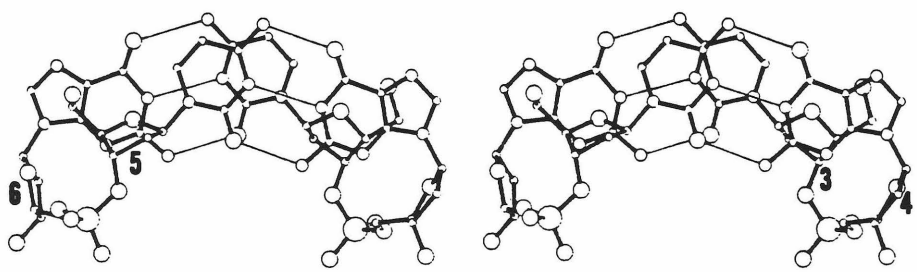


Figure 4a

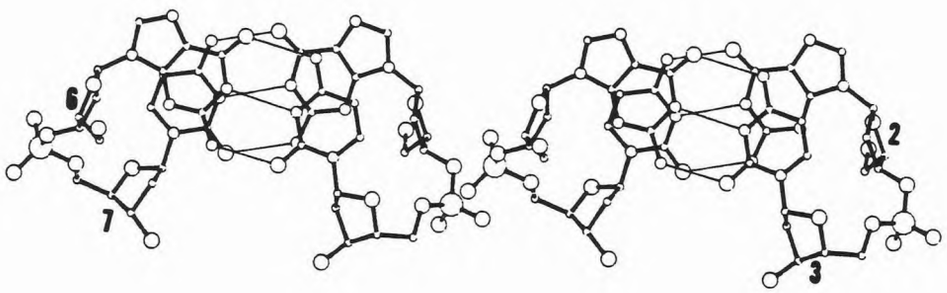


Figure 4b

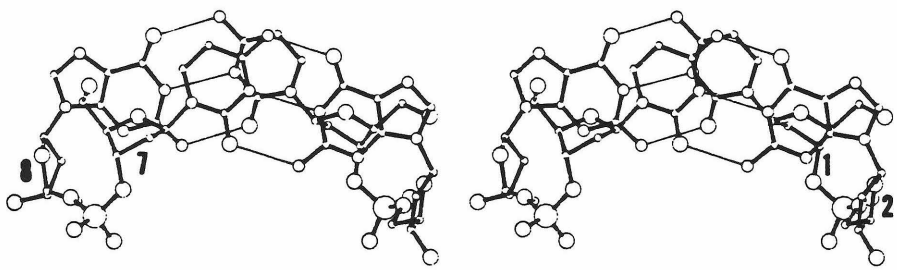


Figure 4c

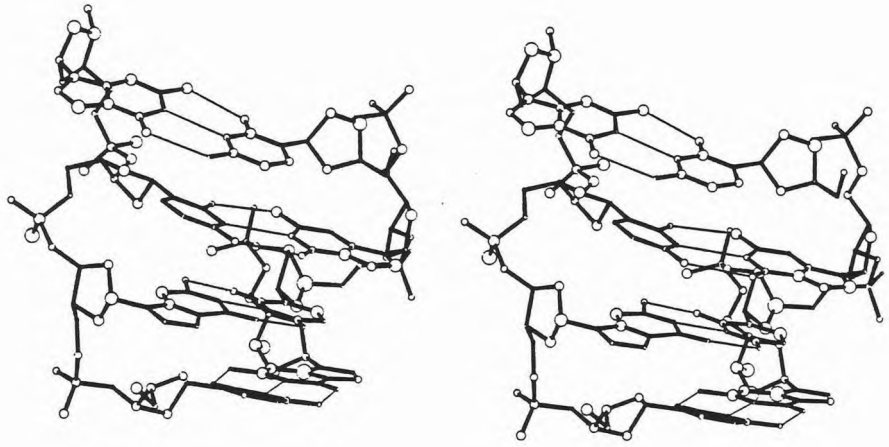


Figure 5a

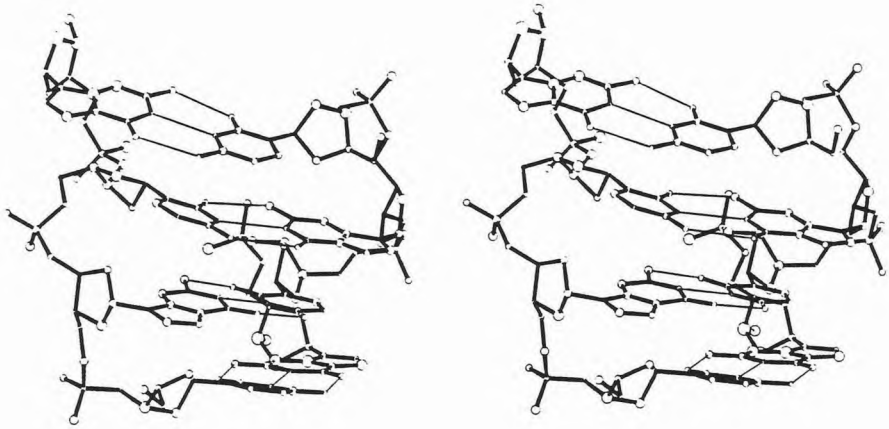


Figure 5b

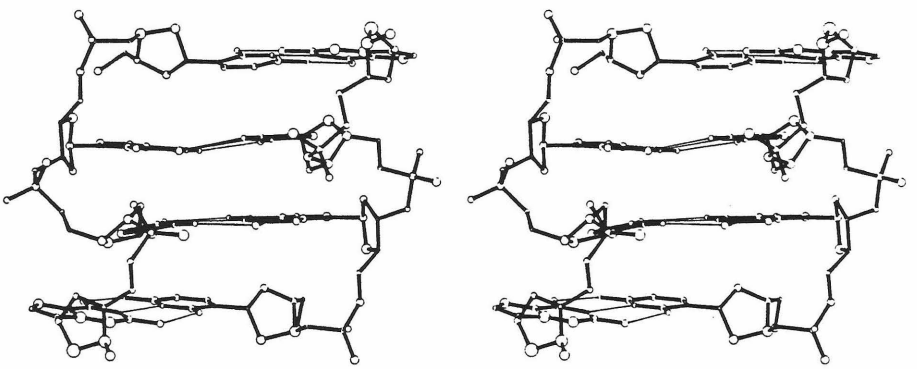


Figure 5c

1.5 CONCLUSIONS

Z-DNA is unlikely to be important in the biological function of the DNA polymer, both because of its requirement for a perfectly alternating pyrimidine-purine sequence and because of the instability of a left-right interface. Nevertheless, something may be learned about biologically active right-handed B-DNA by a comparison of B and Z structures. For example, the torsional rigidity of Z-DNA appears to be greater than that of the B form, and this may have consequences for the adoption of the Z form under conditions of negative superhelicity. The B to Z transition, like the B to A, is associated with a decrease in the relative water activity, and as will be shown below in Section 3.5 this may be because of highly specific interactions between B-DNA and ordered water molecules.

CHAPTER 2

CGCGAATTCGCG: A FULL TURN OF B HELIX

"Ille sinistrorsum, hic dextrorsum abit;
Unus utique error, sed varies illudit partibit."

--Horace

(This to the right, that to the left hand strays,
and both are wrong, but wrong in different ways.)

2.1 INTRODUCTION

When Watson and Crick first proposed a right-handed double-helical structure for DNA in 1953 (1), their model was based in part on a simple-minded interpretation of fiber x-ray diffraction patterns taken by Franklin and Gosling some months before (2). The initial agreement index between the Watson-Crick model and the fiber data was quite poor (R = 0.83), but least-squares refinement procedures were able to reduce the residual to a more acceptable R = 0.36 (3). Thus, the correctness of the right-handed model was apparently confirmed.

It was not until 1980 that someone thought to optimize the fit of alternative models to the same set of fiber diffraction intensities. Gupta et al (4) did this for a series of models: left-handed B, right-handed B, and mixed-sugar-pucker B. They found that, regardless of the molecular details (such as handedness), the R factor could be reduced into the range 0.31-0.36 by systematic refinement. This work has gone largely unreferenced (probably because the authors then go on to argue that left and right-handed B-DNA duplexes are equally possible), but it serves to point out the insensitivity of these fiber diffraction data to details of molecular structure. Actually, if the B-DNA models of Gupta et al are examined in closer detail, it can be seen that the left and right-handed models are nearly mirror images of each other in radial projection. This is probably why the poorly resolved fiber intensities fit either equally well.

Single-crystal x-ray studies of DNA structure have become possible only in the last few years, as a side benefit of improvements in DNA

synthesis technology (5). The synthetic dimer d(pTpT) crystallizes as a fragment of right-handed 8 base per turn single helix (6), while the tetramer d(pApTpApT) melts in the middle to form two 11 base pair per turn ApT dinucleotide steps (7). As discussed above in Chapter 2, both CGCGCG and two salt modifications of CGCG crystallize as short stretches of a 12 base pair per turn left-handed zig-zag double helix (8,9,10).

All of the more recent single-crystal x-ray structures of oligomeric DNA have produced right-handed structures. The tetramer CCGG forms a four-base-pair fragment of A helix (11); the hexamer CGTACG, doubly-intercalated by daunomycin, adopts a B-like conformation (12). This chapter deals with the crystal structure analysis of CGCGAATTCGCG, a dodecamer which crystallizes as slightly more than a full turn of Watson-Crick B-DNA.

REFERENCES

1. Watson, J. D. and Crick, F. H. C. (1953) Nature 171, 737-738.

2. Franklin, R.E. and Gosling, R. G. (1953) Acta Cryst. 6, 673-677.

3. Arnott, S., Dover, S. D. and Wonacott, A. J. (1969) Acta Cryst. B25, 2192-2206.

4. Gupta, G., Bansal, M. and Sasisekharan, V. (1980) Proc. Natl. Acad. Sci. USA 77, 6486-6490.

5. Itakura, K. and Riggs, A. D. (1980) Science 209, 1401-1405.

6. Camerman, N., Fawcett, J. K. and Camerman, A. (1976) J. Mol. Biol. 107, 601-621.

7. Viswamitra, M. A., Kennard, O., Jones, P. G., Sheldrick, G. M., Salisbury, S., Falvello, L. and Shakked, Z. (1978) Nature 273, 687-688.

8. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. (1979) Nature 282, 680-686.

9. Drew, H. R., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 286, 567-573.

10. Crawford, J. L., Kolpak, F. J., Wang, A. H.-J., Quigley, G. J., van Boom, J. H., van der Marel, G. and Rich, A. (1980) Proc. Natl. Acad. Sci. USA 77, 4016-4020.

11. Ben Conner, personal communication.

12. Quigley, G. J., Wang, A. H.-J., Ughetto, G., van der Marel, G., van Boom, J. H. and Rich, A. (1980) Proc. Natl. Acad. Sci. USA 77, 7204-7208.

2.2 A PRELIMINARY REPORT

Reprinted from Nature, Vol. 287, No. 5784, pp. 755-758, October 23 1980

Crystal structure analysis of a complete turn of B-DNA

Richard Wing*, Horace Drew, Tsunehiro Takano, Chris Broka, Shoji Tanaka, Keiichi Itakura† & Richard E. Dickerson

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

DNA is probably the most discussed and least observed of all biological macromolecules. Although its role in biology is a central one, with many examples such as operators and restriction sites where specific base sequences have control functions or interact with specific enzymes, the structures that DNA can adopt have been based until now only on sequence-averaged fibre diffraction patterns. Recent improvements in triester synthesis methods have made possible the preparation of sufficient homogeneous DNA of predetermined sequence for crystallization and X-ray structure analysis. We report here the first single-crystal structure analysis of more than a complete turn of right-handed B-DNA, with the self-complementary dodecamer sequence d(CpGpCpGpApApTpTpCpGpCpG) or CGCGAATTCGCG.

Two of the first crystal structure analyses of DNA prepared by the triester method, CGCGCG (ref. 1) and high-salt CGCG (ref. 2), produced variants of a totally unexpected left-handed zigzag or Z helix, of considerable structural interest but uncertain biological significance. The dodecamer sequence reported here was originally chosen to incorporate an *Eco*RI restriction endonuclease site, GAATTC, within ends related to the CGCG tetramer that was already being investigated in our laboratory. The discovery of the Z helix made this sequence potentially even more interesting as an example of two Z-compatible segments, CGCG, bracketing a Z-incompatible AATT. We fully expected the structure to contain left-handed Z ends and a right-handed or melted-out centre. Instead, we found a right-handed B helix with slightly more than 10 base pairs per turn.

Crystals were grown from a *p*H 7.5 solution of DNA with a 2:1 molar excess of magnesium acetate, and one spermine hydrochloride molecule per eight base pairs, by vapour diffusion from 20 to 30% 2-methyl-2,4-pentanediol (MPD). Crystals 3 mm × 0.7 mm × 0.5 mm could be grown in a matter of weeks. These were broken up and used for heavy atom soaking trials; more perfect small crystals were used for data collection.

The dodecamer crystallizes in space group $P2_12_12_1$ with cell dimensions $a = 24.87$ Å, $b = 40.39$ Å and $c = 66.20$ Å. An assumed crystal density of 1.5 g cm^{-3} leads to one double-stranded dodecamer per asymmetric unit in a cell 47% DNA by weight. Assuming 22 negative backbone phosphate charges per dodecamer, the above unit cell dimensions indicate that the crystals are 2.2 M in negative charge even if no account is taken of possible anions contributed by magnesium acetate. Because all of these must be counterbalanced by spermine or other cations, the dodecamer crystals are fully as 'high-salt' in local DNA environment as are CGCG (2.0 M) or CGCGCG (2.6 M, see ref. 2). If there had been a significant tendency for CGCGAATTCGCG to adopt a partial Z-helical conformation, it should have done so in the crystal. Instead, a right-handed B helix is formed, with not even an indication of local structure irregularity in the CGCG regions.

The distribution of X-ray intensities on survey precession photographs suggests a cross pattern centred on the *c* axis and shows a cluster of very strong reflections at ~3.5 Å along the *c* axis, both implying that this is close to the direction of the DNA helix axis. Intensities remain strong in all directions out to 2.9 Å, and then exhibit a rapid decline until essentially no data can be obtained beyond 1.9 Å. Of the 5,691 possible reflections to 1.9 Å resolution, 2,818 were found to have an intensity greater than 2σ and were used in the analysis. Two isomorphous heavy atom derivatives were used: *cis*-dichlorodiamino platinum (II) obtained by diffusion, and a 3-Br derivative obtained by *de novo* synthesis of the dodecamer with 5-bromocytosine in the third position along each chain. The 1-Br derivative was crystallized but proved not to be isomorphous, and the 9-Br derivative was synthesized but not needed. Isomorphism in the *cis*-Pt derivative began to fail beyond 4-Å resolution, but the 3-Br derivative remains isomorphous to 2.7 Å.

The present report describes the partially refined structure obtained from multiple isomorphous replacement (MIR) analysis at 2.7 Å (mean figure of merit 57%), followed by Jack–Levitt refinement procedures[3] using 2,725 2σ intensities between 8.0 and 1.9 Å. The current residual error or *R* factor is 24.8% for a DNA molecule of 486 atoms and 9 initial water molecules. The structure of the DNA itself is essentially correct and is reported now because of its general interest. Refinement will continue with the addition of more solvent and spermine atoms, and some improvement in local nucleotide conformations.

A skeletal drawing of CGCGAATTCGCG is presented in Fig. 1, and a space-filling version from the same orientation in Fig. 2. The molecule is a Watson–Crick B helix with an average rise per residue of 3.4 Å, and 10.1 base pairs per turn. This is somewhat less than the 10.4 ± 0.1 base pairs per turn as measured in solution by Wang[4] or calculated by Levitt[5] from energy considerations.

The deoxyribose rings in our structure have been fitted to electron density in the MIR map and then allowed to refine; no attempt has been made to impose a uniform C2′-*endo* or C3′-*exo* geometry on them. At this intermediate stage of refinement they are about equally distributed between C2′-*endo* and the O1′-*endo* configuration that is related by rocking about the glycosyl bond, but no conclusions about sugar puckering should be drawn until refinement is completed. No sequence-dependent variations in the uniform B helix have been seen, but these will be watched for as refinement continues.

Even at this stage of the analysis, we can be confident about two interesting departures from classical B-helix geometry: each base pair has a propeller twist that increases the overlap between one base and its neighbours up and down the same chain, and the overall helix axis is slightly curved, concave to the right in Fig. 1. Levitt has proposed such a propeller twist in base pairs from energy calculations[5], and has found it to be persistent and not especially sensitive to the particular choice of energy parameters for minimization. Moreover, Hogan *et al.*[6] have found direct experimental evidence for propeller twist in solution from their transient electric dichroism measurements. This agreement gives us confidence that the DNA conformation we observe in the crystal is very close to that which occurs in solution.

The axis of the dodecamer helix bends by 19° over 11 base steps, for a radius of curvature of 112 Å. Moreover, the plane of bending does not contain the internal 2-fold symmetry axis of the base sequence; it is roughly 90° away from it (Fig. 1). This non-use of sequence symmetry means that equivalent positions in the two chains experience different local stresses, and suggests that the curvature is induced by external influences rather than being inherent in the molecule. Little energy would be required to produce such curvature. If the molecule is regarded as an elastically deformable rod, the energy required for uniform bending is proportional to the rod length and inversely proportional to the square of the radius of curvature[7,8]. The proportionality constant can be evaluated from the persistence length of DNA in solution, and lies between 85 and 170 kcal Å mol^{-1} (refs 5, 8). This means that only 0.25–0.50 kcal of energy would be required per mol of dodecamer to produce the deformation

* Present address: Department of Chemistry, University of California at Riverside, Riverside California 92521.
† City of Hope National Medical Center, Duarte, California 91010.

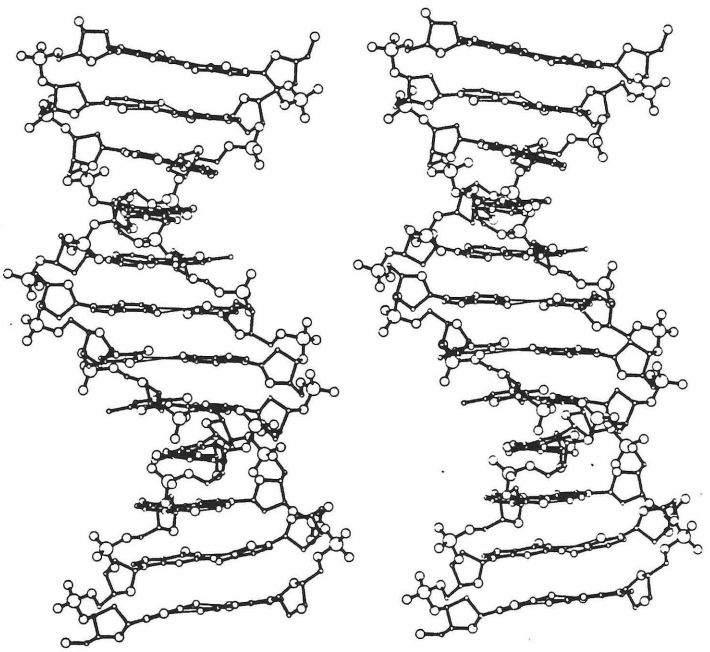


Fig. 1 Skeletal stereo drawing of CGCGAATTCGCG looking into the wide groove of the B double helix. Bases 1 and 24 are at the upper right and left in this view, and 12 and 13 are at the lower right and left. Note the uniform propeller twist in base pairs so as to maximize stacking contact with adjacent bases on the same strand, and the gentle curvature of the helix axis, concave to the right. For comments see text. Atoms in increasing order of size are C, N, O and P.

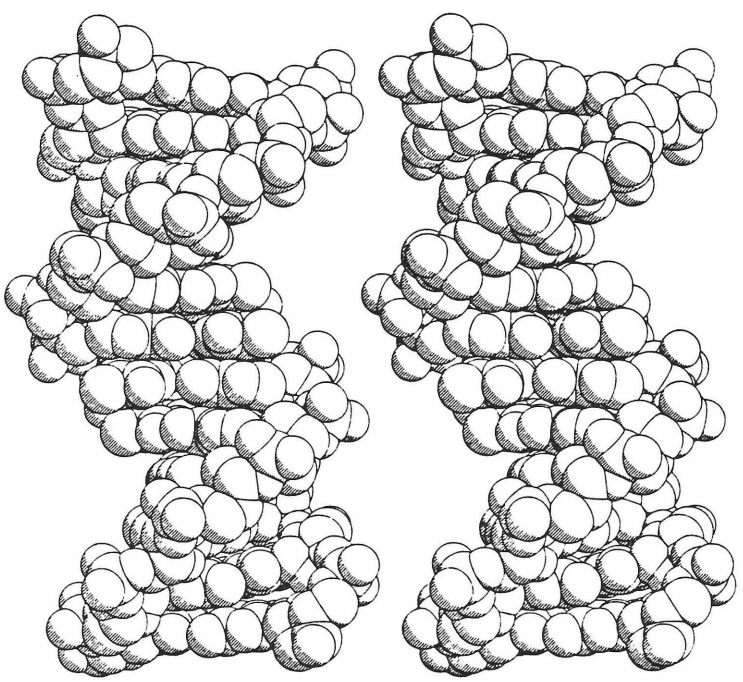


Fig. 2 Space-filling representation of CGCGAATTCGCG from the same viewpoint as Fig. 1. Atomic radii have been set at 85% of their true value for clarity. The advantage of propeller twist in maximizing adjacent base stacking can be seen in this packing drawing.

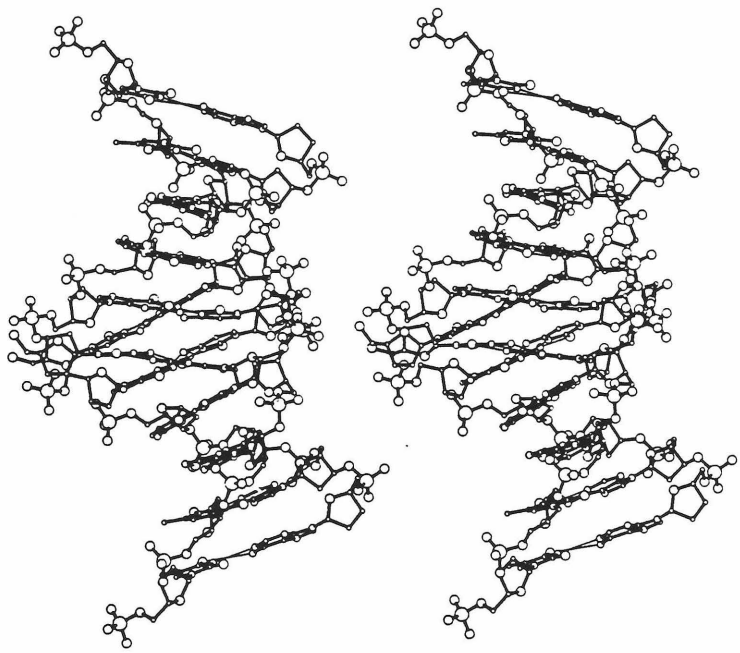


Fig. 3 Overlap of ends of molecules related by 2_1 axis along c (vertical). The upper half of one molecule (lower half not shown) sits in front of the lower half of another molecule (upper half omitted), with narrow grooves interlocked but with base planes tilted 35° to one another. The lower molecule shows bases 19–24 in front and 6–1 behind, and the upper molecule has bases 13–18 in front and 12–7 behind. The final base pair in each helix is packed parallel to, and in van der Waals contact with, the sugar–phosphate backbone of its neighbour. The narrow grooves at the ends of the molecules fit against one another like the palms of two cupped hands.

seen in Fig. 1. As a comparison, the barrier to rotation about a single bond in ethane is nearly an order of magnitude greater, 3 kcal mol^{-1}. As crystal packing forces in proteins do frequently lead to different side-chain conformations in crystals where the asymmetric unit contains two molecules[9,10], such forces would be more than adequate to account for the observed curvature.

The probable origin of the crystal packing forces that produce the curvature is to be found in the association of neighbouring molecules up the 2_1 screw axis parallel to c as shown in Fig. 3. The molecules are not stacked on top of one another like cylindrical drums, instead they are staggered, with each molecule overlapping by three base pairs with its neighbours above and below. The overlap involves contact between minor grooves, with hydrogen bonds connecting ring N3 and amino N2 atoms in adjacent guanines from the two helices. (This pairing of guanines in the minor groove has been encountered in the structure of the complex of 9-ethylguanine and 1-methylcytosine[11], and has been proposed as a possible model for the

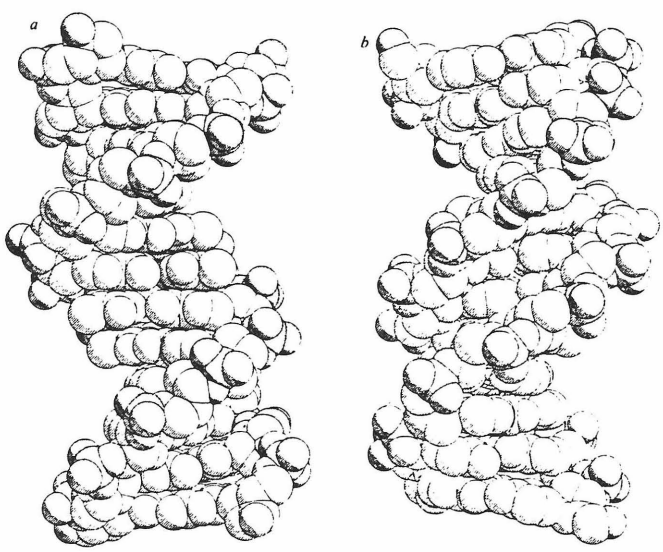


Fig. 4 Comparison of the wide groove (*a*) and narrow groove (*b*) in the B-DNA helix of CGCGAATTCGCG. Phosphate groups are shaded.

recognition of guanine by asparagine or glutamine side chains of a protein (Fig. 3 of ref. 12).) Specifically, the N2 and N3 of guanine 14 in the upper molecule are hydrogen bonded to N3 and N2 of guanine 24 in the lower, and N2 and N3 of guanine 12 in the upper molecule are bonded to N3 and N2 of guanine 2 in the lower. However, base pairs in neighbouring molecules are sharply canted or tilted relative to one another by 35°, a tilt that brings the last base pair of one helix parallel to, and in packing contact with, the sugar–phosphate backbone of the other helix. The guanine N2 . . . N3 hydrogen bonds are twisted but not broken, and the tilted structure is locked into place by one more hydrogen bond—from the 3′-terminal OH of the upper helix in fig. 3 to the N2 of guanine 22 in the lower helix. The extra stability provided by these five bonds is more than enough to produce a 19° curvature in the helix axis. This interlocking of helix termini, a function of the CGCG sequences with which the molecule begins and ends, is probably also responsible for the unusual ease and rapidity with which crystals of CGCGAATTCGCG can be grown, in comparison with other DNA molecules whose crystallization has been attempted in our laboratory and elsewhere.

The contrast between major and minor from grooves can be appreciated from Figure 4, in which the molecule is viewed from diametrically opposed directions. The ratio of widths of major and minor grooves along the helix axis is almost exactly 7 : 4, and the minor groove gives, in this space-filling model, the impression of being quite constricted.

The mode of binding of *cis*-dichlorodiamino platinum (II) to DNA is of considerable medical interest because the complex shows promise as a carcinostatic agent in cancer chemotherapy[13,14]. The *cis*-Pt site in the dodecamer is in the vicinity of N7 and O6 of guanines 4 and 16 in the wide groove of the helix. As the *cis*-Pt complex is isomorphous with the native dodecamer only out to 4 Å resolution, it will be refined as an independent structural problem. The *trans*-Pt complex is even less isomorphous as judged by cell dimensions and intensity changes. All three refined structures will ultimately be reported, as a study of the mode of binding of platinum complexes to DNA.

CGCGAATTCGCG was originally synthesized as a substrate for the *Eco*RI restriction endonuclease. Its presence in solutions from which crystals of the endonuclease are being grown leads to a change in crystal form from that of the enzyme alone, fostering the hope that the crystals in the absence of Mg^{2+} contain an abortive binary complex of dodecamer and enzyme suitable for X-ray analysis. If Mg^{2+} is present, the dodecamer is cleaved well by the enzyme (J. Rosenberg, personal communication).

The B helix has been shown here to be the stable conformation of CGCGAATTCGCG in the crystalline state in salt conditions for which CGCG and CGCGCG adopt the left-handed Z configuration. Moreover, the value of 10.1 base pairs per turn and the propeller twist of bases indicate that this crystal structure is close to that which exists in solution. Although the bend in helix axis is probably induced by crystal packing, its presence illustrates the flexibility of the B-DNA structure, and the ease with which the molecule can be continuously deformed to wrap around objects such as the histone core in nucleosomes. Details of the hydration of B-DNA will be of interest as the transitions between B and other forms are generally considered to be a function of water activity[2]. These details, and any sequence-specific modifications of the uniform helix, will be matters of prime concern as refinement continues.

This work was supported by NIH grants GM-12121 and GM-24393, and NSF grant PCM79-13959. H. D. was also the recipient of an NIH Predoctoral Traineeship.

Note added in proof: Further refinement to an R factor of 17.8% has led to a regularization of the O5′-C5′-C4′-C3′ backbone torsion angles to a standard *gauche*$^-$ (about +60°) conformation.

Received 28 July; accepted 10 September 1980.

1. Wang, A. H.-J. *et al. Nature* **282**, 680–686 (1979).
2. Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. *Nature* **286**, 567–573 (1980).
3. Jack, A. & Levitt, M. *Acta crystallogr.* **A34**, 931–935 (1978).
4. Wang, J. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 200–203 (1979).
5. Levitt, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 640–644 (1978).
6. Hogan, M., Dattagupta, N. & Crothers, D. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 195–199 (1978).
7. Schellman, J. A. *Biopolymers* **13**, 217–226 (1974).
8. Camerini-Otero, R. D. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1708–1712 (1978).
9. Tulinsky, A., Vandlen, R. L., Morimoto, C. N., Mani, N. V. & Wright, L. H. *Biochemistry* **12**, 4185–4192 (1973).
10. Takano, T. & Dickerson, R. E. in *Interaction Between Iron and Proteins in Oxygen and Electron Transport* (ed. Chien Ho) (Elsevier, Amsterdam, in the press).
11. O'Brien, E. J. *Acta crystallogr.* **23**, 92–106 (1967).
12. Seeman, N. C., Rosenberg, J. M. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 804–808 (1976).
13. Khan, A. (ed. *Proc. 3rd int. Symp. Platinum Coordination Complexes in Cancer Chemotherapy*) Vols 1, 2 (Wadley Institutes of Molecular Medicine, Dallas, 1977).
14. Dorr, R. T. & Fritz, W. L. *Cancer Chemotherapy Handbook* (Elsevier, Amsterdam, 1980).

2.3 CONFORMATION AND DYNAMICS IN B-DNA

Structure of a B-DNA Dodecamer

I. Conformation and Dynamics

HORACE R. DREW, RICHARD M. WING[1], TSUNEHIRO TAKANO,
CHRISTOPHER BROKA[2], SHOJI TANAKA[3],KEIICHI ITAKURA*
and RICHARD E. DICKERSON

California Institute of Technology, Pasadena,
California U.S.A. 91125

[1]Present address: Department of Chemistry, University of California at Riverside, Riverside, California 92521

[2]Present address: Department of Chemistry, University of California at San Diego, La Jolla, California 92093

[3]Present address: Suntory Institute for Biomedical Research, Osaka, Japan

Submitted to the Proceedings of the National Academy of Science

Accepted December 1980

ABSTRACT

The crystal structure of the synthetic DNA dodecamer d(CpGpCpGpApApTpTpCpGpCpG) or CGCGAATTCGCG has been refined to a residual error of R = 17.8% at 1.9 Å resolution (two-sigma data). The molecule forms slightly more than one complete turn of right-handed double-stranded B helix. The two ends of the helix overlap and interlock minor grooves with neighboring molecules up and down a 2_1 screw axis, producing a 19° bend in helix axis over the 11 base pair steps of the dodecamer. In the center of the molecule where perturbation is least, the helix has a mean rotation of 36.9° per step, or 9.8 base pairs per turn. The mean propellor twist (total dihedral angle between base planes) between A/T base pairs in the center of the molecule is 17.3°, and that between C/G pairs on the two ends averages 11.5°. Individual deoxyribose ring conformations, as measured by the C5'-C4'-C3'-O3' torsion angle δ, exhibit an approximately Gaussian distribution centered around the C1'-exo position with δ_{avg} = 123° and a range of 79° to 157°. Purine sugars cluster at high δ values, and pyrimidine sugars at lower δ. A tendency toward twofold symmetry in sugar conformation about the center of the molecule is detectable in spite of the destruction of ideal twofold symmetry by the molecular bending. More strikingly, sugar conformations of paired bases appear to follow a Principle of Anticorrelation, with δ values lying approximately the same distance to either side of the central δ = 123° value. This same anticorrelation is also observed in other DNA and DNA/RNA structures.

1. Introduction

In the twenty-seven years since a double-helical model for B-DNA was first proposed by Watson and Crick (1), direct evidence for its structure has been based on refinement of models having standard bond parameters against x-ray diffraction data from oriented fibers (2-5). This has had two disadvantages: loss of information because of rotational disorder about the fiber axis, and lack of information about the effect of specific base sequences, aside from a limited number of experiments on homopolymers and alternating copolymers. Recent advances in triester methods of DNA synthesis have made possible the preparation of molecules of predetermined base sequence, in quantities and purities suitable for single-crystal x-ray analysis. This paper presents the results of the first structure analysis and refinement of a complete turn of right-handed B-DNA, with the sequence: d(CpGpCpGpApApTpTpCpGpCpG) or CGCGAATTCGCG. A preliminary report of the partially refined structure has already appeared (6).

Our structure analysis of the dodecamer CGCGAATTCGCG is a logical extension of the earlier analyses of CGCGCG (7) and CGCG (8). Both of the latter adopted a left-handed zigzag or Z helix, a totally unexpected conformation whose relevance to biological DNA of more varied sequence became particularly interesting. In December 1979 we were fortunate in growing large single crystals of dodecamer, synthesized at Caltech and the City of Hope. This sequence is of particular significance both because it contains an EcoRI restriction site, GAATTC, and because it brackets a Z-incompatible AATT segment with two Z-compatible CGCG ends,

offering a test of the tendency of mixed-sequence DNA to adopt the Z conformation. In spite of favorable salt conditions, the central AATT segment apparently is sufficient to counteract the Z-forming tendencies of the tetrameric CGCG ends, resulting in a classical Watson-Crick B helix throughout.

Structure Analysis and Refinement

The dodecamer crystallizes in space group $P2_12_12_1$ with a = 24.87 Å, b = 40.39 Å, c = 66.20 Å, and two single strands or one double helix per asymmetric unit. Two isomorphous heavy atom derivatives were obtained, by de novo triester synthesis with 5-bromodeoxycytidine at the third position along each strand, and by diffusion of the anticancer agent cisplatin (cis-dichlorodiamminoplatinum(II)) into pregrown crystals. Phases from this analysis were used to obtain a starting model for restrained least-squares refinement (9).

Initial energy parameters for DNA as obtained from Michael Levitt were modified so that conformational energy would restrain the molecule to a sterically acceptable structure, but would allow the x-ray data to determine this structure. Internal sugar ring bond angles were set to the average of their C2'-endo and C3'-endo values, but were left flexible enough that other conformations could be obtained. The residual error or R factor for 2725 two-sigma data between 8.0 and 1.9 Å was reduced from 42% to 18.1% by 50 cycles of position and temperature factor refinement, during which process ten superimposed Fourier/Difference Fourier maps were inspected in order to make manual corrections and

introduce solvent molecules. At the end of the 50 cycles, all 5534 zero-sigma data between 8.0 and 1.9 Å were included in refinement. (Data out to 2.2 Å resolution were measured at least twice on different crystals.) After 62 cycles the final R factor for 486 DNA atoms and 80 ordered solvent molecules is 23.9% for the complete zero-sigma data or 17.8% for the two-sigma data (calculated with the same parameter set). The worst bond length in the structure deviates by 0.03 Å from its ideal value, and the worst bond angle deviates by 4.4°. Both intensity data and final coordinates have been deposited with the Brookhaven Protein Data Bank.

Nonuniform Motion in the Helix

The refined dodecamer structure is shown in Figure 1. It is a right-handed, Watson-Crick B double helix with an average of 10.1 base pairs per turn over the entire helix. Two striking departures from the simplest classical B helix, disucussed in the preliminary report (6), are the propellor twist of each individual base pair (total dihedral angle between base planes) and the 19° bend in helix axis over 11 base pair steps. This bending is to the right in Figure 1a, and concave toward the viewer in Figure 1b. As discussed earlier (6), this bending, although probably induced by hydrogen bonds between molecules in the crystal, requires less than one-half kilocalorie of energy per mole relative to a straight helix, and illustrates the inherent flexibility of the DNA double helix.

The double-helical structure is depicted in Figure 1 in a manner that illustrates the relative displacement of atoms as calculated from individual isotropic temperature factors, B. The larger the atom, the greater is its mean displacement in the crystal structure, whether this is dynamic (thermal vibration) or static (positional variation between crystallographically equivalent molecules). Deoxyribose and phosphate group atoms on the outside of the helix are less restricted to a fixed position (B_{avg} = 42 and 51, respectively) than are atoms within base pairs nearer the helix center (B_{avg} = 28). Part of this effect may arise from the inherent flexibility of a deoxyribose ring. Lacking a hydroxyl group on the 2' carbon atom, it does not experience the steric clash that a ribose ring does in shifting between C3'-endo and C2'-endo conformations (10). For comparison purposes, in the only available example of a RNA polymer, tRNA (11), the temperature factors are more similar between phosphate groups (B_{avg} = 48) and base pairs (B_{avg} = 39).

The motion that we see in the crystal is entirely consistent with, but not necessarily identical to, the coupled sugar-phosphate motion proposed to explain the nanosecond NMR relaxation times of DNA in solution (12). If the displacement seen in Figure 1 is interpreted as molecular vibration, then the sugar-phosphate backbone on the outside vibrates with a greater amplitude than does the core of the helix, and base pairs at the upper and lower ends of the helix move more than those at the center. Atoms involved in the outer hydrogen bonds of any one base pair frequently are vibrating more than those in the central hydrogen bond, consistent with a propellor twist motion. In many cases the

deoxyribose ring appears to be rocking about the C1'-N bond to the base, as suggested by the average B values of 34 for C1', 40 for C2', 42 for C3' and O1', and 44 for C4'. Such a C1'-N rocking motion could lead to a continuum of sugar conformations between C2'-endo and O1'-endo and, as will be seen in the following section, this is indeed what is observed in the dodecamer. The reasonableness of the vibrational interpretation of the displacements in Figure 1 is sufficient to cause one to favor this explanation over the alternative one of static disorder within the crystal.

Variations in Sugar Conformation

Glycosyl and main-chain torsion angles for the refined dodecamer are listed in Table 1, along with ideal values that have been proposed for A- and B-DNA (5,14). The main chain torsion angles α through ζ are very close to the (gauche$^-$, trans, gauche$^+$, trans, trans, gauche$^-$) values expected from an ideal B helix (10), with the greatest variation in the C5'-C4'-C3'-O3' torsion angle δ. The conformation of a deoxyribose ring is closely related to this latter angle in the manner shown along the top end of Figure 2 (13). For the dodecamer, the clarity of definition of the C5'-C4'-C4'-O3' backbone chain in the electron density map permits the establishment of δ to within roughly 10°. The shape of each deoxyribose ring in the electron density map is compatible with the ring conformation predicted by torsion angle δ, but the precision with which δ can be determined makes it the better guide to sugar conformation at less than atomic resolution (8).

Figure 2 also illustrates the strong correlation between glycosyl angle χ and backbone torsion angle δ. The observed conformations are scattered along a diagonal line in the plot, in a roughly Gaussian distribution about point B, which is one of the models for the B helix obtained by Levitt from energy considerations (14). This point corresponds to the C1'-exo conformation, which was also encountered at each guanine position in left-handed Z-helical CGCG (8). The C2'-endo conformation of a classical B helix is represented by point B_F at the upper right of the plot, and the C3'-endo conformation of an A helix is at the far lower left, A_F, beyond the cluster of points from the dodecamer conformation.

One of the most striking aspects of Figure 2 is the preference of purine sugars for high δ and χ values near the C2'-endo conformation, and of pyrimidine sugars for low values nearer O1'-endo, especially if the C1-G24 base pair at the upper end of the helix is ignored as representing end effects. Examination of space-filling models suggest that a close contact exists between the H atom of a sugar C1' and the O2 of a pyrimidine at $\chi = -120°$. This contact cannot be relieved by making χ less negative without introducing other clashes between the pyrimidine H6 and the hydrogens attached to either C2' or C3', but a shift to more negative χ eliminates the clash. In contrast, the larger angle involved in the connection of a sugar to the five-membered ring of a purine keeps the N3 and H8 atoms far enough away from the sugar ring that no close contacts result. Hence the purine sugars are free to adopt higher χ values and an ideal C2'-endo conformation.

In view of the gentle slope of the total energy curve in the vicinity of point B (Figure 2 of Reference 13), and the evidence in Figure 1 for displacements, probably vibrational, in the sugar-phosphate backbone chain, the scatter in individual conformations seen in Figure 2 is hardly surprising. The "ideal" B helix in DNA is as much an oversimplification as is the "ideal" α helix in proteins. The DNA is not aware of the discrete conformational states listed at the top of Figure 2, only of a smooth and rather shallow potential well centered in the vicinity of point B. Visual inspection of the stereo drawings indicates that conformations can best be described as C2'-endo for all sugars to the right of δ = 129° in Figure 2, C1'-exo for sugars between 105° and 129°, and O1'-endo for sugars 15, 3 and 7, with one lone C3'-endo or A-like sugar puckering at the 3' terminus of the second strand. (This assignment is necessarily arbitrary and subjective at intermediate angles.)

Two other features of the distribution in Figure 2 deserve special mention. Sugars related by an approximate twofold axis through the center of the molecule perpendicular to the helix axis (1 and 13, 2 and 14, 3 and 15, etc.) tend to have similar δ and χ values, indicating local preservation of twofold symmetry in spite of the overall 19° bend in the helix axis visible in Figure 1a (and discussed in Reference 6). Such is to be expected, and the only unusual feature is that this conformational symmetry is preserved in spite of the bending of the helix axis.

The second correlation is more surprising, and potentially of more fundamental significance. The conformations of sugars in paired bases (1/24, 2/23, 3/22, etc.) tend to be anticorrelated: if one conformation lies to the left of center in Figure 2, the other conformation lies a similar distance to the right of center. This is a stronger statement than the previous observation that purines generally prefer higher δ and χ values than pyrimidines do. If the midpoint in δ is found for each base pair, these midpoints have a mean value of 122.8°, near the C1'-exo conformation, with a mean deviation of only 5.6°. Some base pairs such as 5/20, 6/19 and 2/23 have similar sugar conformations centered closely about C1'-exo. Others such as 1/24, 10/15 and 3/22 have quite disparate conformations, one of them nearer C2'-endo or C3'-exo, and the other near O1'-endo or even C3'-endo. If the difference between δ values for two paired bases is defined as the conformational spread, then the C/G base pairs have a larger mean spread (36.9°) than do the A/T base pairs (20.8°), although the statistical significance of this is not clear with only 12 base pairs to compare.

This behavior of sugar conformations in paired bases is sufficiently striking to be formalized by defining it as the Principle of Anticorrelation: Deoxyribose sugars attached to paired bases in B-DNA tend to adopt δ values that are equidistant to either side of a central δ = 123° (or C1'-exo) value. It probably is a consequence of wrapping sugar-phosphate chains having flexible deoxyribose rings around base pairs of fixed dimensions. One would not necessarily

expect to find anticorrelation in sugar conformations of paired bases in RNA, with its more constrained ribose ring, and indeed such anticorrelation is not present in the one available example, tRNA (15). If the midpoints in δ values for paired bases in tRNA are calculated, omitting the first base pair at either end of a double-helical stack to eliminate end effects, then the mean of these values is 82.8° and the mean deviation is 2.4°. But the mean spread or difference in δ values between paired bases is only twice this, 5.4°. The half-spread from the midpoint is no larger than the uncertainty in midpoint position, indicating no statistically significant anticorrelation in sugar conformation. By comparison, in the DNA dodecamer CGCGAATTCGCG the mean is 122.8°, the mean deviation is 5.6°, and the mean spread over all base pairs is 31.5°, six times as great.

The Anticorrelation Principle also is observed in the nonintercalated TA base pair step of the 2:1 intercalation complex of daunomycin with CGTACG (16). Sugar pucker information provided in Reference 16 indicates that the C's and G's bracketing the intercalator molecules would be scattered at the upper right of Figure 2 with the exception of an anomalous C1 conformation at the beginning of the chain. But the unintercalated T3-A4 is a beautiful example of anticorrelation. (Since the complex has an internal twofold axis of crystallographic symmetry, the two strands are identical. T3 on one strand is followed by A4 on the same strand, and hydrogen bonded to A4 on the opposite strand.) T3 has the (χ,δ) conformation (-131°, ca. 96°), and A4 has

(-107°, ca. 144°).* In addition, the tendency, noted by both Sobell (17) and Rich (18), for many intercalating groups (but not daunomycin) to induce a (C3'-endo)-3',5'-(C2'-endo) conformation at the bases bracketing the intercalator, whether in DNA or RNA helices, can be regarded as an extreme limit of the Principle of Anticorrelation. Zimmerman and Pheiffer (19) have recently found an RNA-DNA hybrid, poly(rA)·poly(dT), which can be induced to adopt a B helix under conditions of high humidity. Its δ values as deduced from fiber diffraction patterns are 97° and 152° for rA and dT, respectively. The midpoint of these values, 124.5°, is only 1.7° away from the mean DNA dodecamer value.

Local Helix Parameters

A helix-generating program was used to determine the rotation angle, orientation of the local helix rotation vector, and rise per base pair along the local axis, for each of the 11 base pair steps along the dodecamer. The individual rotation angles and rise per base pair are listed in Table 2, along with the measured propeller twist at each base pair. These numbers probably are more accurate than any other results of the analysis, since these planes are well defined in the map and are established by a large number of atomic positions. The propellor twist of A/T base pairs in the center of the molecule is a remarkably

---

*The χ values quoted in reference 16 should have 180° subtracted from each to bring them into accord with IUPAC-IUB conventions.

uniform 17.3° (0.4°) (mean value, with standard deviation in parentheses), and much of the variation at the two ends probably arises from overlapping of helices up the c axis (Figure 3 of Reference 6). The smaller mean value of 11.5° (5.1°) for the two CGCG ends could also reflect the flattening influence of the third hydrogen bond in each base pair.

The individual helical twist or rotation values for the five central base pair steps that are unperturbed by overlapping ends (from G4 to C9) give an average of 36.9° (3.3°), corresponding to 9.8 base pairs per turn. The three base pair steps at either end are somewhat wound, averaging 37.6° (4.5°), or 9.6 base pairs per turn. The overall value of 35.8° or 10.1 base pairs per turn that was reported in Reference 6 arises because several of the local twist vectors in the overlapping ends of the molecule are sharply tilted from the mean helix axis, enlarging their twist angles in projection. Space does not permit a detailed discussion of local and global helix twist vectors here, but this is the subject of a companion paper (22). It can only be mentioned here that these displacements in local helix rotation vectors show a twofold symmetry about an axis through the center of the molecule, perpendicular to the page in Figure 1a or horizontal in Figure 1b. This ideal symmetry axis is expected from the identity of the two strands, but is destroyed by the 19° bend in the helix axis. Hence these local variations in helix twist vector are observed in spite of the bend in the helix, and not because of it.

Acknowledgements

We would like to acknowledge Peter Dembek for his help in synthesis of the dodecamer. We would like to thank Lillian Casler for her assistance in preparing the Figures, and Charles Ray and the Caltech computing center for their willingness to give special attention to quality control in preparation of the stereo drawings.

This work was carried out with the support of NIH grants GM-12121 and GM-24393, and NSF grant PCM79-13959. H.D. was also the recipient of an NIH Predoctoral Traineeship. This is contribution No. 6317 from the Division of Chemistry and Chemical Engineering.

References

1. Watson, J. D. and Crick, F. H. C. (1953) Nature 171, 737-738.
2. Langridge, R., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. and Hamilton, L. D. (1960) J. Mol. Biol. 2, 19-37.
3. Fuller, W., Wilkins, M. H. F., Wilson, H. R., Hamilton, L. D. and Arnott, S. (1965) J. Mol. Biol. 12, 60-80.
4. Arnott, S. and Hukins, D. W. L. (1972) Bioch. Bioph. Res. Commun. 47, 1504-1509.
5. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. and Ratliff, R. L. (1980) Nature 283, 743-745.
6. Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 287, 755-758.
7. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., Van Boom, J. H., Van der Marel, G. and Rich, A. (1979) Nature 282, 680-686.
8. Drew, H.R., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 286, 567-573.
9. Jack, A. and Levitt, M. (1978) Acta Crystallogr. A34, 931-935.
10. Sundralingam, M. (1975) in Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions (M. Sundaralingam and S. T. Rao, Eds.), University Park Press, Baltimore, pp. 487-524.
11. Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. and Kim, S.-H. (1978) J. Mol. Biol. 123, 607-630.
12. Hogan, M.E. and Jardetzky, O. (1980) Biochemistry 19, 3460-3468.
13. Levitt, M. and Warshel, A. (1978) J. Am. Chem. Soc. 100, 2607-2613.

14. Levitt, M. (1978) Proc. Natl. Acad. Sci. USA 75, 640-644.

15. Holbrook, S. R., Sussman, J. L., Warrant, R. W. and Kim, S.-H. (1978) J. Mol. Biol. 123, 631-660.

16. Quigley, G. J., Wang, A. H.-J., Ughetto, G., van der Marel, G., van Boom, J. H. and Rich, A. (1980) Proc. Natl. Acad. Sci. USA 77, 7204-7208.

17. Jain, S. C., Tsai, C.-C. and Sobell, H. M. (1977) J. Mol. Biol. 114, 317-331.

18. Rich, A., Quigley, G. J. and Wang, A. H.-J. (1979) in Stereodynamics of Molecular Systems (R. H. Sarma, Ed.), Pergamon Press, New York, pp. 315-330. or: Rich, A., Wang, H.-J. and Quigley, G. J. (1980) in Frontiers of Bioorganic Chemistry and Molecular Biology (S. N. Ananchenko, Ed.), Pergamon Press, New York, pp. 327-337.

19. Zimmerman, S. B. and Pheiffer, B. H. (1980) Proc. Natl. Acad. Sci. USA 78, 78-82.

20. Arnott, J., Dover, S. D. and Wonacott, A. J. (1969) Acta Crystallogr. B25, 2192-2206.

21. Arnott, J. and Selsing, E. (1975) J. Mol. Biol. 98, 265-269.

22. Dickerson, R. E. and Drew, H. R. (1981) J. Mol. Biol.--submitted for publication.

Table 1

MAIN CHAIN AND GLYCOSYL CONFORMATION ANGLES (°)

Residue	χ	α	β	γ	δ	ε	ζ	Notes
C1	-105	--	--	174	157	-141	-144	
G2	-111	-66	170	40	128	-186	- 98	
C3	-135	-63	172	59	98	-177	- 88	
G4	- 93	-63	180	57	156	-155	-153	
A5	-126	-43	143	52	120	-180	- 92	
A6	-122	-73	180	66	121	-186	- 89	
T7	-127	-57	181	52	99	-186	- 86	
T8	-126	-59	173	64	109	-189	- 89	
C9	-120	-58	180	60	129	-157	- 94	
G10	- 90	-67	169	47	143	-103	-210	
C11	-125	-74	139	56	136	-162	- 90	
G12	-112	-82	176	57	111	--	--	
C13	-128	--	--	56	137	-159	-125	
G14	-116	-51	164	49	122	-182	- 93	
C15	-134	-63	169	60	86	-185	- 86	
G16	-115	-69	171	73	136	-186	- 98	
A17	-106	-57	190	54	147	-183	- 97	
A18	-108	-57	186	48	130	-186	-101	
T19	-131	-58	174	60	109	-181	- 88	
T20	-120	-59	179	55	122	-181	- 94	
C21	-114	-59	185	45	110	-177	- 86	
G22	- 88	-67	179	50	150	-100	-188	
C23	-125	-72	139	45	113	-174	- 97	
G24	-135	-65	171	47	79	--	--	
Avg.	-117	-63	171	54*	123	-169	-108	
Std.Dev.	(14)	(8)	(14)	(8)	(21)	(25)	(34)	
B DNA	-119	-61	180	57	122	-187	- 91	(1)
B_F DNA	-102	-41	136	38	139	-133	-157	(2)
A_F DNA	-154	-90	-149	47	83	-175	- 45	(3)

Notes to Table 1:

Main chain conformation angles are defined as:

$$P\overset{\alpha}{\text{——}}O5'\overset{\beta}{\text{——}}C5'\overset{\gamma}{\text{——}}C4'\overset{\delta}{\text{——}}C3'\overset{\varepsilon}{\text{——}}O3'\overset{\zeta}{\text{——}}P$$

with zero at the fully eclipsed or cis position and positive clockwise rotation of the farther pair of atoms. The glycosyl angle χ is similarly defined in terms of atoms O1'——C1'$\overset{\chi}{\text{——}}$N1——C2 for pyrimidines (C,T) or O1'——C1'——N9——C4 for purines (G,A).

(1) Model chosen from Levitt 1978 (14) energy refinement for best agreement with our results: 10 base pairs per turn and C——O——C deoxyribose angle set at 115°

(2) Arnott et al 1980 (5) fiber data with 10 base pairs per turn

(3) Arnott et al 1980 (5) fiber data with 11 base pairs per turn

*Omitting C1 value

Table 2

LOCAL HELIX PARAMETERS

Base pairs	Propellor twist (°) ψ	($\Delta\psi$)	Helix twist angle* (°) Θ	($\Delta\Theta$)	Base pairs per turn $n=360/\Theta$	(Δn)	Rise per base pair (Å) h	(Δh)
C1/G24	13.2	(2.0)						
			38.3	(1.1)	9.40	(0.27)	3.36	(0.01)
G2/C23	11.7	(2.1)						
			39.6	(6.1)	9.09	(1.40)	3.38	(0.08)
C3/G22	7.2	(2.1)						
			33.5	(2.1)	10.75	(0.67)	3.26	(0.05)
G4/C21	13.2	(1.9)						
			37.4	(1.7)	9.63	(0.44)	3.30	(0.10)
A5/T20	17.1	(2.1)						
			37.5	(0.9)	9.60	(0.23)	3.27	(0.02)
A6/T19	17.8	(2.1)						
			32.2	(2.1)	11.18	(0.73)	3.31	(0.03)
T7/A18	17.1	(1.9)						
			36.0	(2.8)	10.00	(0.78	3.29	(0.01)
T8/A17	17.1	(2.0)						
			41.4	(2.1)	8.70	(0.42)	3.14	(0.02)
C9/G16	18.6	(1.9)						
			32.3	(1.3)	11.11	(0.45)	3.56	(0.07)
G10/C15	4.9	(1.9)						
			44.7	(5.4)	8.05	(0.97)	3.21	(0.18)
C11/G14	17.2	(1.9)						
			37.0	(1.9)	9.73	(0.50)	3.54	(0.19)
G12/C13	6.2	(2.3)						
Averages:	13.4	[4.9]	37.3	[3.8]	9.75	[0.98]	3.33	[0.13]
A DNA (4,20)			32.7		11.0		2.56	
B DNA (4,20,21)			36.0		10.0		3.38	
C DNA (21)			38.6		9.33		3.31	
D DNA (21)			45.0		8.0		3.03	

Notes to Table 2:

Helical parameters were found using vectors between atoms C1' and the attached N of one base, and the equivalent atoms of the next base up the same chain, with a program kindly provided by John Rosenberg. Standard deviations at individual steps in parentheses, statistical variation over entire helix in brackets. Propellor twist is the dihedral angle between base planes in the same base pair, and is twice the fiber twist of reference 20 and 21.

*Rotation per base pair.

Figure Captions

FIGURE 1. "Vibration diagram" representation of the CGCGAATTCGCG double helix viewed (a) into the wide groove and (b) 90° to the right. Chains are identified in text and tables by consecutive base numbering C1...G12 for one chain and C13...G24 for the other. C1 is paired with G24, G2 with C23, ...A5 with T20, etc. Base pair C1/G24 is at the top of each helix, and G12/C13 is at the bottom. The radius of each atom has been set to $u/3$, where u is the r.m.s. displacement as obtained from its individual isotropic temperature factor parameter, $B = 8\pi^2 u^2$.

FIGURE 2. Correlation plot between glycosyl torsion angle χ and the C5'-C4'-C4'-O3' torsion angle δ (or ψ'). Each sugar is represented by a numbered circle, with angles taken from Table 1. A_F, B_F and B locate ideal A and B helices as derived from fiber diffraction (5) or energy refinement (14). The distribution of conformations is close to a Gaussian centered around point B. Purines are distinguished by heavy circles, and pyrimidines by light. Notice the correlation (similar δ values) between bases related by a twofold axis normal to the helix (1/13, 2/14, etc.), and the anticorrelation (constant δ sum) between bases that are paired together on one step of the helix (1/24, 2/23, etc.). The δ values for the ten standard endo and exo conformations, and their energies in kcal/mole relative to C3'-endo, are given along the top of the plot (13).

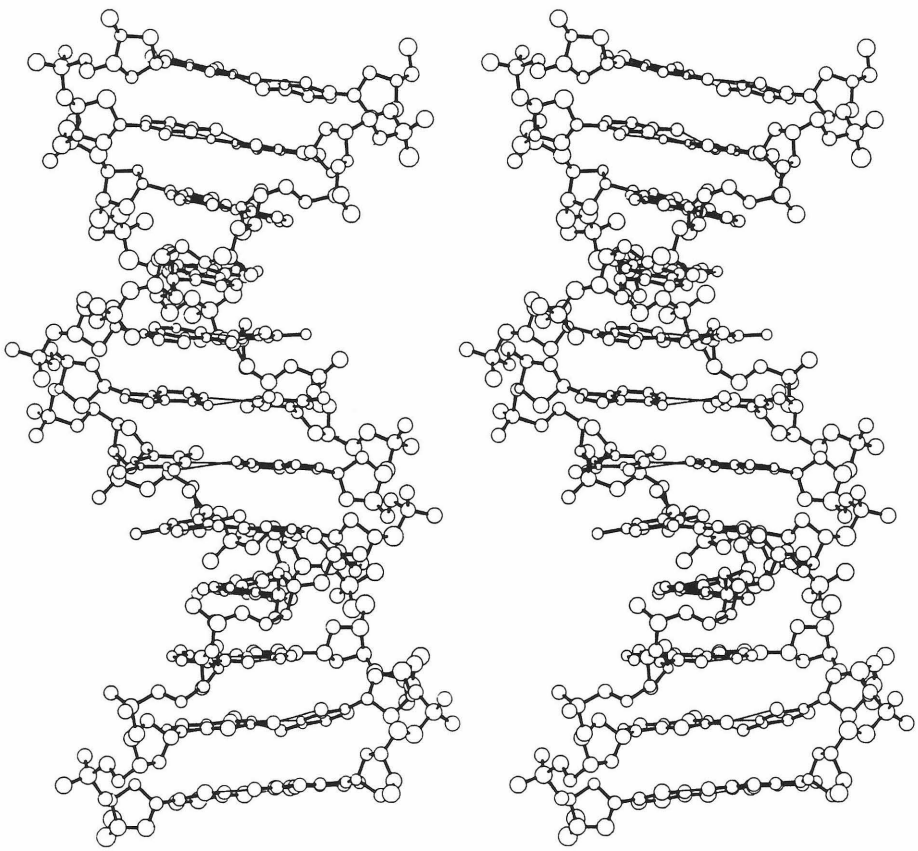


Figure 1a

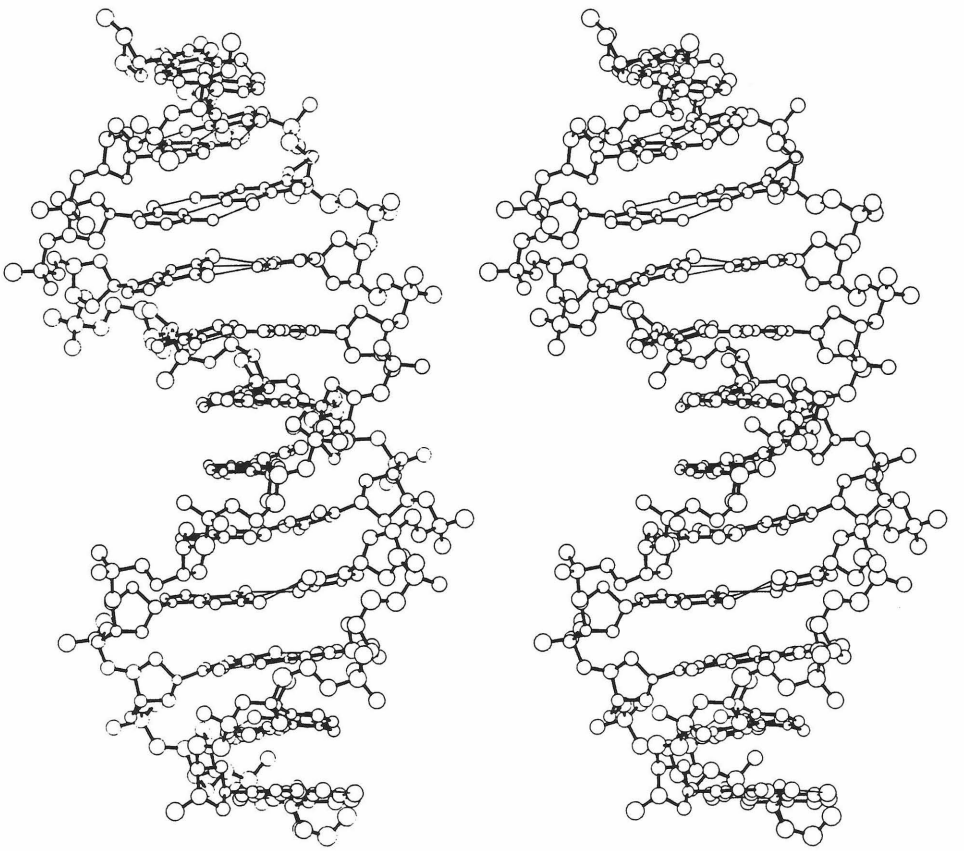


Figure 1b

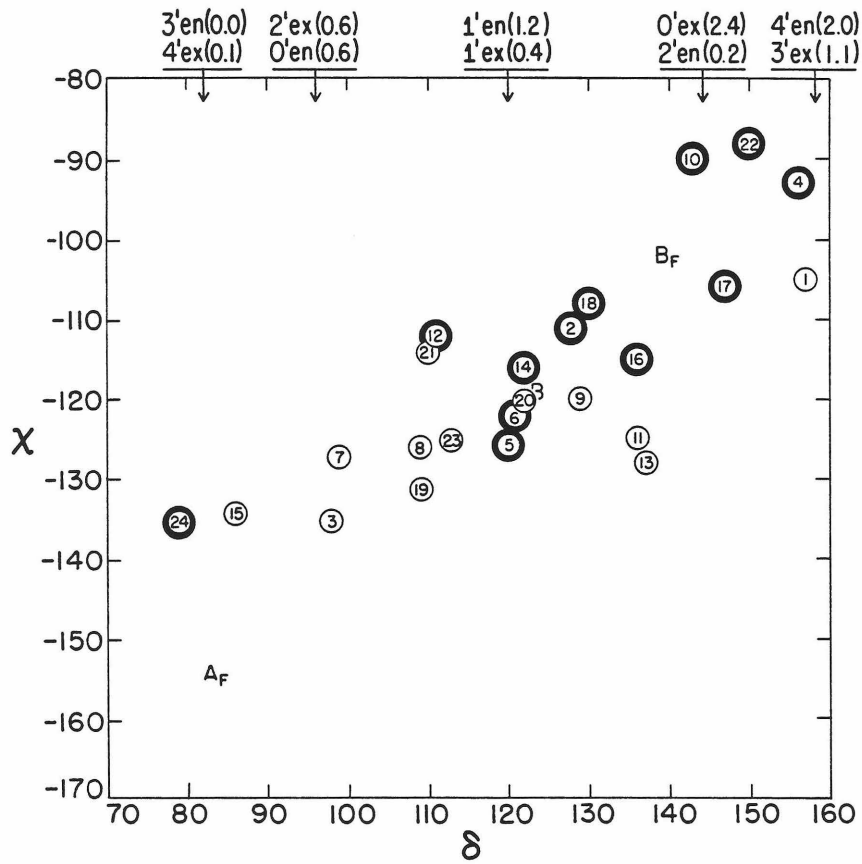

3'en(0.0)
4'ex(0.1)
2'ex(0.6)
0'en(0.6)
1'en(1.2)
1'ex(0.4)
0'ex(2.4)
2'en(0.2)
4'en(2.0)
3'ex(1.1)
-80
-90
-100
-110
-120
-130
-140
-150
-160
-170
χ
70
80
90
100
110
120
130
140
150
160
δ
A_F
B_F

Figure 2

2.4 SEQUENCE AND STRUCTURE IN B-DNA

Structure of a B-DNA Dodecamer.
II. Influence of Base Sequence on Helix Structure*

RICHARD E. DICKERSON and HORACE R. DREW

Norman W. Church Laboratory of Chemical Biology,
California Institute of Technology,
Pasadena, California 91125

Submitted to the Journal of Molecular Biology

*Paper I of this series is Drew et al (1981).

Running Title: Sequence and Structure in B-DNA

SUMMARY

Detailed examination of the structure of the B-DNA dodecamer CGCGAATTCGCG, obtained by single-crystal x-ray analysis (Drew et al (1981) Proc. Natl. Acad. Sci. USA--in press) reveals that the local helix parameters: twist, tilt and roll, are much more strongly influenced by base sequence than by crystal packing or any other external forces. The central EcoRI restriction endonuclease recognition site, GAATTC, is a B-helix with an average of 9.8 base pairs per turn. It is flanked on either side by single base pair steps having aspects of an A-like helix character. The dodecamer structure suggests several general principles, whose validity must be tested by other B-DNA analyses: (1) When an external bending moment is applied to a B-DNA double helix, it bends smoothly, without kinks or breaks, and with relatively little effect on local helix parameters. (2) Purine-3',5'-pyrimidine steps open their base planes toward the major groove, pyrimidine-purine steps open toward the minor groove, and homopolymer (pur-pur, pyr-pyr) steps resist rolling in either direction. This behavior is related to the preference of pyrimidines for more negative glycosyl torsion angles. (3) CpG steps have smaller helical twist angles than do GpC, as though in compensation for their smaller intrinsic base overlap. Data on A/T steps are insufficient for generalization. (4) G/C base pairs have smaller propellor twist than A/T, and this arises mainly from interstrand base overlap rather than the presence of the third hydrogen bond. (5) DNAse I cuts

preferentially at positions of high helical twist, perhaps because of increased exposure of backbone to attack. The correlation of digestion pattern in solution and helical twist in the crystal argues for the essential identity of the helix structure in the two environments. (6) In the two places where the sequence TpCpG occurs, the C slips from under T in order to stack more efficiently over G. At the paired bases of this CpG step, the G and C are tilted so the angle between base planes is splayed out to the outside of the helix. This TpC is the most favored cutting site for DNAse I by a factor of 4.5 (Lomonosoff et al (1981), J. Mol. Biol.--in press). (7) The EcoRI restriction endonuclease and methylase both appear to prefer a cutting site of the type: purine-purine-A-T-T-pyrimidine, involving two adjacent homopolymer triplets, and this may be a consequence of the relative stiffness of homopolymer base stacking observed in the dodecamer.

1. Introduction

The first single-crystal x-ray structure analysis of a complete turn of double-helical B-DNA, that of d(CpGpCpGpApApTpTpCpGpCpG) or CGCGAATTCGCG (Wing et al, 1980; Drew et al, 1981), has revealed many structural details that are not easily obtained by other methods of analysis. As Figure 1 illustrates, this dodecamer has a nearly classical Watson-Crick right-handed B-DNA structure. The overall helix axis has a bend of 19° in 11 base pair steps (numbered for reference in Figure 1), induced by the intermeshing of minor grooves of neighboring molecules along a 2_1 screw symmetry axis in the crystal as depicted in Figure 2. As discussed in Wing et al (1980), less than 0.5 kcal per mole of energy is required to produce such a bend in a straight helix, and this is well within the range of possible crystal forces.

The overlapping ends, at first regarded only as an undesirable perturbation of the structure that the dodecamer would have when free in solution, actually convey additional information about the B helix. In addition to making the crystals easy to grow, strong, and ordered, they allow us to see the effect on a B helix of applying a bending moment in a direction that destroys a natural twofold axis of symmetry of the isolated molecule. This lateral twofold axis, a reflection of the chemical identity of the two polynucleotide chains, would pass perpendicular to the plane of Figure 1 through the numeral 6, superimposing the upper half of an unbent helix on its lower half.

This paper addresses the following questions: What is the detailed structure of a B-DNA helix? How is this structure affected by the particular sequence of bases involved? What effect is produced by external bending forces, and to what extent is the intrinsic two-fold axis retained in spite of the bending? Are there sequence-dependent aspects of structure that might contribute to the recognition of a particular sequence by control proteins such as restriction enzymes and repressors? The sequence prepared for this x-ray analysis was chosen because it contains an EcoRI restriction endonuclease cutting site, GAATTC, in the hope that the structure would shed light on the problem of sequence recognition.

2. Local and Global Helical Twist

A view down the best linear approximation to the overall helix axis in Figure 3 reveals that the CGCGAATTCGCG molecule has an average helical twist, defined as the projected angle between C1'-C1' vectors in successive base pairs, of 35.8°, corresponding to 10.1 base pairs per turn. It was this value that initially was reported in Wing et al (1980). Individual values for this angle at the 11 base pair steps marked in Figure 1 are listed in Table 1, where they are designated as t_g for "global twist angles." (Throughout this paper, the term twist by itself will be considered synonymous with rotation per base pair step about a helical axis. Propellor twist is the total dihedral angle between base planes in one hydrogen-bonded base pair.) In

contrast to these t_g values, a local helix twist program that uses vectors between the two C1' atoms and their attached base nitrogens of one base pair, and the corresponding atoms of the next base pair, to define a local rotation and translation vector, leads to a quite different average of 37.3° per base pair or 9.65 base pairs per turn. The individual values, designated t_ℓ for "local twist angles" are also given in Table 1 along with the rise per base pair along the local helix axis. The two approaches differ in that t_g measures rotation angles projected along a best overall helix axis, whereas t_ℓ considers each set of two successive base pairs in isolation, defining rotation axis, rotation angle and translation as though the remainder of the molecule did not exist.

The apparent discrepancy between an average of 10.1 base pairs per turn from t_g and 9.65 base pairs from t_ℓ initially was a cause for concern, and attempts were made to account for it in terms of writhe and supercoiling. However, as often happens, reality is simpler than a theoretical approach would lead one to expect. The t_ℓ are the true local helical twist angles, and the increase or decrease needed to obtain t_g arises because the local rotation vector frequently is tilted relative to the overall helix axis (the direction of view in Figure 3), foreshortening or enlarging the true twist angle slightly as viewed in projection. Figure 4 shows that t_g and t_ℓ vary in the same manner along the helix, with minima at steps 3, 6 and 9, and a maximum at step 10, to be discussed in the next section.

3. Symmetry in Local Helix Rotation Vectors

The simple t_ℓ and h values listed in Table 1 do not show the directions in space of the local helix rotation axes, and these are important for an understanding of the molecule. Figure 5 displays the eleven local helix rotation vectors for individual base pair steps on a skeletal view of the dodecamer. In the center of the molecule at steps 4-8 these local vectors are nearly aligned head to tail along a single straight axis. For this central GAATTC region, which also is the restriction endonuclease site, the average local twist, t_ℓ, is 36.9°(3.3°) (mean value, with standard deviation in parentheses), corresponding to 9.8 base pairs per turn. The average global twist, t_g, is nearly identical throughout this region, 36.5°(2.9°), as would be expected for a straight helix. This value of 9.8 base pairs per turn probably can be regarded as the best overall value for an unperturbed EcoRI recognition site.

The greatest variation in local helix vectors occurs in the first and the last three steps of the dodecamer, where the vectors are both tilted and shifted relative to the central helix axis. At first one is tempted to ascribe this variation to the molecular bending produced by overlapping helix ends shown in Figure 2. But a closer examination of the actual vector displacements reveals that they exhibit a twofold symmetry perpendicular to the helix axis, of a type that actually is broken by the bending of the molecule in the crystal. The local rotation vectors of steps 3 and 9 are both shifted into the major groove

(Figure 5a) and tilted along it (Figure 5b). The vectors of steps 1 and 11 show the same effects--displacement into the major groove and alignment along it--to a lesser extent (Figure 5b). In contrast, the vectors of steps 2 and 10 are tilted in a contrary direction to those just above and below them. But vector pairs 1 and 11, 2 and 10, 3 and 9, etc. still are nearly related by a lateral twofold axis through the center of the molecule: perpendicular to the page in Figure 5a and horizontal in Figure 5b. Hence the surprising conclusion that these effects probably are observed in the crystal in spite of the bending of the molecule and not because of it. This displacement and tilting of local rotation vectors may well be present in the dodecamer molecules when isolated in solution. If so, then this represents one of the first clear-cut examples of sequence-induced variation in a B helix structure.

4. Variation in Local Helix Type

The rotations at steps 3 and 9 are particularly interesting in that they have some aspects of an ideal A helix rather than B (Arnott et al, 1974, 1975; Sarma et al, 1979). Displacement of the rotation axis farther from the base pairs into the major groove is a characteristic of the A helix (compare Figures 8 and 10 of Arnott et al, 1975). Furthermore, the base pairs at steps 3 and 9 are tilted clockwise relative to the local helix rotation axis by 33° and 21° respectively when viewed into the minor groove. A-DNA has a 19° base tilt in this

same direction. Finally, the local helical rotations t_ℓ at steps 3 and 9 correspond to 10.8 and 11.1 base pairs per turn, bracketing the 11 base pairs of an ideal A helix. One has the impression that an EcoRI restriction site in the B conformation with an average of 9.8 base pairs per turn has been terminated at either end by one base pair step in something close to an A-DNA conformation. This raises interesting questions, which cannot yet be answered, as to whether such local interruptions in helix type are a regular feature of some kinds of recognition sites in DNA.

Step 10 also is unusual in resembling a variant of the B family of helices: the eightfold D-DNA helix (Arnott et al, 1974). Its local helix rotation vector (Figure 5b) is even closer to the base pairs than for a B helix (compare Figures 10 and 13 of Arnott et al, 1975), and is tilted 25° to the base planes in a sense opposite to that of steps 3 and 9. The ideal D helix has approximately a 16° tilt in this direction. Furthermore, the t_ℓ value of 44.7° for step 10, the largest in the entire molecule, corresponds exactly to the 8 base pairs per turn proposed for D-DNA.

The only other unusual feature of local rotations occurs at step 6, where t_ℓ = 32.2°, or 11.2 base pairs per turn, suggesting an A helix. But neither the orientation nor the displacement of the local helix rotation vector supports this assignment, so the EcoRI site in the center of the molecule must be regarded as a helix of the B family (Arnott, 1976) with individual rotations corresponding to as few as 8.7 or as many as 11.2 base pairs per turn, with an average of 9.8.

In summary, eight of the eleven steps along the entire dodecamer are B-like, another has the character of a D helix (also a member of the B family), and two symmetrically related steps have local A-DNA character. These must be regarded, however, as no more than local variations in an overall B-helical framework. The distances between successive phosphorus atoms along the sugar-phosphate backbone, listed in Table 2, remain as expected of B-DNA even at the variable steps.

Although the individual twist values must inevitably be perturbed by the bending of the helix axis in the crystal, the t_ℓ values in Table 1 still show evidence of a lateral twofold symmetry axis. Symmetry-related t_ℓ for steps 1 and 11, 2 and 10, 3 and 9, etc. differ by an average of only 3% from their respective means. The 19° bend has surprisingly little effect on local helix rotation along the molecule.

5. Base Pair Stacking and DNAse I Digestion

Stereo drawings of the eleven individual base pair steps are shown in Figures 6a-k). In each case the direction of view is down the best overall helix axis of Figure 3, rather than the individual helix rotation vectors of Figure 5. The main chain and glycosyl torsion angles, and deoxyribose sugar conformations, are especially clear in these views, and have been listed for reference in Table 2. The main chain angles α through ζ generally exhibit the (gauche$^-$, trans, gauche$^+$, trans, trans, gauche$^-$) conformations expected for

B-DNA (Arnott and Hukins, 1972; Arnott *et al*, 1974). Torsion angle ζ (C3'-O3'-P-O5') nearly always occurs in the *gauche*$^-$ conformation as predicted by Arnott and coworkers, rather than the *trans* conformation proposed by Gupta *et al* (1980). Two anomalies at bases G10 and G22--a less negative ε and a more negative ζ (the *trans* conformation)--can be seen from Figures 6b and 6j to arise because of a twisting of the phosphate groups so their O3' and O5' atoms lie more nearly in the plane of the page. Apart from these two, which themselves reflect the lateral twofold symmetry of the free dodecamer molecule, the greatest variation occurs in angle δ, which is correlated with deoxyribose sugar conformation.

The extent of overlap of successive base pairs varies considerably from one step to another. In general one finds a greater overlap for purine-pyrimidine steps (GpC, ApT) than for pyrimidine-purine (only CpG in this molecule), as would be expected in an ideal B helix (Arnott *et al*, 1969). The one D-like and two A-like steps, however, are exceptions showing lesser and greater overlap, respectively, than expected from sequence. In Figure 6j, the phosphate group connecting sugars 10 and 11 is turned so its O3' and O5' atoms lie nearly in the plane of the page, as noted earlier. This opens up the twist angle t_ℓ to a D-like 44.7°, so that bases G10 and C11 are hardly stacked atop one another at all. The effect between G22 and C23 in the symmetry-related position at the other end of the molecule (Figure 6b), although similar, is not as extreme.

The A-like steps 3 and 9 show greater than expected overlap, and as we shall see, also provide strong evidence linking the crystal structure of the dodecamer to that in solution. Bases C21 and G22 are stacked over one another more extensively than would be expected for a pyrimidine-purine step in an ideal B helix, as are C9 and G10. As the top view of the helix in Figure 3 reveals, a broader than average helical step from T20 to C21 is followed by a smaller step to G22, and at the symmetry-related position at the other end of the molecule, a broad step from T8 to C9 is followed by a smaller step to G10. This inequality of rotation leads to the two threefold intersections of C1'-C1' vectors that are circled in Figure 3. In the two places where the sequence T-C-G occurs in this dodecamer, C appears to prefer to slip away from T in order to make a better overlap with G. The other ends of these two trios of base pairs show a more normal helical progression, from C3 to G4 and A5, and from C15 to G16 and A17. As mentioned earlier in connection with phosphate separations, steps 3 and 9 should be regarded only as A-like perturbations in a B helix, not as true changes of helix type.

Lomonossoff et al (1981) have observed from kinetic studies in solution that this opened-out TpC step in our CGCGAATTCGCG dodecamer is the preferred point for cutting by the DNAse I endonuclease. The relative rate constant for cutting at step 8 (T8-C9) is 771 ($x10^{-3}$ min^{-1}), whereas the next highest rate constant is 172 for step 10, and all others lie in the range of 6 to 60 ($x10^{-3}$ min^{-1}). This strong preference for attack at TpC may arise precisely because of the

slipping away of C from T in order to stack more efficiently over G, leaving the TpC bond open and more exposed to the enzyme. The logarithms of the cutting rate constants measured by Lomonosoff et al in solution (Table 1) correlate well with the twist angles t_ℓ observed in the crystal, suggesting that the crystal structure is a faithful picture of the solution conformation of the dodecamer, and that the perturbing effects of overlapping ends are small. In Figure 7, both ln k and the end-for-end averaged twist angles t_ℓ are plotted against base step number. This end-for-end averaging of t_ℓ for steps 1 and 11, 2 and 10, etc., is necessary because, no matter what the local crystal environment, the two ends of the molecule certainly will be identical in solution. The left half of the ln k curve in Figure 7 is artificially low because DNAse I cuts shorter pieces of DNA less well (A. Klug, private communication). But maxima and minima of t_ℓ and ln k correlate well over the entire plot, allowing two conclusions to be drawn:

1) Cutting by DNAse I is enhanced by increasing the twist angle of the double helix, presumably making the sugar-phosphate bond more accessible to the enzyme. Since the energy of activation for the cutting process is proportional to -ln k, one can say that the smaller the helical twist angle at a given step, the greater is the activation energy for cutting by DNAse I.

2) The individual helix twist angles measured in the crystal, after

end-for-end averaging, reflect a predominant state of the dodecamer molecule in solution, and not just an accidental freezing-out of one of a large spectrum of vibrational states of the molecule.

The latter point is especially important since the standard deviation in t_ℓ of 3.8° (Table 1) actually is smaller than the 5.9° root mean square thermal fluctuation in twist angle measured by Millar et al (1980) for calf thymus DNA by picosecond spectroscopy. Without the DNAse I cutting evidence, one could not rule out the possibility that the crystal structure was only a sampling or "snapshot" of one of a great number of vibrational states in solution. However, although there may be other conformational variants in solution, that corresponding to the crystal structure clearly is the one that the cutting enzyme sees.

6. Flattening of Propellor Twists

Propellor twist values (Table 2) for A/T base pairs in the center of the molecule have a quite uniform mean and standard deviation of 17.3°(0.4°), whereas those for the G/C base pairs on the ends are both smaller on the average and much more variable, 11.5°(5.1°). One contribution to this effect could be the presence of the third hydrogen bond in G/C base pairs, but an alternative explanation involving base

stacking explains both the smaller average value and the particular distribution of low values seen in Table 2.

In an ideal B helix with uniform propellor twists, each base would be packed against its neighbors up and down the same strand, and this packing could be maintained whether the actual propellor twist angle was large or small. If, however, the base pairs were slid across one another so that a base on one strand overlapped in part with a base from the opposite strand on the next step along the helix, then large propellor twists could produce steric clashes between these two bases, which would have propellor rotations in opposite directions. This is what is happening in the present dodecamer: An overlap in projection between bases on opposite strands leads to a damping-down of propellor twist in at least one of the base pairs involved.

This effect is particularly striking for the base pair having the smallest propellor twist, G10/C15. As Figures 6i and 6j show, G10 overlaps in projection with both G16 above it and G14 below it on the opposite strand. Atom N2 on base G10 is only 3.16 Å from the equivalent atom on base G16, and atom C6 on G10 is only 3.35 Å from atom O6 on G14. These are the dotted vertical distances labeled d and e in Figure 8. Together they constrain base pair G10/C15 to a propellor twist no greater than its actual 4.9°.

In a similar manner, the two short contacts labeled b and c in Figure 8 help to flatten out base pair C3/G22 in the symmetry-related position at the other end of the molecule. Individual atom identifications and distances for these contacts are given in Table 3. Contact

a helps to flatten the propellor twist in both planes C1/G24 and G2/C23. Contact f may be partly responsible for the small 6.2° propellor twist value of G12/C13, but another factor probably is the slightly closer packing of C13 against the sugar-phosphate backbone of its neighboring molecule than is the case for the symmetry-related base C1 at the other end of the helix. Distance g in Figure 8 is only 3.23 Å, whereas distance h is a longer 3.38 Å even in the presence of propellor twist in C1/G24.

In contrast to the foregoing, Figures 6d through 6h show no interstrand overlap between base steps in the central portion of the dodecamer, and the propellor twists here are uniformly 17° to 18°. A/T base pairs should have greater propellor twists in general, both because deletion of one hydrogen bond permits easier twisting, and because the absence of a side group on the C2 of adenine decreases the likelihood of interstrand overlap. If imaginary N2 atoms are added to all the adenines in the dodecamer, then Figure 6 shows that one might expect overlap and damping-down of propellor twist at every step except the central A6-T7 (Figure 6f).

The CGCGAATTCGCG dodecamer even has a built-in control that permits us to conclude that interstrand base overlap is more significant in lowering the propellor twist of G/C base pairs than is the third hydrogen bond. The GpA sequences at G4-A5 and G16-A17 (Figures 6d and 6h) have three hydrogen bonds in the G4/C21 and G16/C9 base pairs, but lack an interfering N2 atom on A5 and A17. The propellor twists at G4/C21 and G16/C9 remain large, 13.2° and 18.6°. Hence base overlap

appears to dominate, with number of hydrogen bonds in a base pair as only a secondary factor. Every propellor twist value observed in the dodecamer can be explained in terms of interstrand base overlap.

7. Sequence Dependence of Helix Conformation

Several aspects of the geometry of the dodecamer helix show a dependence upon base sequence, or at least on purine-pyrimidine base type, that could be important in the recognition process. Individual bases or base pairs have three rotational degrees of freedom corresponding to the nautical roll, pitch and yaw. As shown in Figure 9, twist (t) is rotation about the helix axis (either local or global), tilt (θ_T) is rotation about the pseudo-dyad axis lying in the plane of the bases, and roll (θ_R) is rotation about an axis perpendicular to the pseudo-dyad in the plane of the bases. Global (t_g) and local (t_ℓ) twist have been introduced earlier; this section deals with twist, roll and tilt, and the way in which they are influenced by both base sequence and the 19° crystal-induced bend in the molecule.

As we shall see, the influence of the 19° bend on these local helical parameters is minimal and almost nonexistent. The twofold symmetry that arises from chemical identity of the two strands is much more influential. In the early stages of structure analysis we were concerned about the presence of the bend, as a possibly disruptive influence on the helix. Later we became intrigued by it as a possible source of information about the mechanics of bending of a B helix.

But the refined structure shows that the externally induced bend contributes little by way of either helix disruption or new information. The dodecamer molecule behaves like a smoothly deformable rod without breaks or kinks, and the local twist, roll and tilt parameters are relatively little affected by bending.

The end-for-end averaged twist values plotted in Figure 7 show a regular alternation of maxima and minima with one conspicuous exception at the center of the molecule. Helical rotations at CpG steps (1, 3, 9, 11) are systematically smaller than average, and those at GpC steps (2, 10) are larger. This suggests a form of molecular compensation, by which the pyrimidine-purine steps with their lesser base overlap in an ideally regular B helix are strengthened by decreased twist, at the expense of the intrinsically stronger purine-pyrimidine steps. This is the inverse of a proposal by Klug et al (1979) for an "alternating-B" structure for poly(dA-dT). They suggest that the energy contributed by TpA stacking is so small that little would be lost by eliminating it altogether with an increased twist, permitting enhanced ApT stacking at adjacent positions by virtue of smaller twists there. The dodecamer has only one AT sequence, ApT at step 6, and this does have a small t_ℓ value consistent with the Klug et al proposal. But for alternating GC polymers, Nature seems to have adopted the opposite strategy: not of cutting its losses at pyrimidine-purine sites, but of fine-tuning the twist angle to do the best it can at each base pair. The difference in GC and AT stacking behavior, if further evidence from other structures shows it to be real, may arise because of the third hydrogen

bond in G/C, bringing with it a flatter base pair and better stacking interactions. But clearly, generalizations from one type of pyrimidine-purine base set are not necessarily transferrable to the other.

Table 4 shows how the angles between base plane normals of successive bases change as one progresses along the helix. Angle θ is the magnitude of the total angle between successive base plane normals, and is the $\Delta\kappa$ of Levitt (1978). As defined in Figure 9, the roll, θ_R, is the component of θ along the direction of helix rotation, and the tilt, θ_T, is the component of θ radially outward from the center of the helix. The three sets of values in Table 4 are for best mean planes through both base pairs, and for the bases of strands 1 and 2 considered individually. θ_R is positive if the two planes being compared open out toward the minor groove, and negative if toward the major. For individual strands, θ_T is positive if the base planes open toward the outside of the helix, and negative if toward the center. For the best mean plane comparison, θ_T^{mean} is positive if the angle between planes opens out toward the strand 1 side of the helix. The mean plane values eliminate effects of propellor twist and of any bowing or arching that the base pairs might have, but the individual strand values show the behavior of the bases in more detail.

Most of the roll and tilt angles greater than 4° in Table 4 can be discerned in Figure 1, or even more easily in the overlapping-ends closeup of Figure 8. For example, the alternation of large positive and negative roll values at steps 9-11 of strand 1 can be seen from

Figure 8 to arise because base pairs G10/C15 and G12/C13 are nearly flat (for reasons discussed in the previous section), whereas C9/G16 and C11/G14 have large propellor twists. If all base pairs had the same propellor twists, regardless of what that value might be, then roll angles between successive bases would be uniformly close to zero. Alternation of nearly flat and propellor-twisted base pairs along the helix, however, means alternating positive and negative roll angles. A symmetry-related alternation of θ_R in strand 2 occurs at steps 1-3 at the other end of the molecule. It is less extreme because base pairs C1/G24 and C3/G22 are not as flattened as G10/C15 and G12/C13 are.

A plot of roll angles between best mean planes through successive base pairs, in Figure 10, shows a striking sequence effect: Purine-pyrimidine base steps, whether GpC or ApT, systematically open out toward the major groove of the helix (negative θ_R), whereas pyrimidine-purine steps tend to open the other way toward the minor groove. This is entirely consistent with the observation in Drew et al (1981) that purines prefer higher (less negative) glycosyl χ values than pyrimidines. Careful examination of Figure 6 shows that, if the course of the backbone chain is regarded as being relatively fixed, then tilting the sugars to accommodate the preference of purines for less negative χ values than pyrimidines leads directly to the observed situation: the opening of purine-pyrimidine base pair steps toward the major groove and opening of pyrimidine-purine steps toward the minor. The preference of pyrimidines for lower (more negative) χ values was

explained in Drew et al (1981) as a consequence of close contacts between atom H6 of the base and a C2' hydrogen of the sugar; a contact that is eliminated in purines because of the greater rotational freedom associated with attachment to a five-membered ring. Hence the series of three minima in Figure 10 ultimately can be attributed to differential degrees of steric hindrance around the glycosyl bond.

The downward slope to the right of the entire curve in Figure 10 can be seen from Figure 1 to be a consequence of the bend in the helix, and in fact one of the few detectable consequences. Pushing the top third of the molecule to the left in Figure 1 would have the effect of opening all of the θ_R angles toward the minor groove and making them more positive, whereas pushing the bottom third of the molecule in the same direction opens up the major groove and makes all of the local θ_R values more negative. Were it not for this bend, all three of the purine-pyrimidine minima in Figure 10 presumably would occur around -6°.

Individual roll angles are plotted for strands 1 and 2 in Figure 11. The propellor twists now obscure but do not efface the preference of pur-pyr steps for negative θ_R values, and the positive θ_R at pyr-pur steps are even more pronounced. But Figure 11 shows another structural correlation: Homopolymer steps of the type GpA, ApA, TpT and TpC tend to be resistant to roll in either direction. This observation may help to explain the finding that nucleosome cores can be reconstituted from polynucleotides and histone cores if the DNA is of alternating or mixed sequence, but not with the homopolymers

poly(dA)·poly(dT) or poly(dG)·poly(dC) (Simpson and Künzler, 1979; Rhodes, 1979). The authors account for this behavior by invoking the tendencies of homopolymers to adopt triple-strand or A-helix conformations. But it may also be that the resistance to roll seen here in stacked homopolymers makes them stiffer and less easily wrapped around a nucleosome core with radius of only 44 Å.

Strand 2 has been reversed in Figure 11 because this leads to a better match with strand 1. The low -12° value at the end of strand 2 (step 1 in Figure 11) arises because base pair G12/C13 is flattened out more than C1/G24, and the particular packing of helix ends in the crystal has already been discussed as an explanation. Hence, except for this one end effect, the bringing of chemically equivalent bases in the two strands together by reversing strand 2 in the figure emphasizes that base sequence is more important in determining the roll angles than is any outside influence. The correlation of behavior across the lateral twofold axis relating the two ends of the molecule is stronger than the correlation around the helix axis itself. This would be an obvious statement if the helix were unbent and isolated in solution, but its applicability to the crystal as well illustrates the persistence of sequence-dependent structural features.

In each of the two strands, the roll angle oscillates in the two CGCG regions, but the oscillations are more exaggerated in the second such region than in the first. Why should this be, and what is responsible for the great swing in θ_R between steps 9 and 10 in each strand? Figure 12 demonstrates that this is a consequence of the

reinforcement or the opposition of two different effects: the inherent tendency of pyr-pur sequences to open one way and pur-pyr steps the other, and the alternation of large and small propellor twists within the CGCG regions. If the first and third base pairs are flat while the other two are propellor twisted, then the two effects tend to cancel, and oscillations in θ_R are decreased, as is the case in the first half of each chain (C1-G4 and C13-G16 in Figure 12). If, in contrast, it is the second and fourth base pairs that are flattened, then the effects are mutually reinforcing and roll angle oscillations are exaggeraged, as in the second halves of the two chains (C9-G12 and C21-G24).

The base tilt, θ_T, shows a lesser degree of sequence dependence (Figure 13). Again one strand has been reversed to illustrate the sequence dependence of the curves. Both strands show a high maximum at their third step, followed by a minimum, and then intermediate values leading to a final minimum and rise. The curves show less end-for-end symmetry than the roll curves of Figure 11; without the reversal of one strand they would not correlate at all well. The most striking feature--the maximum at step 3--indicates a splaying-out of the base pairs that are opposite to the tightly overlapped C and G of the two T-C-G sequences that were first noted in section 5 as the most favored DNAse I cutting sites. It does not seem unreasonable, examining Figure 8, to propose that sliding base C21 up into close contact with G22 would tip the other ends of the two base pairs apart, or that the same effect should occur at the other end of the molecule with C9

and G10, but this is only speculation based on visual inspection of the model.

8. Homopolymer Stacking and EcoRI Cutting Specificity

The comparative resistance of homopolymer (pur-pur and pyr-pyr) base stacks to rolling in the direction of either helix groove may be significant in the specificity of the EcoRI restriction endonuclease cutting site. The "canonical" sequence at which cutting normally takes place is GAATTC, which was incorporated into the center of this dodecamer. But increasing pH and decreasing ionic strength, or alternatively adding Mn^{2+} ions, can lead to less selective cutting at other similar but "noncanonical" sites (Goodman et al, 1977; Woodbury et al, 1980a,b). These are listed in Table 5, ranked in order of decreasing susceptibility to cutting by the EcoRI enzyme. The standard GAATTC site is best, but GGATTT and AAATTT are next, followed by GAATTT and GAATTA. AGATTT is also reported as being cut. In contrast, all those sequences preceded by an x) in Table 5 are not cut by the restriction enzyme even under the more permissive conditions. Goodman et al (1977) have proposed that the enzyme prefers the sequence GAATTx, with x = C for normal conditions and x = G, T or A for the more relaxed conditions. However, Woodbury et al (1980b) find that GAATTG is recalcitrant, leaving only C, T or A at the final position.

With the relative stiffness of homopolymer stacking in our dodecamer in mind, we were struck by the observation that all of the

successful cutting sites in Table 5 with the exception of GAATTA consist of two homopolymer triplets: pur-pur-pur-pyr-pyr-pyr. Also with this one exception, every potential site having only one homopolymer triplet, or none at all, fails as a cutting point for the EcoRI enzyme or as a substrate for the related methylase. Perhaps the structural integrity of stacked homopolymers, reflected in the low θ_R values in Table 4, is important in recognition and binding by these two enzymes. The data in Table 5 suggest that a generalized EcoRI site should be: pur-pur-A-T-T-pyr, with A being favored over G at the second position. Apparently A can also be tolerated, reluctantly, at the pyr site providing that it is also present as the most favored alternative in the second position.

9. Sequence-Structure Generalizations from CGCGAATTCGCG

This x-ray structure analysis of an EcoRI restriction endonuclease site, GAATTC, embedded in an alternating CG copolymer, has produced several generalizations about sequence-structure correlations in B-DNA that can be tested against future structure analyses of other sequences:

1) The tendency of an alternating CG copolymer to adopt the left-handed Z helical conformation is small, and can be upset by intervening sequences that are incompatible with the Z structure. (The dodecamer crystals are fully as "high salt" as were crystals

of CGCG or CGCGCG (Wing et al, 1980). Moreover, Dinshaw Patel (private communication) has been unable to observe a salt-induced transition in CGCGAATTCGCG in solution by NMR methods.) Hence the left-handed Z helix probably can be dismissed from consideration within regions of DNA that ultimately are expressed as amino acid chains in proteins.

2) When a bending moment is applied to B-DNA, the helix deforms smoothly and continuously, without kinking, in such a manner as to have little effect on the local helix parameters such as twist, roll or tilt. The smooth-bending calculations of Levitt (1978) accordingly are probably correct, and the energy required to produce a 19° bend in a straight, 11-base-step helix is around 0.5 kcal/mole as calculated by Wing et al (1980).

3) Purine-pyrimidine steps show a tendency to open their base planes toward the major groove of the helix, pyrimidine-purine steps open preferentially toward the minor groove, and homopolymer steps are relatively resistant to rolling in either direction. The pur-pyr and pyr-pur effects ultimately can be traced back to a preference of pyrimidines for very negative glycosyl torsion angles χ, itself a consequence of steric hindrance about the C1'-N bond to the six-membered ring.

4) CpG steps have smaller helix twist angles than do GpC steps, as if a kind of internal compensation were at work, making up for the inherently smaller degree of overlap in pyrimidine-purine base steps in a perfectly regular B helix. The situation with A and T is unclear; the sole representative in this dodecamer shows exactly the reverse effect with a small twist at ApT.

5) G/C base pairs have an inherently smaller propellor twist than A/T pairs. Interstrand overlap of bases in successive steps of the helix is more important in producing this effect than is the presence of the third hydrogen bond in G/C.

6) Considerable sequence-induced local variation in helix parameters t_{ℓ}, θ_R and θ_T can be observed within the overall framework of a B helix, even to the extent of finding steps with distinctly A-like or D-like character. The EcoRI restriction site itself has an average of 9.8 base pairs per turn, but is flanked by two steps having both the base tilt (relative to the local helix rotation vector) and helical twist expected from an A helix. This conceivably could represent one component in base sequence recognition by the EcoRI endonuclease.

7) Within the base sequence: T-C-G, the C slips away from T and stacks more tightly over G. At the other ends of the latter two

base pairs, the planes of bases G and C are opened up radially toward the outside of the helix (θ_T positive).

8) The enzyme DNAse I shows a strong preference for cutting at sites having large helical twists, presumably because the helix backbone at these positions is more open and exposed to the enzyme. In the present dodecamer, cutting is much more rapid at the T-C-G site just mentioned, by a factor of 4.5. This correlation of twist and cutting rate is evidence for the essential identity of dodecamer helix structure in the crystal and in solution.

9) The EcoRI endonuclease and methylase show a preference for the sequence: pur-pur-A-T-T-pyr, which contains two adjacent homopolymer triplets. A can be tolerated at the pyr position only if it, rather than G, is also present at the second position.

After this first B-DNA single-crystal structure analysis, we are in roughly the same position that protein crystallographers were in 1959 after the first globular protein structure analysis (Kendrew et al, 1960). We know a lot about one molecule, but are unsure as to how much of this can safely be generalized to similar molecules with different sequences. All of the above principles must be tested against other B-DNA molecules, and the CCGG and CCCAAATTTGGG analyses now under way in this laboratory will furnish a start. If the history of protein crystallography is any guide, the present simplicity in DNA

structure that comes mainly from ignorance will shortly be replaced by a bewildering complexity of new data, before it ultimately settles down again into the simplicity that means that we truly understand matters. The CGCGAATTCGCG structure is a beginning.

Acknowledgments

We should like to thank Mary Kopka for her willingness to read and re-read endless drafts, criticizing our presentation and contributing much to whatever clarity the paper might now have. We are grateful to Dinshaw Patel, P. J. G. Butler, G. P. Lomonossoff and Aaron Klug for making their experimental results with this dodecamer available to us prior to publication. We should also like to thank Lillian Casler for her patience and skill in preparing the line drawings as illustrations. This work was carried out with the support of National Institutes of Health grants GM-12121 and GM-24393, and National Science Foundation grant PCM79-13959. H.D. was also the recipient of a National Institutes of Health Predoctoral Traineeship. This is contribution No. 6350 from the Division of Chemistry and Chemical Engineering.

REFERENCES

Arnott, S. (1976). In "Organization and Expression of Chromosomes" (Allfrey, V. G., Bautz, E. K. F., McCarthy, B. J., Schimke, R. T. and Tissieres, A., Eds.), Dahlem Konferenzen, Berlin, pp. 209-222.

Arnott, S. and Hukins, D. W. L. (1972). Biochem. Biophys. Res. Commun. 47, 1504-1509.

Arnott, S., Dover, S. D. and Wonacott, A. J. (1969). Acta Crystallogr. B25, 2192-2206.

Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C. and Watts, L. (1974). J. Mol. Biol. 88, 524-533.

Arnott, S., Chandrasekaran, R. and Selsing, E. (1975). In "Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions" (Sundaralingam, M. and Rao, S. T., Eds.), University Park Press, Baltimore, pp. 577-596.

Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. and Ratliff, R. L. (1980). Nature 283, 743-745.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1981). Proc. Natl. Acad. Sci. USA--in press.

Goodman, H. M., Greene, P. J., Garfin, D. E. and Boyer, H. W. (1977). In "Nucleic Acid-Protein Recognition" (Vogel, H. J., Ed.), Academic Press, New York, pp. 239-259.

Gupta, G., Bansal, M. and Sasisekharan, V. (1980). Proc. Natl. Acad. Sci. USA 77, 6486-6490.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. B., Davies, D. R., Phillips, D. C. and Shore, V. E. (1960). Nature 185, 422-427.

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

Levitt, M. (1978). Proc. Natl. Acad. Sci. USA 75, 640-644.

Lomonossoff, G. P., Butler, P. J. G. and Klug, A. (1981). J. Mol. Biol.--submitted.

Millar, D. P., Robbins, R. J. and Zewail, A. H. (1980). Proc. Natl. Acad. Sci. USA 77, 5593-5597.

Rhodes, D. (1979). Nucleic Acids. Res. 6, 1805-1816.

Sarma, R. H., Dhingra, M. M. and Feldmann, R. J. (1979). In "Stereodynamics of Molecular Systems" (Sarma, R. H., Ed.), Pergamon Press, New York, Plates 1-16.

Simpson, R. T. and Künzler, P. (1979). Nucleic Acids. Res. 6, 1387-1415.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980). Nature 287, 755-758.

Woodbury, C. P. Jr., Downey, R. L. and von Hippel, P. H. (1980a). J. Biol. Chem. 255, 11526-11533.

Woodbury, C. P. Jr., Hagenbüchle, O. and von Hippel, P. H. (1980b). J. Biol. Chem. 255, 11534-11546.

TABLE 1. Helical Twist, Rise, and Rate of DNAse I Cutting at Successive Base Steps

Base Step	Type	t_g	t_ℓ	Δt	Rise Per Base Pair, $\underline{h}^\dagger$	$\ell n\ k^\ddagger$
1	CG	39.4°	38.3°	+1.1°	3.36 Å	--
2	GC	37.4	39.6	-2.2	3.38	3.61
3	CG	27.4	33.5	-6.1	3.26	1.79
4	GA(=TC)	36.8	37.4	-0.6	3.30	3.54
5	AA(=TT)	37.7	37.5	+0.2	3.27	1.72
6	AT	32.3	32.2	+0.1	3.31	2.05
7	TT(=AA)	35.6	36.0	-0.4	3.29	3.82
8	TC(=GA)	40.3	41.4	-1.1	3.14	6.65
9	CG	30.3	32.3	-2.0	3.56	3.96
10	GC	39.8	44.7	-4.9	3.21	5.14
11	CG	37.2	37.0	+0.2	3.54	2.23
All Steps: Mean:		35.8	37.3	-1.4	3.33	
Std. dev.:		(4.1)	(3.8)	(2.2)	(0.13)	
Steps 4-8 only: Mean:		36.5	36.9	-0.4	3.26	
Std. dev.:		(2.9)	(3.3)	(0.5)	(0.07)	

†Rise along local helix rotation vector, not overall best helix axis.

‡Logarithm of relative rate constants for DNAse I bond cleavage, from Lomonosoff et al (1981).

TABLE 2. Torsion and Propellor Twist Angles, and Sugar Conformation

Residue	Glycosyl: χ	Main Chain Torsion Angles: α	β	γ	δ	ε	ζ	Adjacent Phosphorus Atom Separation (Å)	Observed Sugar Conformation	Base Pair Propellor Twist: ψ
C1	-105°	--	--	174°	157°	-141°	-144°	--	C2'-endo	13.2°
G2	-111	-66°	170°	40	128	-186	- 98	6.64	C1'-exo	11.7
C3	-135	-63	172	59	98	-177	- 88	6.47	O1'-endo	7.2
G4	- 93	-63	180	57	156	-155	-153	6.83	C2'-endo	13.2
A5	-126	-43	143	52	120	-180	- 92	6.88	C1'-exo	17.1
A6	-122	-73	180	66	121	-186	- 89	6.90	C1'-exo	17.8
T7	-127	-57	181	52	99	-186	- 86	6.29	O1'-endo	17.1
T8	-126	-59	173	64	109	-189	- 89	6.87	C1'-exo	17.1
C9	-120	-58	180	60	129	-157	- 94	6.70	C1'-exo	18.6
G10	- 90	-67	169	47	143	-103	-210	6.55	C2'-endo	4.9
C11	-125	-74	139	56	136	-162	- 90	7.05	C2'-endo	17.2
G12	-112	-82	176	57	111	--	--	--	C1'-exo	6.2
C13	-128	--	--	56	137	-159	-125	--	C2'-endo	6.2
G14	-116	-51	164	49	122	-182	- 93	6.62	C1'-exo	17.2
C15	-134	-63	169	60	86	-185	- 86	6.45	O1'-endo	4.9
G16	-115	-69	171	73	136	-186	- 98	7.12	C2'-endo	18.6
A17	-106	-57	190	54	147	-183	- 97	6.77	C2'-endo	17.1
A18	-108	-57	186	48	130	-186	-101	6.71	C2'-endo	17.1
T19	-131	-58	174	60	109	-181	- 88	6.70	C1'-exo	17.8
T20	-120	-59	179	55	122	-181	- 94	6.70	C1'-exo	17.1
C21	-114	-59	185	45	110	-177	- 86	6.17	C1'-exo	13.2
G22	- 88	-67	179	50	150	-100	-188	6.60	C2'-endo	7.2
C23	-125	-72	139	45	113	-174	- 97	6.68	C1'-exo	11.7
G24	-135	-65	171	47	79	--	--	--	C3'-endo	13.2
Avg.	-117	-63	171	54[1]	123	-169	-108	6.68		
Std.Dev.	(14)	(8)	(14)	(8)	(21)	(25)	(34)	(0.23)		
B DNA[2]	-119	-61	180	57	122	-187	- 91	--		
B_F DNA[3]	-102	-41	136	38	139	-133	-157	6.46		
A_F DNA[4]	-154	-90	-149	47	83	-175	- 45	5.64		

Notes to Table 2:

Main chain conformation angles are defined as:

P—α—O5'—β—C5'—γ—C4'—δ—C3'—ε—O3'—ζ—P

with zero at the fully eclipsed or cis position and positive angles by clockwise rotation of the farther pair of atoms. The glycosyl angle χ is similarly defined by atoms O1'——C1'—χ—N1——C2 for pyrimidines (C,T) or O1'——C1'—χ—N9——C4 for purines (G,A).

[1]Omitting C1 value as representing an end effect.

[2]Model chosen from Levitt (1978) energy refinement for best agreement with our results: 10 base pairs per turn and C——O——C deoxyribose angle set at 115°.

[3]Arnott et al (1980) fiber data with 10 base pairs per turn.

[4]Arnott et al (1980) fiber data with 11 base pairs per turn.

TABLE 3. Close Interstrand Contacts Between Overlapping Bases

Base Step	From: Base	From: Atom	To: Base	To: Atom	Distance	Label in Figure 8
1	G2	N3	G24	N2	3.50 Å	a
2	G2	O6	G22	O6	3.21	b
3	G4	N2	G22	N2	3.31	c
9	G10	N2	G16	N2	3.16	d
10	G10	C6	G14	O6	3.35	e
11	G12	N2	G14	C2	3.40	f
end	C13	O2	$G2^-$	C4'	3.23	g
end	C1	O2	$G14^+$	C4'	3.38	h

– and + refer to the next molecules down or up the 2_1 axis.

TABLE 4. Angles Between Base Plane Normals

Base Step	Best mean plane: θ^{mean}	θ_R^{mean}	θ_T^{mean}	Strand 1: θ	θ_R	θ_T	Strand 2: θ	θ_R	θ_T
1	7.3	+7.1	-1.6	7.6	+ 7.5	- 1.3	7.2	+ 6.2	+ 3.7
2	3.2	-3.2	+0.1	3.6	- 2.9	+ 2.2	6.1	- 6.0	+ 1.1
3	5.1	+5.0	+0.8	13.2	+ 4.3	+12.4	11.8	+10.9	+ 4.6
4	5.9	+4.5	+3.8	2.7	+ 2.7	0.0	6.9	+ 3.9	+ 5.7
5	1.3	+1.2	-0.5	4.1	+ 1.7	+ 3.7	4.8	+ 0.5	+ 4.8
6	6.3	-6.1	+1.7	6.5	- 1.7	+ 6.3	4.3	- 3.6	+ 2.3
7	2.6	+0.6	+2.5	7.2	+ 1.1	+ 7.1	1.3	+ 0.9	+ 0.9
8	2.7	+2.7	+0.4	1.4	- 0.9	+ 1.1	2.1	+ 1.4	+ 1.6
9	1.8	+1.7	-0.4	9.6	+ 7.9	+ 5.5	10.9	0.0	+10.9
10	10.1	-9.7	-3.0	18.2	-18.0	- 2.8	8.4	- 5.7	+ 6.1
11	1.3	-1.0	+0.8	11.9	+ 7.0	+ 9.6	12.5	-12.0	+ 3.4
Mean	4.33	0.25	0.96	7.82	0.79	3.98	6.94	- 0.32	4.10
Std. Dev.	2.84	4.99	1.63	5.08	7.26	4.70	3.74	6.32	2.88

θ_R is positive when opening toward minor groove.

θ_T is positive when opening to outside of molecule (toward strand 1 for best mean plane values).

TABLE 5. EcoRI Cutting Specificity and Homopolymer Triplet Character

Number of homopolymer triplets:

2	1	0
1) GAA TTC	5) GAA TT.A	x) C.AA TT.G
2) GGA TTT	x) GAA TT.G	x) T.AA TT.G
3) AAA TTT	x) GGA TT.A	
4) GAA TTT	x) GGA TT.G	
-) AGA TTT	x) AAA TT.G	

Leading numbers indicate order of descending reactivity for EcoRI cleavage, with site marked -) also cut although not ranked. Sites marked x) are not cut. Cutting information from Woodbury et al (1980a,b). Homopolymer triplets (pur-pur-pur or pyr-pyr-pyr) are emphasized by heavy overlining, and changes of pur/pyr type within triplets are indicated by dots. Only one strand is shown for non-selfcomplementary sequences.

FIGURE CAPTIONS

FIGURE 1. The structure of the dodecamer CGCGAATTCGCG. The 5' end of strand 1 (bases 1-12) and the 3' end of strand 2 (bases 24-13) appear at upper left and right, respectively, and the 3' end of strand 1 and the 5' end of strand 2 appear at lower left and right. Base pair C1/G24 is at the top of the helix, and C13/G12 is at the bottom. Individual base steps n (from base pair n to n + 1) are numbered. The radius of each atom is proportional to the square root of its vibrational temperature factor, B, obtained from crystallographic refinement to a residual error of R = 17.8% (Drew et al, 1981). Hence the size of the atom as drawn conveys information about the mean displacement of the atom from its equilibrium position in the helix. Displacements are greater along the sugar-phosphate backbone than in base pairs, and are less for base pairs in the center of the molecule than for those at the end. The overall helix axis exhibits a 19° bend to the left as indicated by the curved arrow.

FIGURE 2. Skeletal view of dodecamer packing along the vertical 2_1 screw symmetry axis along c. Only phosphate groups (circles) and sugar C1' atoms (bends) are shown, with base pairs represented by thin connecting lines. The first and last two base pairs of one molecule (open circles) are nested in the minor grooves of the helices above and below (crossed spheres).

FIGURE 3. View down the best overall helix axis, showing only C1' deoxyribose ring atoms as numbered circles. Thin lines connect the C1' atoms of paired bases. Two symmetry-related threefold intersections of C1'-C1' vectors are circled.

FIGURE 4. Plot of global (t_g) and local (t_ℓ) twist or helix rotation as a function of base pair step number. Notice minima at steps 3, 6 and 9, and the maximum at 10.

FIGURE 5. Skeletal drawings of the dodecamer molecule as in Figure 2, with individual local helix rotation vectors relating two successive base pairs superimposed. Each vector shows the position and orientation of the rotation axis, and its length indicates the translation distance, or rise per base pair in Table 1. Base-pair steps are designated 1 to 11 from the top of the molecule as in Figure 1. Angles of local rotation, t_ℓ, are listed in Table 1.
(a) View down the approximate lateral twofold axis as in Figure 1.
(b) View 90° to the left, with the approximate twofold axis horizontal in the plane of the page.

FIGURE 6. Stereo drawings of the 11 individual base-pair steps, viewed along the best overall helix axis of Figure 3. Bases in strand 1 are numbered along the right side, and those in strand 2 are numbered at the left.

FIGURE 7. Correlation between local twist angle (t_ℓ) at a particular base step, and the relative rate constant for cutting at that point by DNAse I. Rate data from Table 1 and Lomonosoff et al (1981).

FIGURE 8. Minor groove overlap of the ends of two molecules up the 2_1 axis parallel to c (vertical). The intersection shown is the lower one of Figure 2 viewed from the left. Selected base numbers are given within sugar rings, and base step numbers are circled. The light molecule continues above the figure, and the dark molecule continues below. Unlike Figure 1, atomic radii here indicate atom type: P, O, N and C in order of descending size. Bases G12 on the light molecule and G2 on the dark are connected by pairs of N2-N3 hydrogen bonds shown as thin lines, as are G14 on the light molecule and G24 on the dark. The 3'-OH of strand 1 of the light molecule is hydrogen bonded to the N2 amino group of G22 on the dark, and the 3'-OH of strand 2 of the dark molecule is similarly bonded to the ring N3 of G16 on the light (and not to N2 of base G10 as strict twofold equivalence would suggest). The six vertical dotted lines labeled a-f indicate close contacts from inter-strand base overlap that flatten out the propellor twists of base pairs G10/C15 and C3/G22, and to a lesser extent, G12/C13 and C1/G24. Dotted lines g and h mark close contacts between two molecules. These eight contact distances are listed in Table 3.

FIGURE 9. The three rotational degrees of freedom of rigid base planes. Twist, t, is rotation about a helix axis, either local or global. Tilt, θ_T, is rotation about the pseudo-dyad axis passing through the base plane. Roll, θ_R, is rotation about an axis in the plane of the bases perpendicular to the pseudo-dyad. The two C'-N arrows indicate the vectors used to define local helix axes and parameters t_ℓ and h. Adapted from Figure 2 of Arnott et al (1969).

FIGURE 10. Variation in roll angle for successive base pairs along the helix. θ_R^{mean} measures the roll angle between two successive base pairs, measured from the best mean plane through all atoms of the entire base pair. It is positive if the angle between base pair planes opens into the minor groove, and negative if toward the major groove. Data from Table 4. Note the strong tendency of purine-pyrimidine steps at 2, 6 and 10 to open to the major groove, and the overall downhill slope of the curve to the right.

FIGURE 11. Roll angles between successive bases, for each strand considered separately. Strand 2 has been reversed to bring the same base sequence for both strands to a given position in the plot, and to emphasize the twofold symmetry in helix conformation. Small arrows indicate the 5'...3' directions of strands 1 and 2. H at bottom indicate homopolymer (pur-pur and pyr-pyr) base steps.

FIGURE 12. Combined effect on roll angles of alternation of purines and pyrimidines, and alternation of flat and propellor-twisted base pairs. The view is into the minor groove. Large + and - signs indicate the tendency for pyr-pur and pur-pyr sequences to open toward the minor and major grooves, respectively. Small signs indicate the perturbing effects of propellor twists only in alternate base pairs. Where these two effects reinforce, at the left, the + and - oscillation in roll angles, θ_R, is unusually large. Where the two effects act in opposition, at the right, oscillations in θ_R are damped down. This is why strands 1 and 2 both have greater oscillations in θ_R in their second halves, even though these occur at opposite ends of the dodecamer molecule. Actual observed roll angles for the two strands are shown at each step.

FIGURE 13. Tilt angles between successive base planes, for each strand considered separately. θ_T is positive if the angle between base planes opens toward the outside of the helix, and negative if it opens toward the center. Strand 2 has been reversed to bring the same base sequence for both strands to a given position in the plot, and to emphasize the twofold symmetry in helix conformation. Small arrows indicate the 5'...3' directions of strands.

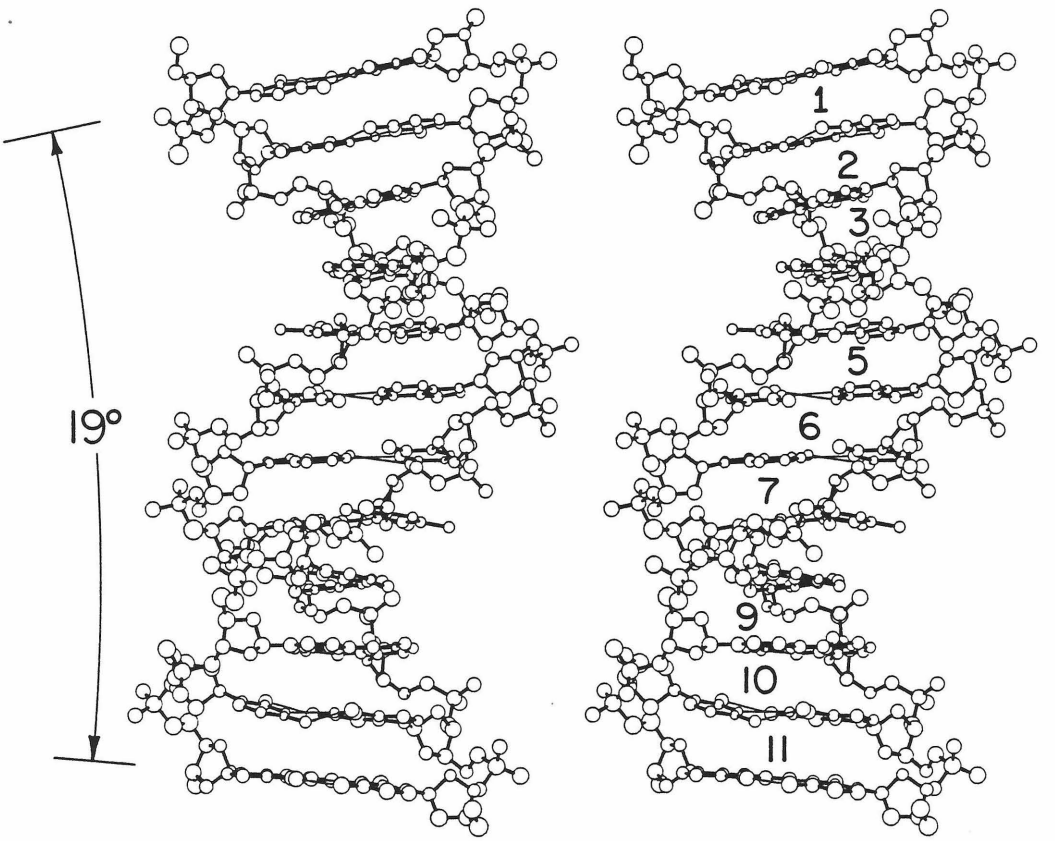


Figure 1

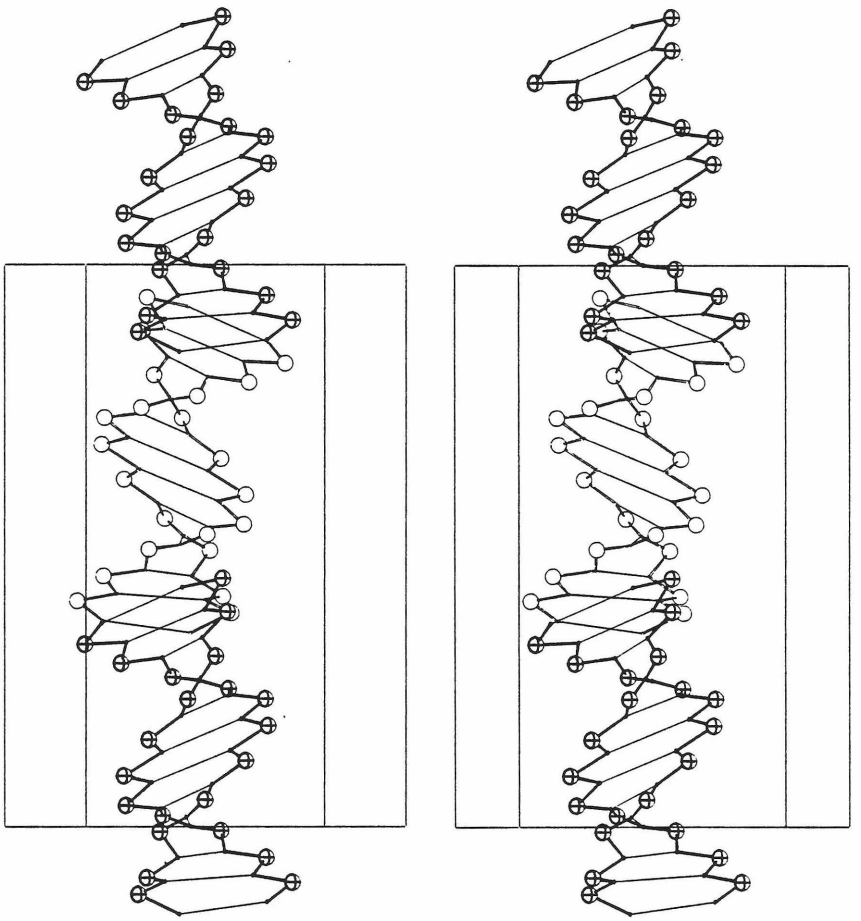


Figure 2

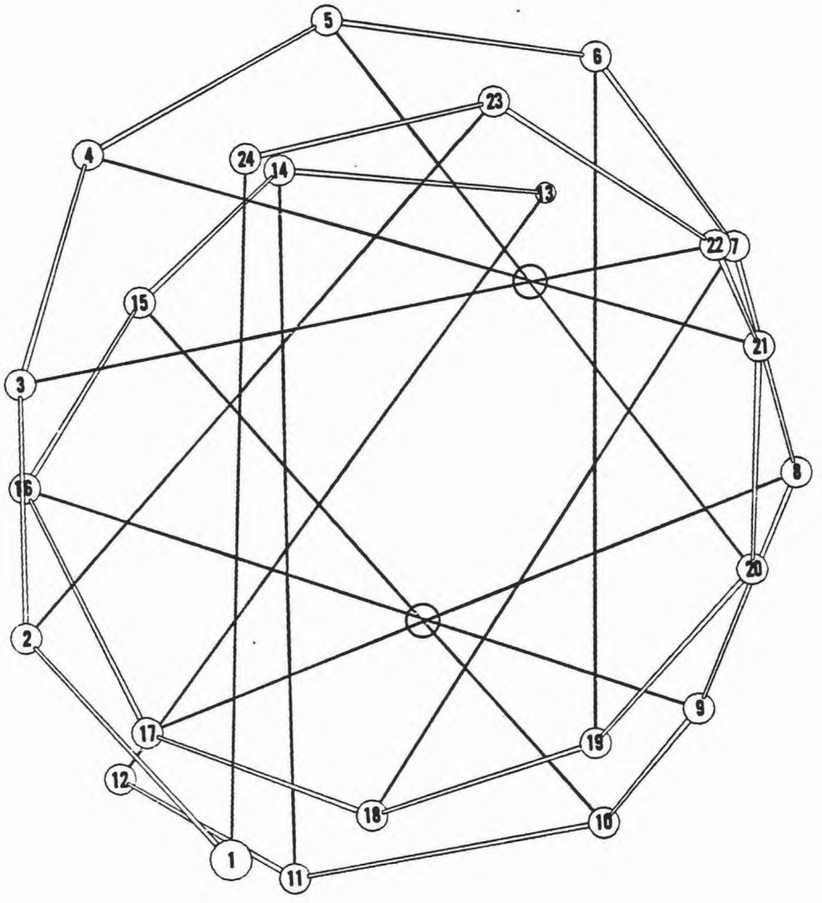


Figure 3

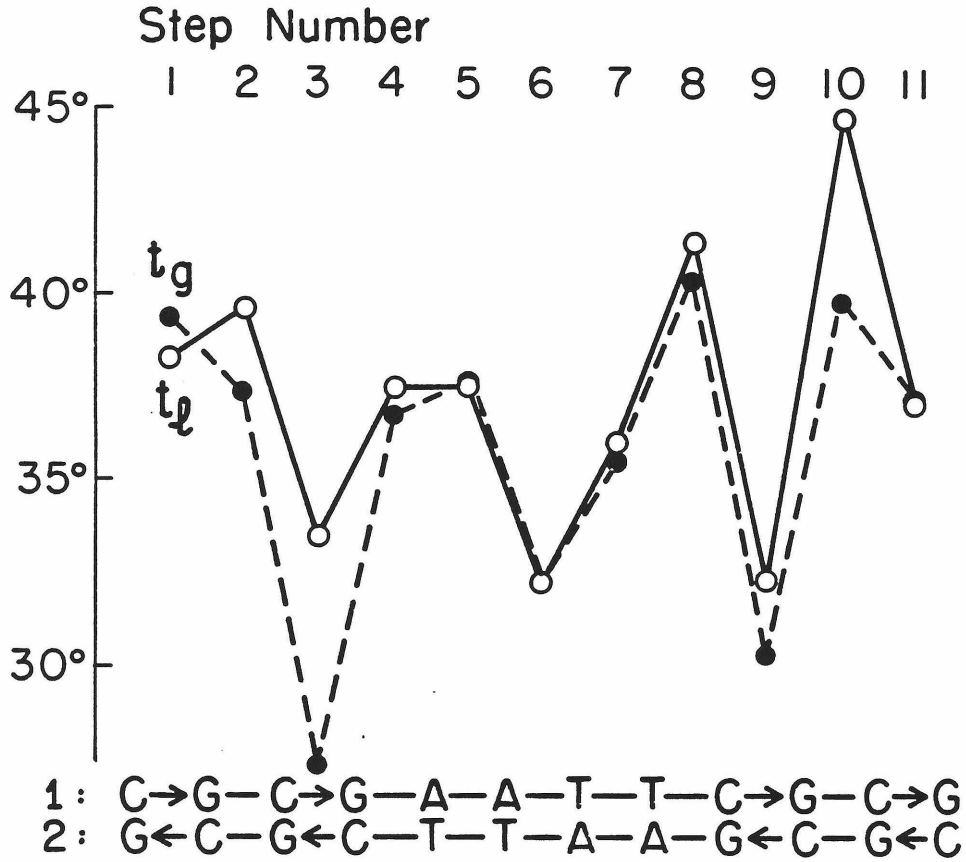

Step Number
1 2 3 4 5 6 7 8 9 10 11
45°
40°
35°
30°
t_g
t_ℓ
1: C→G—C→G—A—A—T—T—C→G—C→G
2: G←C—G←C—T—T—A—A—G←C—G←C

Figure 4

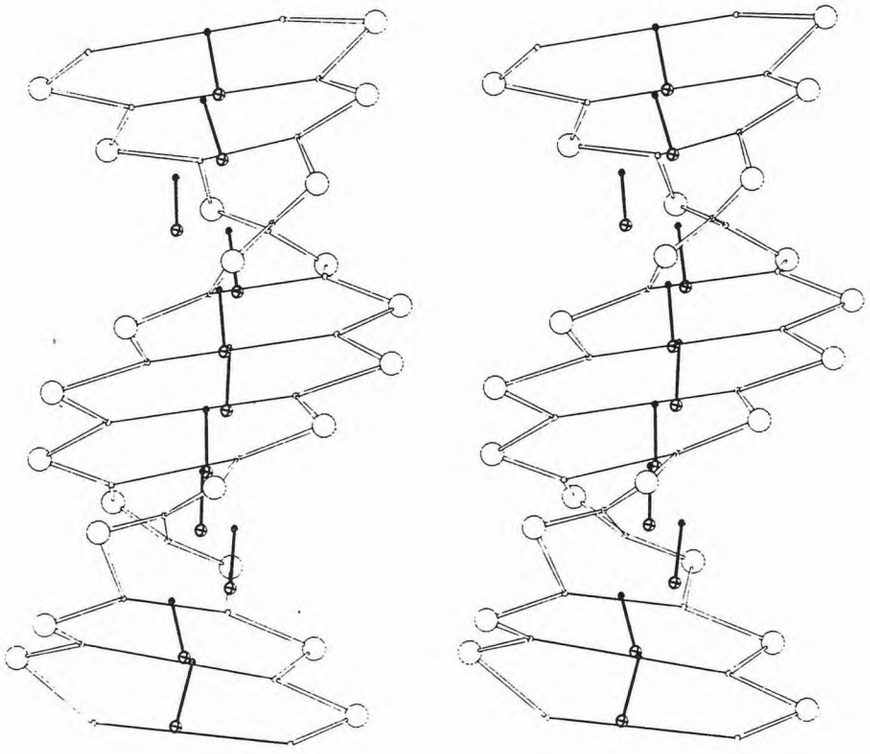


Figure 5a

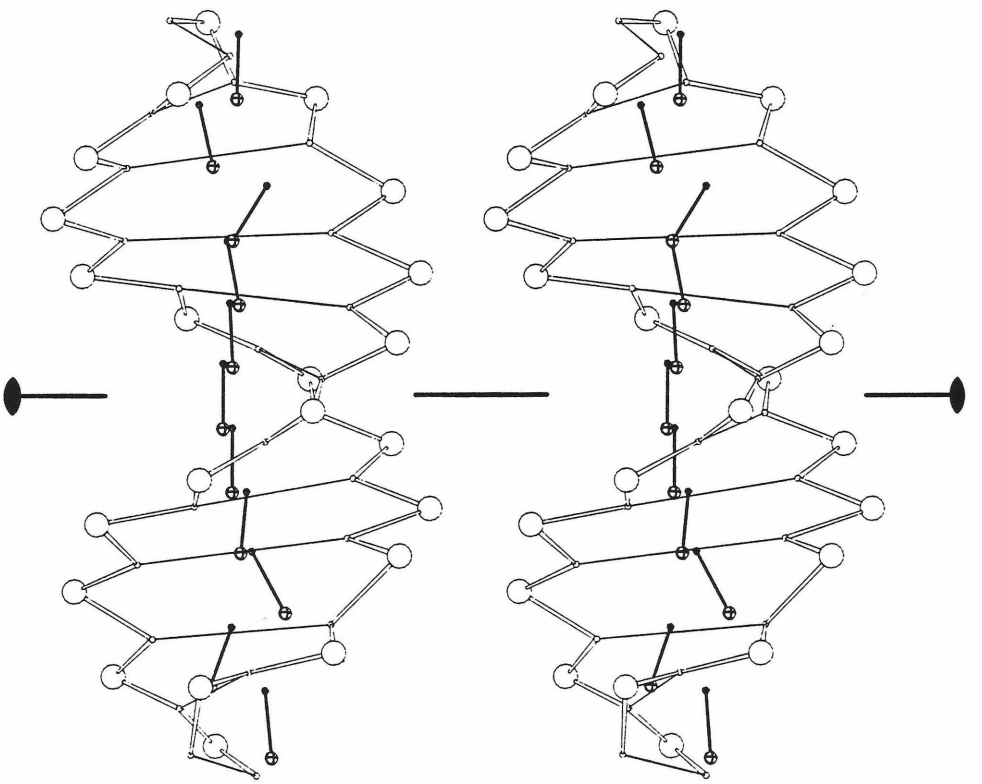


Figure 5b

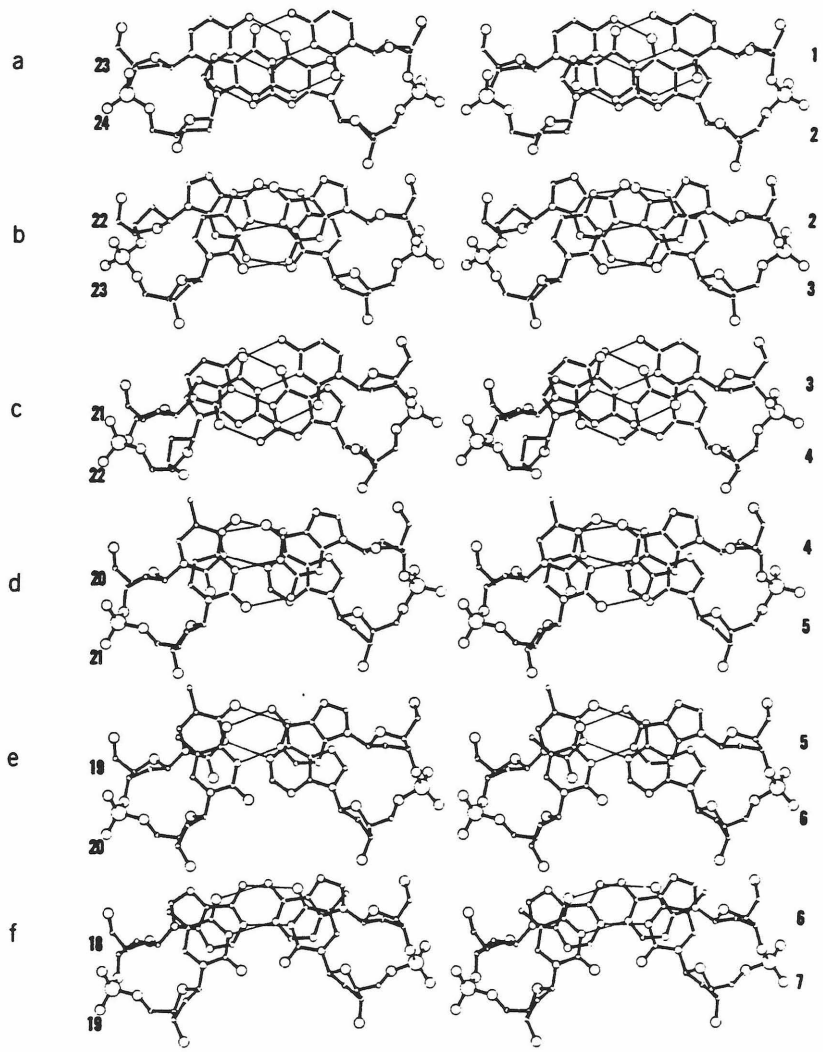


Figure 6a-f

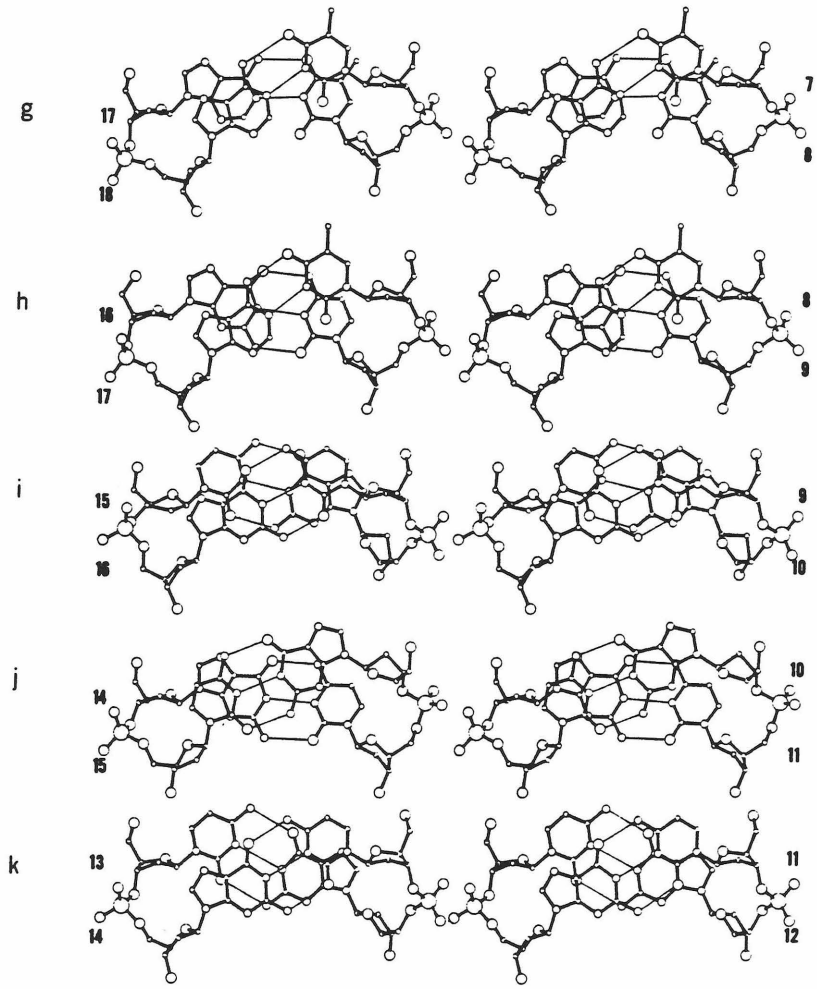


Figure 6g-k

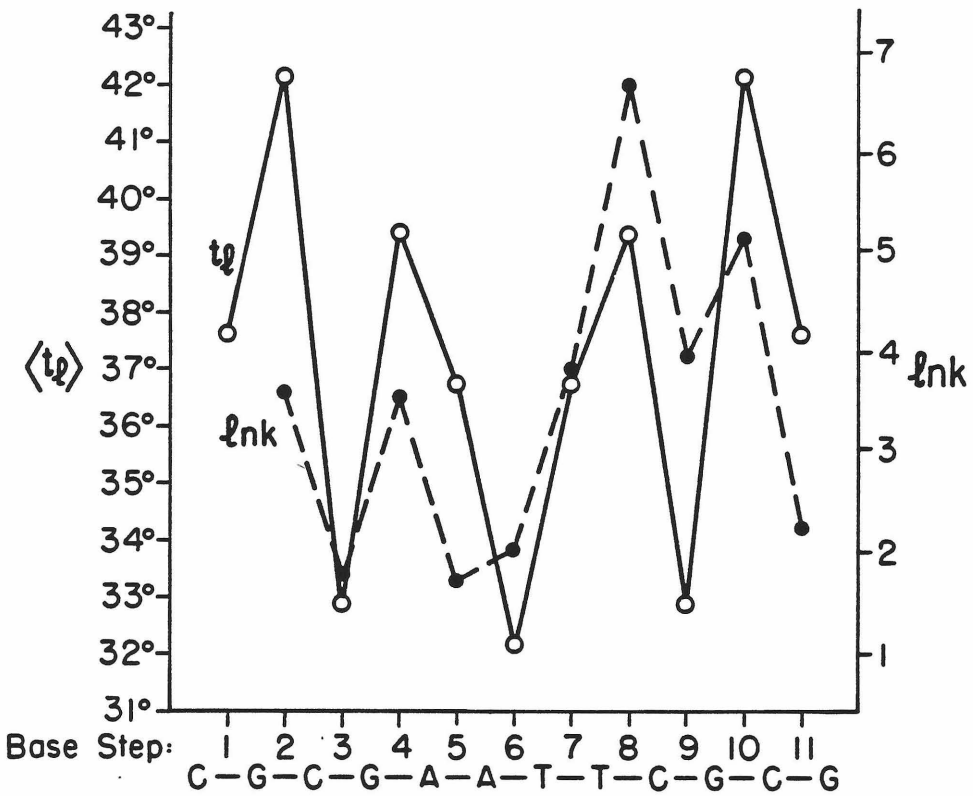

43°
42°
41°
40°
39°
38°
37°
36°
35°
34°
33°
32°
31°
⟨tℓ⟩
tℓ
ℓnk
7
6
5
4
3
2
1
ℓnk
Base Step: 1 2 3 4 5 6 7 8 9 10 11
C—G—C—G—A—A—T—T—C—G—C—G

Figure 7

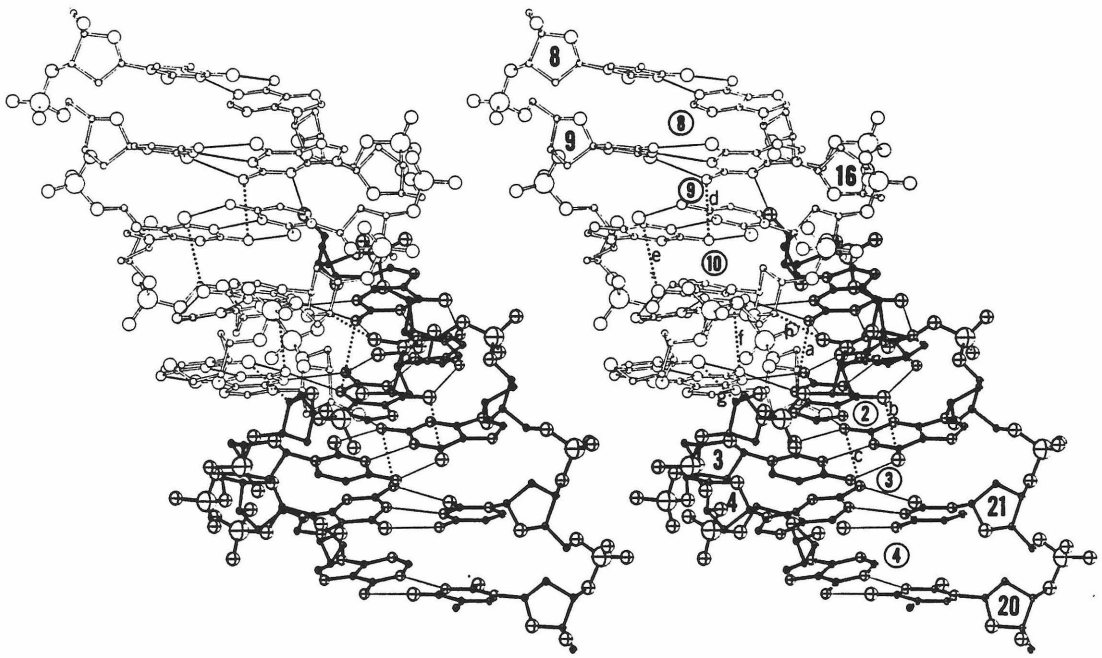


Figure 8

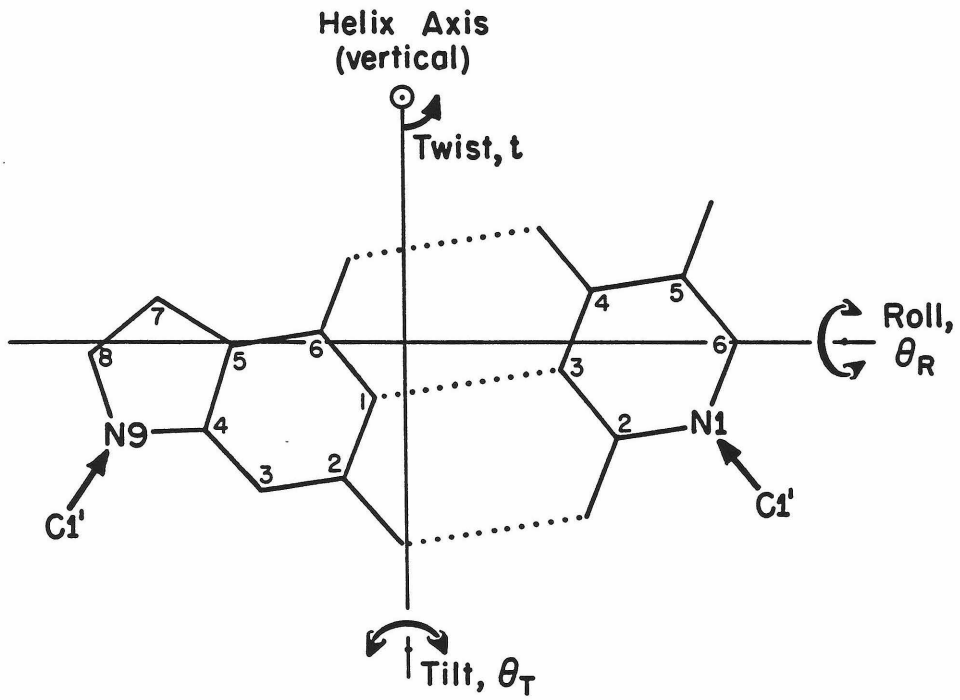

Helix Axis
(vertical)
Twist, t
Roll,
θ_R
Tilt, θ_T
N9
N1
C1'
C1'
1
2
3
4
5
6
7
8
1
2
3
4
5
6

Figure 9

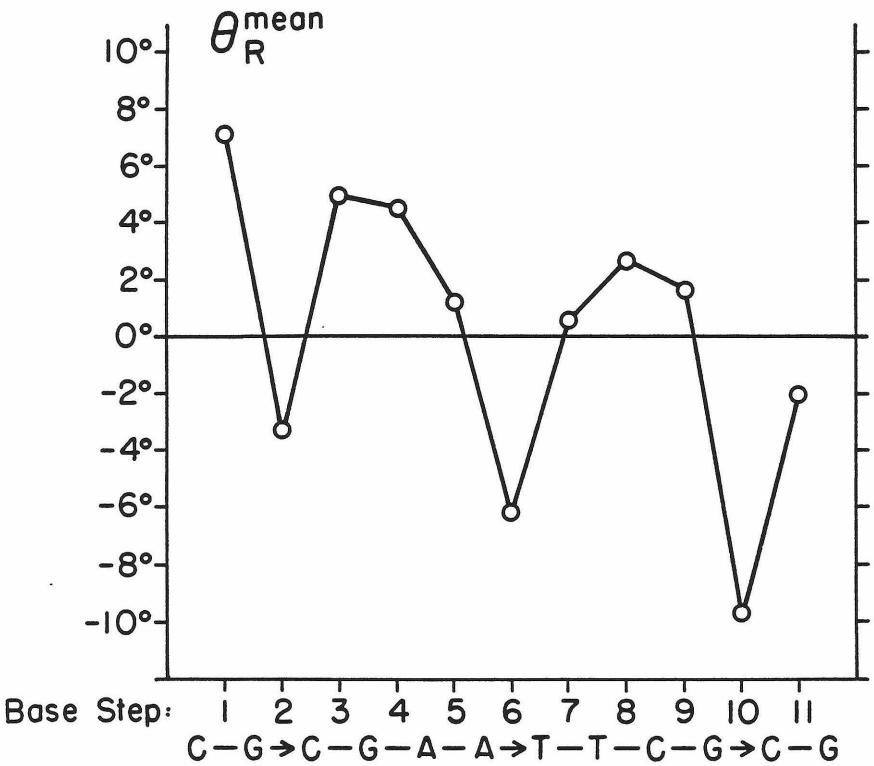

θ_R^{mean}
10°
8°
6°
4°
2°
0°
-2°
-4°
-6°
-8°
-10°
Base Step: 1 2 3 4 5 6 7 8 9 10 11
C—G→C—G—A—A→T—T—C—G→C—G

Figure 10

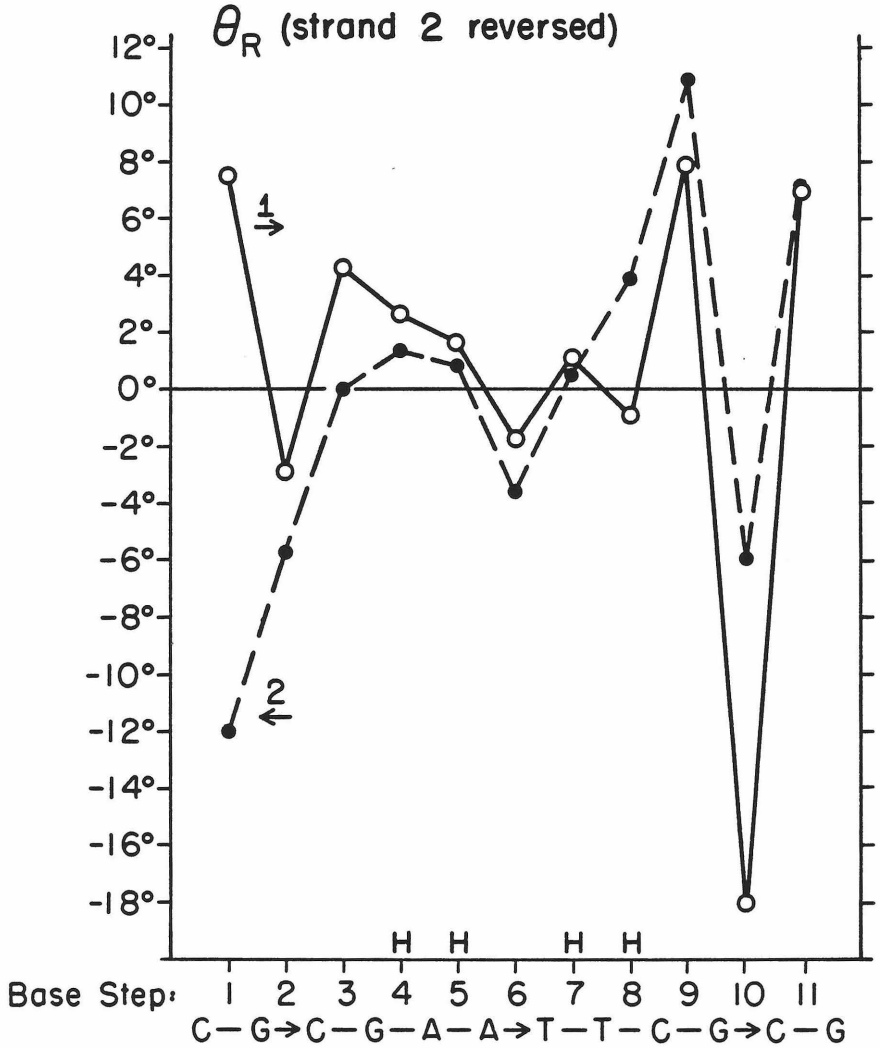

θR (strand 2 reversed)
12°
10°
8°
6°
4°
2°
0°
-2°
-4°
-6°
-8°
-10°
-12°
-14°
-16°
-18°
1
2
H H H H
Base Step: 1 2 3 4 5 6 7 8 9 10 11
C—G→C—G—A—A→T—T—C—G→C—G

Figure 11

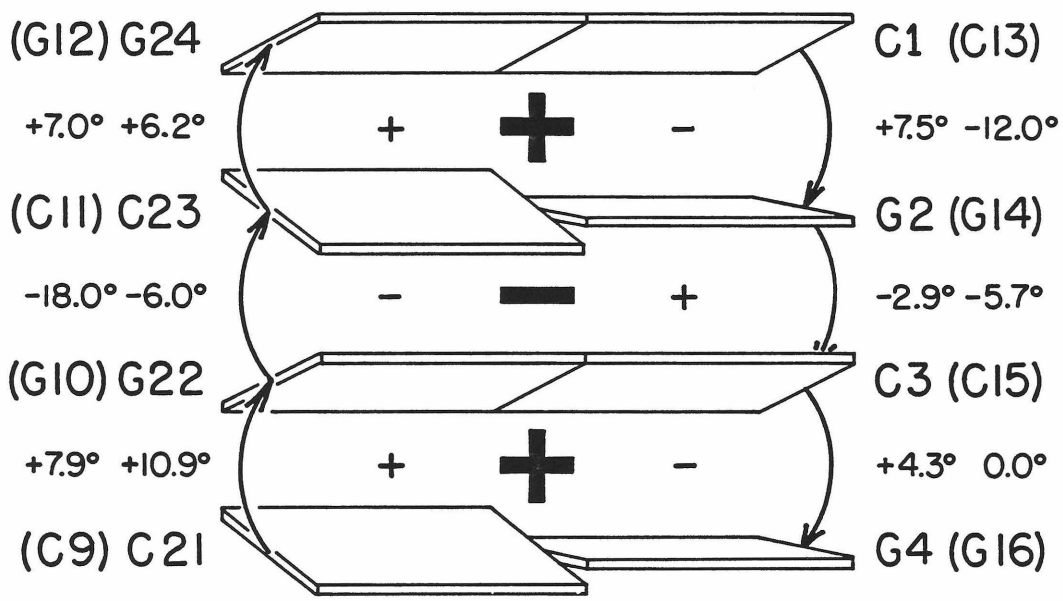

(G12) G24
+7.0° +6.2°
(C11) C23
-18.0° -6.0°
(G10) G22
+7.9° +10.9°
(C9) C21
C1 (C13)
+7.5° -12.0°
G2 (G14)
-2.9° -5.7°
C3 (C15)
+4.3° 0.0°
G4 (G16)

Figure 12

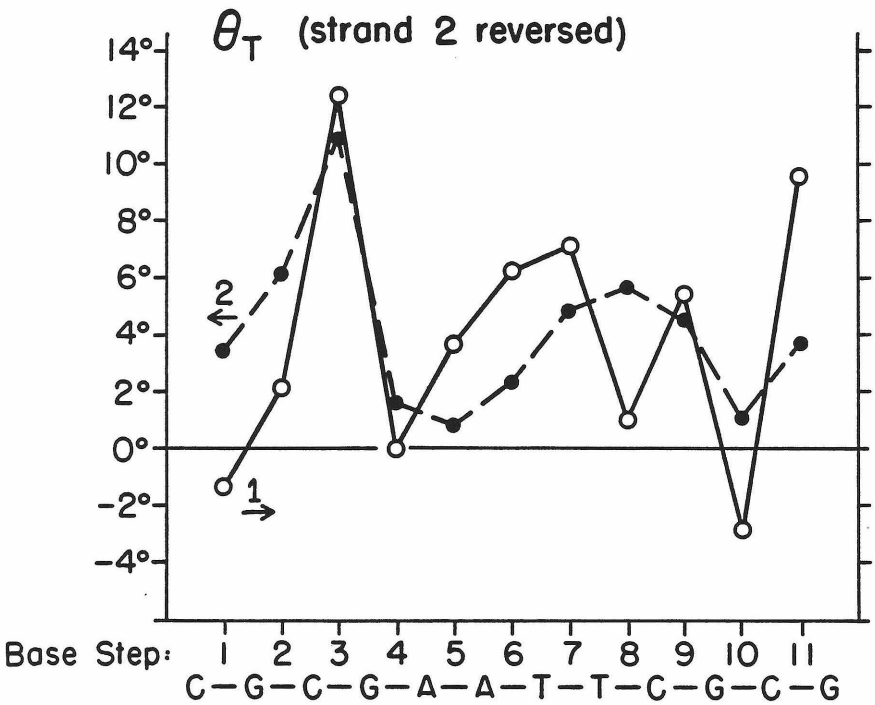

θT (strand 2 reversed)
14°
12°
10°
8°
6°
4°
2°
0°
-2°
-4°
2
1
Base Step: 1 2 3 4 5 6 7 8 9 10 11
C—G—C—G—A—A—T—T—C—G—C—G

Figure 13

2.5 HYDRATION OF B-DNA

Structure of a B-DNA Dodecamer. III. Geometry of Hydration

HORACE R. DREW and RICHARD E. DICKERSON

Norman W. Church Laboratory of Chemical Biology,
California Institute of Technology,
Pasadena, California U.S.A. 91125

Submitted to the Journal of Molecular Biology

Running Title: Hydration of B-DNA

SUMMARY

The dodecamer d(CpGpCpGpApApTpTpCpGpCpG) or CGCGAATTCGCG crystallizes as slightly more than one full turn of right-handed B-DNA. It is surrounded in the crystal by one bound spermine molecule and 72 ordered water molecules, most of which associate with polar N and O atoms at the exposed edges of base pairs. Hydration within the major groove is principally confined to a monolayer of water molecules associated with exposed N and O groups on the bases, with most association being monodentate. Waters hydrating backbone phosphate oxygens tend not to be ordered, except where they are immobilized by 5-methyl groups from nearby thymines. In contrast, the minor groove is hydrated in an extensive and regular manner, with a zigzag "spine" of first- and second-shell hydration along the floor of the groove serving as a foundation for less regular outer shells extending beyond the radius of the phosphate backbone. This spine network bridges purine N3 and pyrimidine O2 atoms in adjacent base pairs. It is particularly regular in the AATT center, and is disrupted at the CGCG ends, in part by the presence of the N2 amino groups on guanines. The minor groove hydration spine may be responsible for the stability of the B form of polymers containing only A/T and I/C base pairs, and its disruption may explain the ease of transition to the A form of polymers with G/C pairs.

1. Introduction

Three families of double helical structures are known for DNA: the A, B and Z helices, and the relative stabilities of these structures depend both on DNA sequence and on water activity (Franklin and Gosling, 1953; Malenkov et al, 1975; Drew et al, 1980). The familiar Watson-Crick B-DNA is a right-handed double helix with 10 base pairs per turn (Wing et al, 1980). A-DNA is also right-handed, but has 11 base pairs per turn and a slightly different disposition of base pairs about the helix axis (Arnott, 1976). Z-DNA, seen so far only in alternating pyrimidine-purine sequences, is a novel left-handed structure with 12 base pairs per turn (Wang et al, 1979; Drew et al, 1980). The B form predominates in aqueous solution to the exclusion of A and Z, but in alcoholic or high-salt solutions the situation is reversed for reasons that are imperfectly understood (Pohl and Jovin, 1972; Ivanov et al, 1974; Pohl, 1975).

There are several ways by which an aqueous environment might favor adoption of the B form. If the exposure of hydrophobic surfaces were less in B-DNA than in A, for example, then the B form would be able to interact more favorably with water molecules. Calculations of relative atom exposures, however, have shown this not to be the case (Alden and Kim, 1979). Alternatively, specific hydrogen-bonded interactions with water might contribute to the stability of B-DNA (Lewin, 1967). Elaborate networks of specifically bound water molecules have been seen in crystals of a dinucleoside phosphate-drug

complex (Neidle et al, 1980), but the structure of water around longer oligomers remains uncertain.

Our single-crystal x-ray analysis of the dodecamer CGCGAATTCGCG has revealed many new details of DNA conformation and solvent structure. As Figures 1a and 1b illustrate, this molecule crystallizes as slightly more than a full turn of right-handed Watson-Crick B helix. The only structural perturbation attributable to crystal packing is a smooth 19° bend in the overall helix axis. This bend, produced by overlapping ends in the crystal, has a surprisingly small effect on local helix parameters, in comparison with that attributable to specific base sequence. The first two papers in this series have dealt with the conformation and solid-state dynamics of the double helix (Drew et al, 1981) and with sequence-induced deformations from ideally regular B-helix geometry (Dickerson and Drew, 1981). In this third paper we report the positions of ordered water molecules around CGCGAATTCGCG, and attempt to evaluate their contributions to helix stability.

2. Methods

a. Solution and Refinement of the Structure

CGCGAATTCGCG crystallizes from aqueous magnesium acetate solution in space group $P2_12_12_1$ with cell dimensions: a = 24.87 Å, b = 40.39 Å and c = 66.20 Å. The asymmetric unit is a complete double

helix (two dodecamer strands), and four of these make up the unit cell. With an assumed density of 1.5 g cm^{-3}, the crystals are 50% solvent by weight.

The structure of the dodecamer was solved by the method of multiple isomorphous replacement (Wing et al, 1980), and refined by simultaneous least-squares minimization of conformational energy and of disagreement with observed x-ray data (Jack and Levitt, 1978; Drew et al, 1981). All CuK_{α} reflections between resolution limits 8.0 Å and 1.9 Å were included in refinement: 5534 reflections, with 2725 above the two-sigma level. Structure factors were calculated by Fourier inversion of a model electron density map (Ten Eyck, 1977). The course of refinement was monitored continually by examination of Fourier and difference Fourier electron density maps, and brought to a close at Cycle 62 only when residual peaks on the difference map had fallen to the $1\frac{1}{2}$ electron range.

Final unweighted R factors are 17.8% for the two-sigma data and 23.9% for the zero-sigma data set. The root mean square error in atomic positions as estimated by the method of Luzzati (1953) is 0.27 Å. The r.m.s. error in distances between any two unconstrained atoms therefore will be $\sqrt{2}$ times this, or ca. 0.4 Å. Observed and calculated structure factors, calculated phases, atomic coordinates and individual isotropic temperature factors have been deposited with the Brookhaven Protein Data Bank. The final superimposed electron density and difference map sections from Cycle 62 have been deposited

with the National Auxiliary Publication Service of the American Society for Information Science, from which copies can be obtained.[1]

b. Identification and Refinement of Ordered Solvent

Solvent molecules were added by examination of Fourier and difference Fourier maps during the course of refinement. They were added only where well-shaped positive peaks appeared in both Fourier and difference Fourier maps, preferably in stereochemically likely hydrogen-bonding positions. Solvent peaks were deleted thereafter if they did not return strongly in subsequent maps. A wrongly included water molecule typically returned in later maps as a weak Fourier peak, encircled by a distinctly negative difference peak. High-occupancy solvent, most of which were found to reside near the DNA, were added first, and low-occupancy solvent, the majority of which lie some distance from the DNA, were added later.

For computational simplicity, all solvent peaks were treated as being appropriately positioned full-occupancy water molecules (8 electrons), with only isotropic temperature factors subjected to refinement. At this limited resolution, occupancies and temperature factors are so strongly coupled that it would have been meaningless to refine both independently. Initial approximate water molecule positions as derived from the maps were refined as free atoms during stereochemically restrained refinement of the DNA. It was found that water molecules having the highest temperature factors did not shift

automatically into their electron density maxima, probably because the difference Fourier gradients around their positions were so shallow. These were repositioned manually if the maps showed them to be misplaced.

Figures 2a-d illustrate the appearance of the Fourier and difference Fourier maps during the course of refinement and addition of water molecules. No waters had yet been added at Cycle 15 in Figure 2a. During subsequent refinement steps culminating in Cycle 62, 80 solvent peaks numbered 25-104 were added to the structure. (Numbers 1-24 were used to identify the 24 bases of the double helix.) These were added slowly where indicated by both Fourier and difference Fourier maps, letting the gradually improved phases develop clearer suggestions of new peaks, rather than assigning water molecules to every positive feature early in refinement and removing mistaken choices later. Water peaks 27, 36 and 56 are clearly indicated both by Fourier and difference Fourier peaks in Cycle 15 (Figure 2a). But peak 56 was not added immediately, since its coordination geometry to other atoms was not yet obvious. The Fourier map at Cycle 15 also shows a peak at what eventually would be water 57, but this water was not added then because a corresponding peak was missing from the difference map. After a difference map peak developed in Cycle 35 (Figure 2b), water 57 was added to the list (Figure 2c). Water 56 finally was added after Cycles 15-35 had established peak 47, through which it could be coordinated by chemically reasonable O-H$\cdots$O distances to peak 27 and from there to two thymine C2 oxygens.

It is significant for the discussion that will follow that the minor groove waters in general developed first, with major groove peaks becoming clearly defined only at later stages of phase improvement. Of the entire crescent of peaks 36-74-84-66-28-91-57 around the major groove in Figure 2d, only peaks 36 and 84 were suggested by both Fourier and difference Fourier maps at Cycle 15. A few other tempting peaks in the solvent region between molecules in Cycle 15 disappeared from the final maps, and introducing them indiscriminately at the stage of Cycle 15 probably would have introduced phasing errors that would be difficult to eliminate later.

The foregoing procedures led to the gradual addition of 80 ordered solvent peaks with isotropic temperature factors (B values) ranging from 40 to 74 Å^2. As mentioned earlier, we elected to assign 8 electron occupancy to all water peaks and to refine B values, rather than fixing a B value and refining occupancy. Partial occupancy therefore is reflected in a smearing out of density via a large B value. The positive features still remaining on the difference map sections after Cycle 62, if interpreted as representing 8 electrons, would require temperature factors greater than 74. For comparison purposes, the average temperature factor for the entire structure is 36, and individual temperature factors from least-square refinement average 28 for bases, 42 for sugar atoms, and 51 for phosphate groups. All of these 80 solvent peaks, except for the eight that were interpreted as representing a spermine molecule (see below), have coordination geometries characteristic of water. Magnesium ions must be

present in the crystal in sufficient numbers to neutralize the 22 negative charges of each DNA double helix (nine Mg^{2+} in addition to the +4 charge from a single spermine molecule), but their positions apparently are disordered, and are not discernible. One reason for this may be that phosphate groups from different strands or different helices do not approach one another closely enough in this crystal to chelate a magnesium ion. In contrast, in high-salt CGCG, several magnesium ions were found where two phosphate groups came into close proximity (H. Drew, unpublished results).

c. Energy Refinement of Spermine

During refinement an extended electron density maximum appeared in the Fourier and difference Fourier maps, and its shape at first was represented in phasing by eight closely spaced water molecules, numbered 25, 54, 62, 63, 76, 77, 79 and 86 in the Brookhaven atomic coordinate list. After 62 cycles of positional and temperature factor refinement, an idealized spermine molecule was fitted manually into this snake-like electron density, taking due account of potential hydrogen-bonded contacts. Coordinates then were refined against bond lengths and angles derived from the x-ray structure analysis of spermine phosphate hexahydrate (Levitt and Lifson, 1969; Iitaka and Huse, 1965). A mild torsional potential was included in the energy parameters to encourage a trans orientation about each of the 13 C-C and C-N single bonds in the molecule. The root mean square shift from

this geometry optimization was 0.5 Å, with the largest shift of 0.8 Å at the disordered end of the chain, farthest from the DNA.

3. General Features of CGCGAATTCGCG Hydration

Relatively few of the water molecules that associate hydrodynamically with DNA in solution exhibit reduced thermal motion (Hearst and Vinograd, 1961; Tunis and Hearst, 1968). These few, and any others that may have been localized by the act of crystallization, are the only solvent that one can expect to position accurately from the information present in single-crystal x-ray data. Because their positions are poorly defined, the more loosely bound solvent molecules that fill the channels between helices do not contribute significantly to x-ray scattering except at low resolution.

In spite of this pessimistic prelude, we found 72 ordered water molecules around each CGCGAATTCGCG double helix in the crystal. Fifty of these, or an average of 4 per base pair, lie within the grooves of the helix or within the first coordination shell of phosphate groups. In Figures 3 and 4, these 50 water peaks are drawn as lightly shaded spheres of diameter 2.8 Å. The caterpillar-like stippled horizontal object spanning the upper end of the helix is the spermine molecule, which can be seen most clearly in Figures 3a and 4b. The wedge-like masses with dashed stippling at the upper end of the helix in Figure 4a and at the lower end in Figure 4b are four-base-pair portions of neighboring double helices up and down the 2_1 screw axis parallel to $\underline{c}$.

a. Disordered Solvation of Phosphate Groups

Techniques such as gravimetric analysis and infrared spectroscopy have established that the negatively charged phosphate oxygens bind water more tightly than do the deoxyribose oxygens or the base pair oxygens and nitrogens (Falk et al, 1962, 1963). This does not require that the positions of waters around the phosphate groups be unique, and as shown in Figures 3a and 3b, relatively few ordered water molecules are to be found occupying the first coordination shells of phosphate oxygens. Most of the well-ordered waters fill the major and minor grooves of the helix, where they are associated with polar N and O atoms on the edges of base pairs. One notable example of ordered water molecules coordinated to phosphate oxygens occurs when the water is immobilized by interaction with the 5-methyl group of thymine, as will be discussed later.

b. Ordered Solvation of Helix Grooves

The easiest introduction to the geography of the ordered water molecules around CGCGAATTCGCG is obtained by following the course of each groove down the entire length of the helix. Beginning at the top of Figure 3a, where spermine stretches from a phosphate on one strand to a phosphate on the other, the major groove winds down across Figures 4b, 3b and 4a, returning to the bottom of 3a once more. Exposed major groove oxygen and nitrogen atoms on both A/T and G/C

base pairs (adenine N7 and N6, thymine O4, guanine N7 and O6, cytosine N4) are solvated uniformly except where blocked by spermine, but relatively few water molecules are found bridging two such atoms in a geometrically regular manner.

A similar tour down the minor groove begins at the top of Figure 3b, and continues with 4a, 3a, 4b and back to the bottom of 3b again. In the AT-rich middle section of this groove (Figure 3a) the water molecules have a highly ordered and regular structure that will be described below, but in the CG-rich ends this regularity is destroyed by the overlap of helices. The double helices are securely fastened to one another in the overlap region by six intermolecular hydrogen bonds (Dickerson and Drew, 1981), and the CGC minor grooves are brought so close by these bonds that no room for solvent remains between them.

The exceptional concentration of ordered water within the minor groove in the AT-rich center cannot be attributed to crystal packing. The nearest neighboring double helix is more than 10 Å away across a channel of disordered water (see Figure 2d, for example), and dehydration of the minor groove at the CGC ends within the same helix does not seem to affect the structure of water at A/T base pairs. Although the space-filling drawings of Figure 3 do not show why ordered waters should cluster in the AT-rich region, a simple explanation can be found based on the arrangement of polar groups within the minor groove.

4. Sequence-Specific Features of CGCGAATTCGCG Hydration

a. Water Geometry in the Minor Groove: The Spine

The AATT minor groove has three to four layers of ordered water, of which the two outermost are easily visible in Figure 3a. Some of these waters occur near phosphate oxygens, others in the space between strands. The inner two layers, on the other hand, are arranged in a regular manner as shown in Figure 5. They provide a well-defined hydrophilic surface to which third- and fourth-shell solvent molecules can adhere readily. First-shell water molecules form bridges between thymine O2 and adenine N3 atoms. However, the bridge does not extend between bases in one base pair. Instead, it occurs between a base on one strand, and the first base following its hydrogen-bonded partner on the other strand in a 5'$\longrightarrow$3' direction. This means that in any paired dinucleotide step, it is the second of the two bases in each strand that is bridged; in an ApT step, the two thymines are connected. As Figure 5 shows, the twist of the right-handed helix brings the O2 or N3 groups of these bases into close proximity within the minor groove. These first-shell water molecules in turn are bridged by second-shell waters that give them a local near-tetrahedral coordination. Beyond this layer, the geometrical regularity of solvent structure begins to break down. Third- and fourth-shell water molecules (not shown in Figure 5) fill the remainder

of the minor groove in an irregular manner, frequently bridging across to phosphate oxygens along the walls of the groove.

If the helical minor groove were to be unrolled and stretched flat, then the first two hydration shells and the base atoms to which they are bound would appear as in Figure 6. In the center of this zigzag "spine" of water molecules the geometry is approximately tetrahedral except for a closing-down of the angle between bonds to the base O or N atoms, and a corresponding opening-up of the angles along the top of the spine (e.g., Nos. 35-27-33). Distances in the central region are acceptable as hydrogen bonds, if allowance is made for a r.m.s. uncertainty of 0.4 Å in distances between unconstrained atoms. But as one moves to either end of the helix, the first hydration shell pulls away from the floor of the minor groove. The angle between bonds to the bases becomes more acute, and the interatomic distances become too large to be described as hydrogen bonds. Even farther to the left and right of Figure 6, the water network frays away from the minor groove entirely, as the groove becomes blocked by interlocking ends of neighboring helices.

b. AT-Specificity of Minor Groove Hydration

The observed regularity of hydration geometry in the AT-rich center of the spine probably arises from the near-equivalence of adenine N3 and thymine O2, to which the water molecules are attached.

As Figure 7 shows, N3 and O2 occupy nearly isomorphous positions relative to the glycosyl bonds leading to the helix framework. Replacing an A/T base pair by T/A changes the N and O positions by less than 0.4 Å (Arnott and Hukins, 1972). Such a replacement would have little effect on the geometry of polar nucleophilic groups within the minor groove. The water spine seen here may be a common feature of all AT-rich segments of B-DNA, regardless of their particular sequence.

Introduction of a N2 amino group on the purine ring in guanine (Figure 7b) is doubly disruptive: It brings a bulky side group into the minor groove, and one that by nature is a hydrogen donor, not an acceptor. It can be seen in Figures 3 and 4 that the minor groove becomes drier and drier as it winds from the AATT center toward either CGCG end, with fewer third- and fourth-shell water molecules present. This trend also is reflected in the relative occupancies of first- and second-shell waters as measured by refined B values. In the center of the spine, B's are typically 40-50 (Figure 6b), whereas they rise to 70 and greater at the two ends, indicating both positional smearing and decreased occupancy. Moreover, the response of the first hydration shell to the introduction of N2 amino groups is to pull away from the bottom of the minor groove, as coordination shifts from the N3 of a purine to the N2 (compare the two ends of Figure 6a). Inspection of the electron density map suggests that both ends of the water network actually continue beyond water molecules 88 and 71, but that the additional solvent molecules appear only as two great plumes of disordered waters.

c. Water Geometry in the Major Groove

Water molecules elsewhere than in the minor groove usually coordinate to the DNA in a monodentate fashion. Waters in the second hydration shell, where they are ordered enough to be visible, show no geometrical regularity comparable with the spine in the minor groove. Each base pair contributes three polar groups to the major groove: N7-O6/N4 for G/C pairs and N7-N6/O4 for A/T (Figure 7). Table 2 is an attempt to show these three sites at each base pair as though one were looking into the major groove, with the sugar-phosphate backbones to right and left, and a lattice of polar groups between them. The numbers identify the water molecules associated with each polar site, N or O. These sites nearly always are associated with nearby water molecules, at distances that are not unreasonable for hydrogen bonds. Hydration in the AATT center of the major groove is shown in Figure 8. At three of the four adenines, a single water molecule forms a bi-dentate association with both N3 and N2 (water Nos. 49, 66 and 81). In addition, three of the four thymines appear to use their 5-methyl groups to freeze out the thermal motion of a water molecule that is bound to a nearby phosphate oxygen (Nos. 36, 57 and 58). Such an interaction is of potential importance in understanding the structural consequences of DNA methylation (Razin and Riggs, 1980), but its meaning is not yet clear. It is possible that a longer stretch of pyrimidines on the same strand, all with 5-methyl groups, might induce

formation of a cooperatively ordered solvent complex comparable with the minor groove spine.

The general pattern at G/C base pairs is for cytosine N4 and guanine O6 to bind water molecules independently, and for guanine N7 to be less hydrated. Of the eight guanine N7 positions listed in Table 2, two are blocked by spermine, three have no ordered water at all, one has only a water peak a full 3.9 Å away, and two share a water molecule with an adjacent base pair (G14) or with a neighboring molecule (G24). Some of these water molecules in CG-rich regions of the major groove are shown in Figures 9 and 10.

d. The Spermine Complex

Only one spermine molecule binds in an ordered manner per asymmetric unit, spanning the upper end of the major groove as depicted in Figure 9. The two $-NH_3^+$ ends of the spermine molecule form salt links with phosphate oxygens, and one of the two interior $-NH_2^+-$ groups forms a hydrogen bond to a guanine O6, and via a bridging water molecule, to a cytosine N4 on the next base pair down. The right end of the spermine molecule as shown in Figure 8 is 4 Å away from its phosphate oxygen and is positionally disordered, probably because it is equally attracted to a phosphate group on a neighboring molecule in the crystal. It might be a reasonable inference that in solution, away from the competition with neighboring molecules, the right end of the spermine would sit closer to its phosphate on this molecule.

The other end of the CGCGAATTCGCG dodecamer helix is chemically if not crystallographically identical to the spermine-binding end, and we can see no obvious reason why spermine should not bind there also. But as the electron density maps and Figure 10 indicate, it does not appear to do so. Attempts were made to interpret the solvent peaks in the corresponding area as a disordered image of a partially occupied spermine site, but these were not especially convincing. Inversion of the helix in Figure 10 to bring it into congruence with Figure 9, however, does point out similarities in binding of spermine and of water molecules, in particular the association of water 52 with a phosphate group on the left, and the vertical interaction (too long for a proper hydrogen bond) between waters 38 and 50, which in turn are connected to guanine O6 and cytosine N4 in a way reminiscent of the bridge through water 26 in Figure 9.

5. Discussion

This analysis of hydration around the B-DNA dodecamer CGCGAATTCGCG has shown several striking features:

i) Hydration within the minor groove is ordered, regular and apparently cooperative, with a geometric spine or network of water molecules in a quasi-tetrahedral environment, and with outer hydration shells filling the groove and frequently bridging to phosphate oxygens. The regularity of the network appears

to be favored by the absence of a purine N2 amino group and disrupted by its presence, suggesting that this spine of hydration occurs preferentially in AT-rich regions of DNA.

ii) Hydration within the major groove is generally monodentate and noncooperative, with first hydration shell waters associated with polar N and O groups, but little suggestion of a network of further hydration shells. Beyond the immediate ligands of the N and O atoms, most of the other water molecules filling the major groove seem to be positionally disordered.

iii) Water molecules associated with the phosphate backbone also appear to be structurally disordered, with multiple positions that are not well defined in the crystal structure analysis.

Lewin (1967) once postulated that structural water would contribute to the stability of B-DNA. The specific geometries that he then proposed from model-building considerations do not agree with our observations from the dodecamer, but the principle remains viable. All present data regarding the sequence-specific conformational stability of B-DNA suggest that water structure does have a part in stabilizing that form. Dehydration generally disfavors the B form relative to A or Z, but the transition away from the B family

is easiest in GC-rich sequences, and may be difficult or impossible in AT-rich polymers (Bram and Tougard, 1972; Pilet and Brahms, 1973; Pilet et al, 1975; Arnott, 1976; Drew et al, 1980). Leslie et al (1980) have carried out the most systematic study of this phenomenon, and their results are summarized in Table 3. No polymer lacking guanine could be induced to adopt a structure not in the B family except for poly(dA-dT)•poly(dA-dT), which occasionally gave an A-like pattern in a nonreproducible fashion but which always annealed with time to a B-family helix. Conversely, every polymer containing guanine could be driven into an A helix with the exception of poly(dA-dG)•poly(dC-dT), and those consisting of alternating purines and pyrimidines could also adopt the left-handed Z structure. Significantly, inosine (hypoxanthine), which is like guanine except for deletion of the N2 amino group (Figure 7), behaves like adenine. All polymers containing I/C base pairs remain in the B family, and cannot be driven to A.

These results of Leslie et al (1980) and earlier workers, combined with our observations on the hydration state of the dodecamer CGCGAATTCGCG, suggest that the regular, ordered spine of water molecules in the minor groove of B-DNA is one of the main elements in its helix stability relative to A or Z. In the absence of a disruptive purine N2 amino group in A/T and I/C base pairs this spine is intact. The presence of the N2 amino in G/C, however, weakens and disrupts hydration within the minor groove, making the helix more susceptible to dehydration and to eventual transition to the A form, or to Z if

an alternation of purines and pyrimidines permits. An interesting test of this line of argument would be the preparation of DNA polymers containing 2-amino adenine (Figure 7a) in place of adenine. Such 2-aA/T base pairs should be as disruptive of the minor groove hydration spine as are G/C base pairs, and it should be possible to convert polymers containing them to the A form readily.

How cooperative is the formation and disruption of the spine? Figure 6 suggests that the integrity of the spine might depend heavily on nearest-neighbor interactions, and that interruption of the hydration pattern at any one point by intrusion of a G/C base pair could lead to fraying of the ends of the spine on either side. Hence this minor groove hydration could also contribute to the cooperativity of the B-to-A transition. A test of this hypothesis would be the preparation of two polymers of the same overall base composition, but with G/C base pairs distributed evenly down the sequence in one case, and clustered together in the other form so as to leave long regions of only A/T pairs. If the spine does indeed contribute significantly to cooperativity in the helix transition, then the distributed-G polymer should be easier to convert into the A form than that with long uninterrupted regions of A and T.

Acknowledgments

We should like to thank Lillian Casler for preparing all of the line drawings in this paper, and Charles Ray and the Caltech Booth

Computing Center for their willingness to consider exorbitant demands during the preparation of stereo drawings. This work was carried out with the support of National Institutes of Health grants GM-12121 and GM-24393, and National Science Foundation grant PCM79-13959. H.D. was also the recipient of a National Institutes of Health Predoctoral Traineeship. This is contribution No. 6395 from the Division of Chemistry and Chemical Engineering.

References

Alden, C. J. and Kim, S.-H. (1979) J. Mol. Biol. 132, 411-434.

Arnott, S. (1976) in "Organization and Expression of Chromosomes" (Allfrey, V. G., Bautz, E. K. F., McCarthy, B. J., Schimke, R. T. and Tissieres, A., eds.), pp. 209-222, Dahlem Konferenzen, Berlin.

Arnott, S. and Hukins, D. W. L. (1972) Biochem. Biophys. Res. Commun. 47, 1504-1509.

Bram, S. and Tougard, P. (1972) Nature New Biol. 239, 128-131.

Dickerson, R. E. and Drew, H. R. (1981) J. Mol. Biol., submitted.

Drew, H. R., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 286, 567-573.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1981) Proc. Natl. Acad. Sci. USA, in press.

Falk, M., Hartman, K. A. and Lord, R. C. (1962) J. Am. Chem. Soc. 84, 3843-3846.

Falk, M., Hartman, K. A. and Lord, R. C. (1963) J. Am. Chem. Soc. 85, 387-391.

Franklin, R. E. and Gosling, R. G. (1953) Acta Crystallogr. 6, 673-677.

Hearst, J. E. and Vinograd, J. (1961) Proc. Natl. Acad. Sci. USA 47, 1005-1014.

Iitaka, Y. and Huse, Y. (1965) Acta Crystallogr. 18, 110-121.

Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D. and Schyolkina, A. K. (1974) J. Mol. Biol. 87, 817-833.

Jack, A. and Levitt, M. (1978) Acta Crystallogr. A34, 931-935.

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

Levitt, M. and Lifson, S. (1969) J. Mol. Biol. 46, 269-279.

Lewin, S. (1967) J. Theoret. Biol. 17, 181-212.

Luzzati, V. (1953) Acta Crystallogr. 5, 802-810.

Malenkov, G., Minchenkova, L., Minyat, E., Schyolkina, A. and Ivanov, V. (1975) FEBS Lett. 51, 38-42.

Neidle, S., Berman, H. M. and Shieh, H. S. (1980) Nature 288, 129-133.

Pilet, J. and Brahms, J. (1973) Biopolymers 12, 387-403.

Pilet, J., Blicharski, J. and Brahms, J. (1975) Biochemistry 14, 1869-1876.

Pohl, F. M. (1975) Nature 260, 365-366.

Pohl, F. M. and Jovin, T. M. (1972) J. Mol. Biol. 67, 375-396.

Razin, A. and Riggs, A. D. (1980) Science 210, 604-610.

Ten Eyck, L. F. (1977) Acta Crystallogr. A33, 486-492.

Tunis, M.-J. B. and Hearst, J. E. (1968) Biopolymers 6, 1325-1344.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. (1979) Nature 282, 680-686.

Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 287, 755-758.

Footnotes

[1]See NAPS Document No. _____ for 66 pages of supplementary material (map sections). Order from ASIS/NAPS, Microfiche Publications; Post Office Box 3513; Grand Central Station; New York, New York 10017. Remit in advance $3.00 for microfiche copy or $17.00 for photocopy. All orders must be prepaid. Institutions and Organizations may order by purchase order. However, there is a billing and handling charge for this service. Foreign orders add $3.00 for postage and handling.

TABLE 1. Progressive Addition of Water Molecules During Refinement

Figure	2a	2b	2c	2d
Refinement cycle	15	35	39	62 (final)
R factor (2σ data)	26.6%	21.4%	20.5%	17.8%
Data included in maps	2σ	2σ	2σ	0σ
Fourier map:				
Contour range	30-510	20-510	20-510	20-440
Contour interval	60	70	70	60
Difference Fourier:				
Contours (±)	40, 70, 100	35, 65, 95	30, 60, 90	40, 70
Water molecules included				
in sections of Fig. 2:	--	27	27	27
	--	36	36	36
	--	47	47	47
	--	--	56	56
	--	--	57	57
	--	--	--	66
	--	--	--	74
	--	--	--	84
	--	--	--	85
	--	--	--	91

TABLE 2. Hydration of Polar Groups on Bases in the Major Groove

Hydration Positions on Bases:

G/C pairs:	N7	O6	N4
C/G pairs:	N4	O6	N7
A/T pairs:	N7	N6	O4
T/A pairs:	O4	N6	N7
	↓	↓	↓

Strand 1			Strand 2	
C1	--	--	42	G24
G2	spermine	spermine	43	C23
C3	spermine	spermine	spermine	G22
G4	--	32	26	C21
A5	81	102	80	T20
A6	66	66	91	T19
T7	(66)	28	103	A18
T8	(37)	49	49	A17
C9	50	37	--	G16
G10	noise	38	38	C15
C11	30	y	38	G14
G12	(99)	z	30	C13

NOTES:

This table is set up as though one were looking into the unrolled major groove, with DNA strand 1 at left and strand 2 at right, and with the three polar groups along the edge of a base pair in their proper relative positions. The identity of these three groups in each of the four base pair combinations is given at the top of the table. The zigzag line separates bases of each pair. Numbers in the body of the table identify water molecules as in the Brookhaven coordinate list. Numbers in parentheses indicate more distant waters or looser association. "Spermine" signifies either direct coordination to a spermine nitrogen, or physical blocking of other solvating molecules. "Noise" indicates low-level contours probably indicative of disordered water, and "--" means that the map is featureless at that point. y and z are two low solvent peaks that were observed in the final map but not included in the solvent list.

TABLE 3. Sequence Dependence of DNA Conformational Behavior

DNA Polymer	Purine N2 Amino Group Present?	Observed Helix Conformations
(A)(T)	No	B
(I)(C)	No	B
(A-I)(C-T)	No	B
(A-T)(A-T)	No	B, (A)
(I-C)(I-C)	No	B
(I-T)(A-C)	No	B
(A-A-T)(A-T-T)	No	B
(A-I-T)(A-C-T)	No	B
(A-I-C)(I-C-T)	No	B
(I-I-T)(A-C-C)	No	B
(G)(C)	Yes	B, A
(A-G)(C-T)	Yes	B
(G-C)(G-C)	Yes	B, A, Z
(G-T)(A-C)	Yes	B, A, Z
(A-A-C)(G-T-T)	Yes	B, A
(A-G-T)(A-C-T)	Yes	B, A
(A-G-C)(G-C-T)	Yes	B, A
(G-A-T)(A-T-C)	Yes	B, A
(G-G-T)(A-C-C)	Yes	B, A

Notes:

Adapted from Leslie et al (1980). The repeating unit in each polymer strand is given in conventional 5'→3' order within parentheses. "B" includes all members of the B family of helices: B, B', C, C', C'', D and E. "Z" includes the S helix. (A) indicates that the A form of poly(dA-dT) could not be obtained reproducibly; Pilet et al (1975) have noted that the A conformation of this sequence is stable only over a narrow range of relative humidity.

FIGURE CAPTIONS

FIGURE 1. Final refined structure of the dodecamer CGCGAATTCGCG. The 5' end of strand 1 (bases 1-12) and the 3' end of strand 2 (bases 24-13) appear at the upper left and right of the helix, respectively. Base pair C1/G24 is at the top, and C13/G12 at the bottom. Radii of different types of atoms decrease in the order: P, O, N, C. (a) Narrow groove view. (b) Wide groove view after 180° rotation about a vertical axis. The overall helix axis shows a 19° bend, concave to the left in (a) and to the right in (b).

FIGURE 2. Superpositions of sections $\underline{z}$ = 12/96, 13/96 and 14/96 for the combined Fourier and difference Fourier maps at four stages of refinement and addition of solvent molecules. (a) Cycle 15, R = 26.6%, no solvent in entire structure. (b) Cycle 35, R = 21.4%, 3 solvent peaks in these sections. (c) Cycle 39, R = 20.5%, 5 solvent peaks in these sections. (d) Final cycle 62, R = 17.8%, 10 solvent peaks in these sections. X = phosphorus, O = carbon, + = oxygen, Δ = nitrogen, ◇ = water peak. Peaks 74, 84 and 85 are water molecules that are labeled formally on symmetry-related map sections. Peaks 28 and 90 are located on sections 11/96 and 15/96, rather than on the sections shown here. Solid contours represent a $F_o\exp(\phi_c)$ electron density map, and short and long dashes are positive and negative contours of a $(F_o-F_c)\exp(\phi_c)$ difference map. Contour intervals and identities of water molecules included are to be found in Table 1. These sections cut obliquely through the center of the molecule, at an angle to the base planes, in a manner that includes the nonbonded bases thymine 7

(above) and thymine 19 (below), but not the adenines to which they are hydrogen bonded. Not all of the atoms of the thymine rings occur on these three sections, but outline bonds have been added for clarity. (Adjacent sections cannot be added to either side without introducing confusing density overlap from adenines 6 and 18.) Deoxyribose atoms C1', C2', C3' and O3' lead off from the left of each thymine ring, toward a dense region whose central phosphorus atom generally lies in other sections. The large density to left of center along the top margin is cytosine 1 of a neighboring molecule, and that to the right of the top margin is the 5'-terminus of strand 2 of another neighbor. One crystallographic repeat along the a axis can be seen at left and right.

FIGURE 3. Space-filling drawings of the dodecamer and its associated solvent molecules. Viewing directions are identical in Figures 1a and 3a, and in 1b and 3b. Individual ordered water molecules are represented by finely shaded spheres with 2.8 Å diameter. The horizontal shape with coarse stippling at the top of (a) is a spermine molecule. Of the 72 ordered water molecules found, only the 50 most intimately associated with the helix are drawn, omitting numbers 29, 40, 51, 59, 60, 64, 65, 68, 72, 73, 75, 83, 84, 85, 87, 89, 92, 93, 94, 95, 100 and 104 of the Brookhaven coordinate list.

FIGURE 4. Space-filling representations of solvent structure as in Figure 3, but with the addition of overlapping CGCG ends from neighboring molecules up and down the 2_1 symmetry axis, filling the

minor groove at the top of (a) and bottom of (b) (dashed stippling). Note that, because of the 2_1 screw symmetry axis, the bottom of the dash-stippled mass at the top of (a) is identical to the bottom of the unshaded helix in (b), and conversely that the top of the dash-stippled mass in (b) corresponds to the upper end of the unshaded helix in (a). Figures 3a, 4a, 3b and 4b represent successive 90° rotation steps about a vertical axis.

FIGURE 5. Stereodiagram of minor groove solvent geometry in the GAATTC middle of the double helix. Base pair G4/C21 is at the top of this hexamer fragment of the full structure, and C9/G16 at the bottom. The sugar rings of other selected bases are numbered. Crossed spheres are oxygen atoms of water molecules, whose presumed hydrogen bond interactions are drawn as thin lines. From upper right to lower left, the water molecules in the zigzag "spine" down the minor groove are numbered 88, x, 96, 78, 34, 35, 27, 33, 41, x, 98, x and 71 in the Brookhaven coordinate list. (x denotes a small peak, not included in that list, which was found later by inspection of the electron density map after Cycle 62.)

FIGURE 6. Unrolled view of the spine or backbone of water molecules that curves down the minor groove. Spheres at bottom are atoms from the paired bases (guanine N3 or N2, adenine N3, cytosine and thymine O2). These atoms are bridged by first hydration-shell water molecules 88-96-34-27-41-98-71, and these in turn are given roughly tetrahedral coordination geometry by second-shell water molecules

x-78-35-33-x-x. Dashed lines at bottom connect members of one base pair; the water bridging spans two base pairs. (a) Bond angles and atom identification numbers. (b) Interatomic distances in Å, and individually refined isotropic B or temperature factors within circles. x denotes a partial-occupancy water peak that would have to have a temperature factor greater than B = 74 Å.

FIGURE 7. Purine-pyrimidine base pairing and minor groove structure in the presence and absence of purine N2 amino groups. This N2 group is absent in a normal A/T base pair (top) but present in 2-amino adenine. Similarly, it is absent in I/C pairs (bottom) but present in normal G/C pairs. The minor groove spine of hydration schematized in Figure 6 apparently is possible whenever this N2 amino group is absent, but is disrupted by its presence. In normal DNA this means it can exist in AT-rich regions but not GC-rich regions.

FIGURE 8. Water coordination geometry in the AT-rich middle of the major groove. Selected base sugar rings are numbered for identification. Crossed spheres are water molecules. From left to right across the edge of A5/T20 these are Nos. 81, 102 and 80. Across A6/T19 are found waters Nos. 66, 91 and 57 (the latter bridging a thymine methyl and a phosphate oxygen). From left to right in the next row below are Nos. 36, 28 and 103, and 58 and 49 interact with T8/A17. Three of the four thymine methyl groups help immobilize a

water molecule (dotted lines to Nos. 36, 58 and 57) in a bridge to a nearby phosphate group.

FIGURE 9. Geometry of spermine binding at the upper end of the major groove. The spermine molecule is drawn with solid bonds and with crossed spheres, larger for N than for C. Its $-NH_3^+$ ends interact with phosphate groups following bases 1 and 21, and one of its central -NH- groups forms hydrogen bonds to the N2 of guanine 22 and to water molecule No. 26, which itself is hydrogen bonded to the N4 of cytosine 21.

FIGURE 10. Major groove water geometry in a CG-rich region. The drawing has been inverted in order to correspond as closely as possible to Figure 9, comparing chemically (but not crystallographically) identical regions in the presence and absence of spermine. In each figure, the top base pair is the end of the double helix, and the helix continues off the bottom of the drawing. The upper trio of water molecules from left to right is numbered 52, 61 and 38, and the bottom three are Nos. 37, 50 and 67, with No. 70 coordinated to a phosphate oxygen at the right. The vertical line connecting Nos. 38 and 50 is 4 Å long, too long for a true hydrogen bond, and is drawn only to call attention to similarity of binding of spermine and water to B-DNA. It is roughly analogous to the hydrogen bond between spermine -NH- and water 26 in Figure 9.

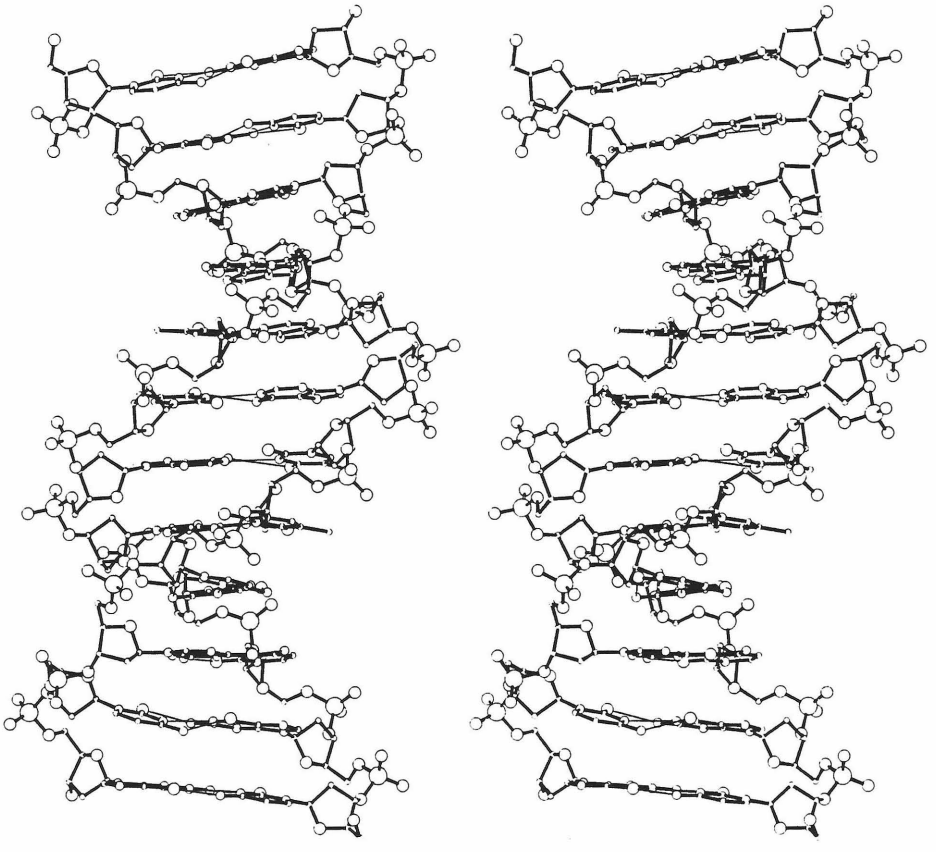


Figure 1a

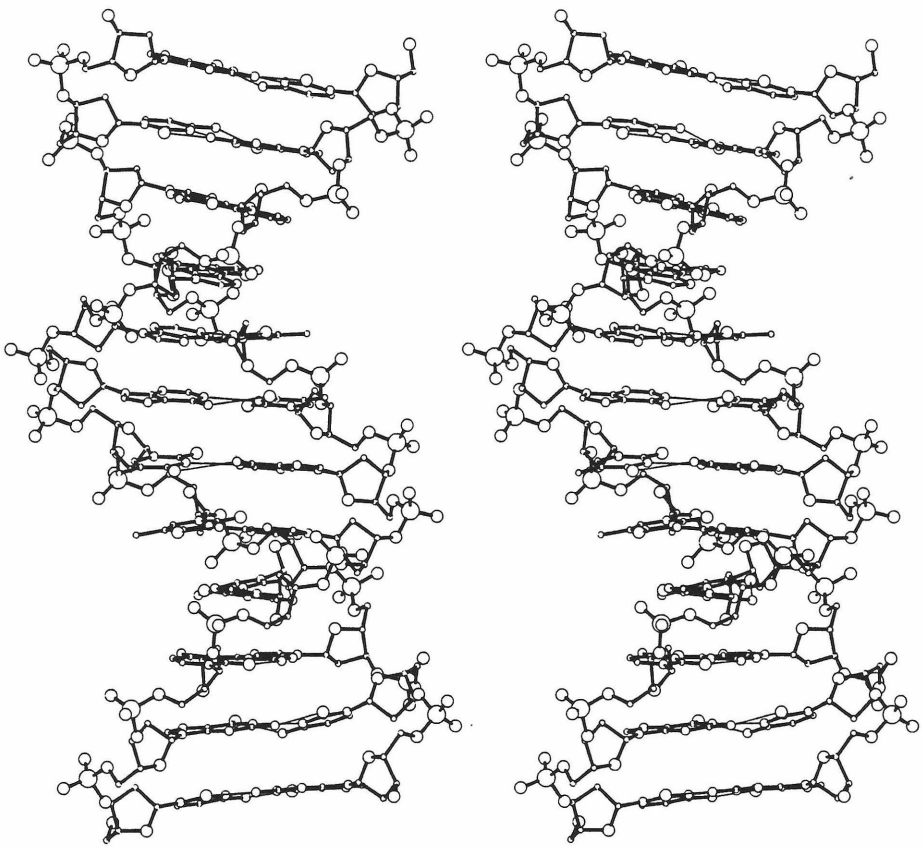


Figure 1b

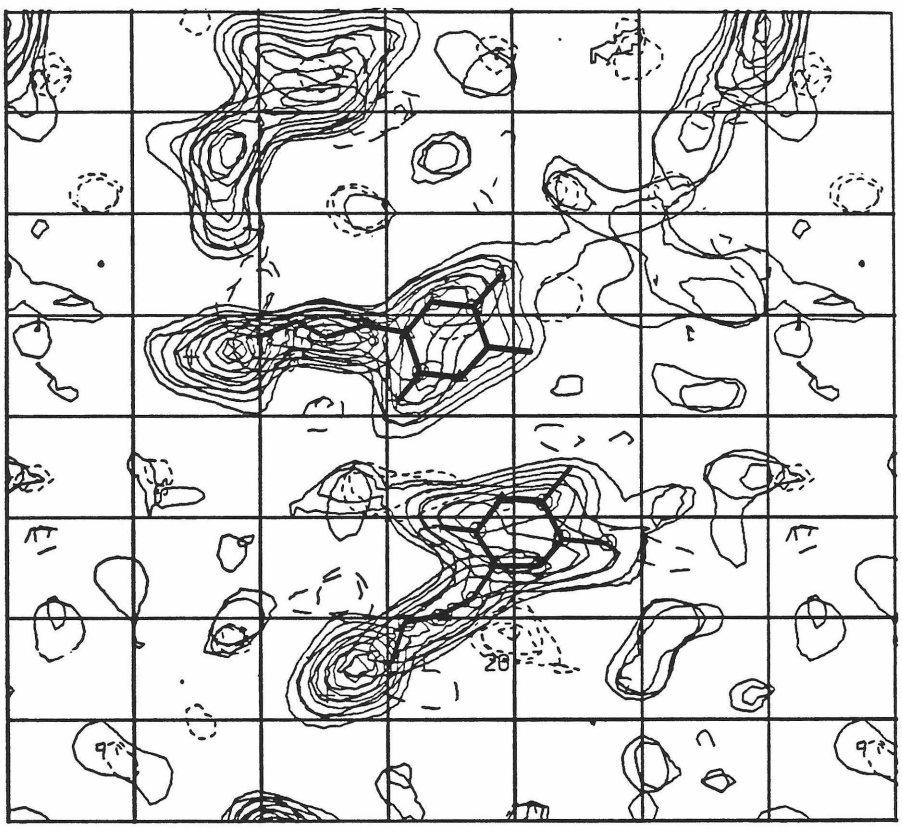


Figure 2a

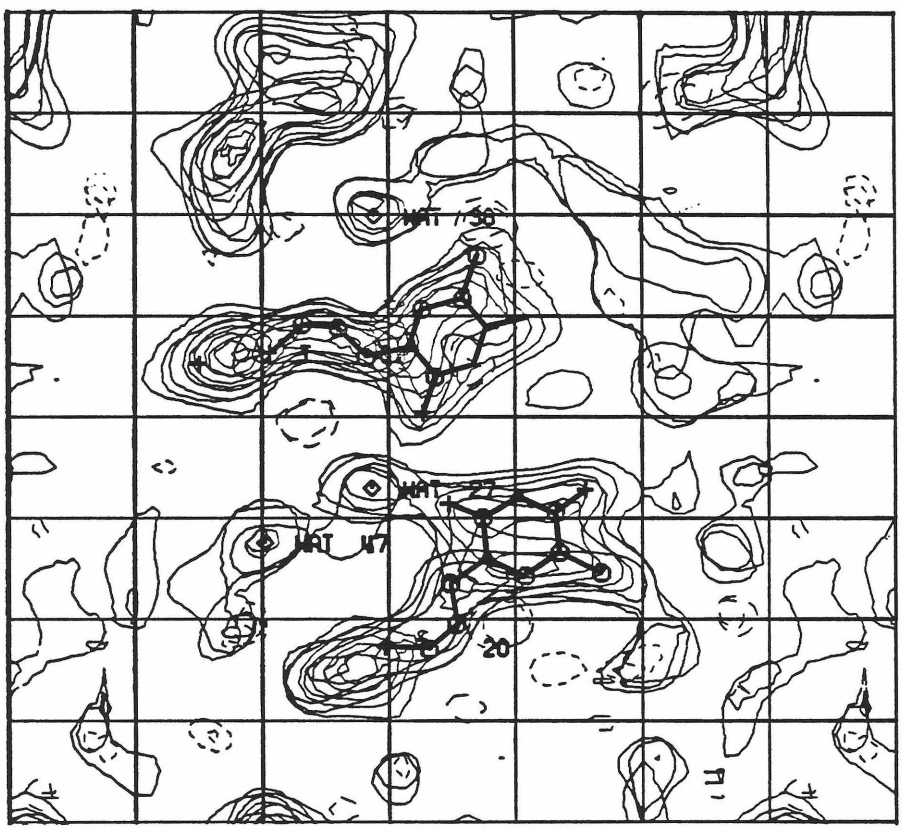


Figure 2b

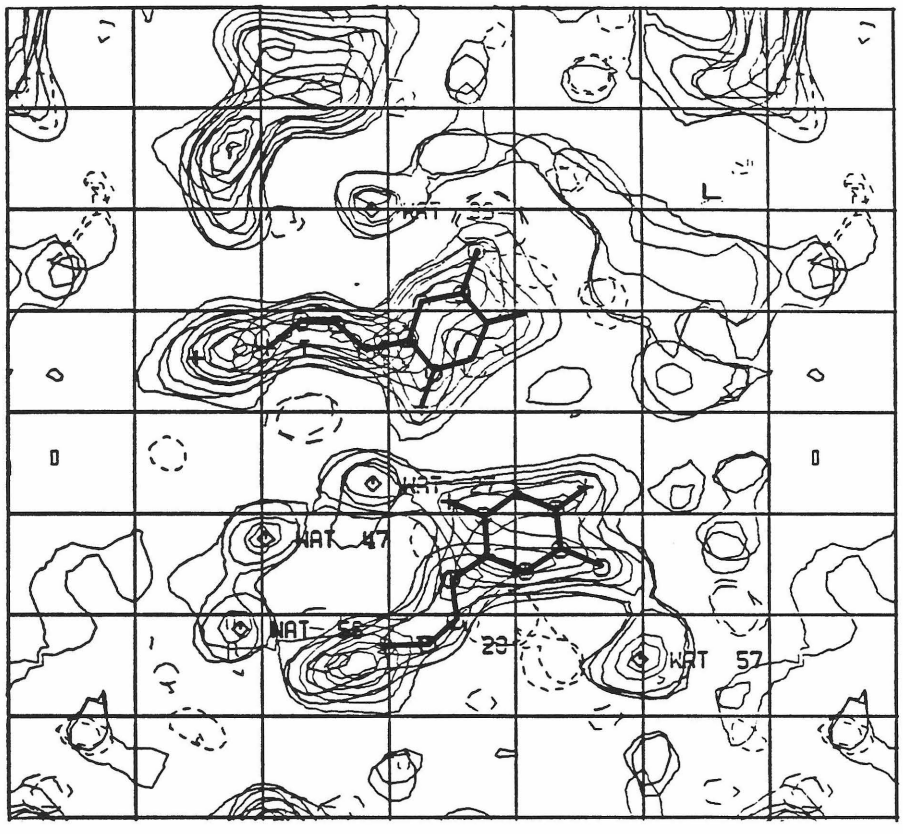


Figure 2c

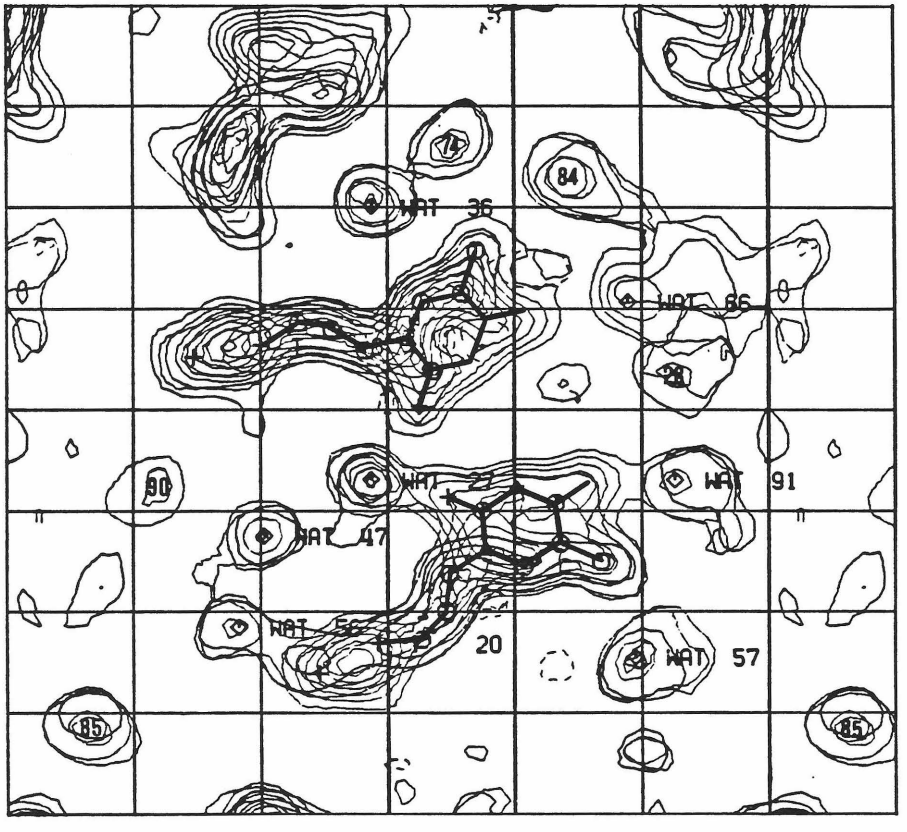


Figure 2d

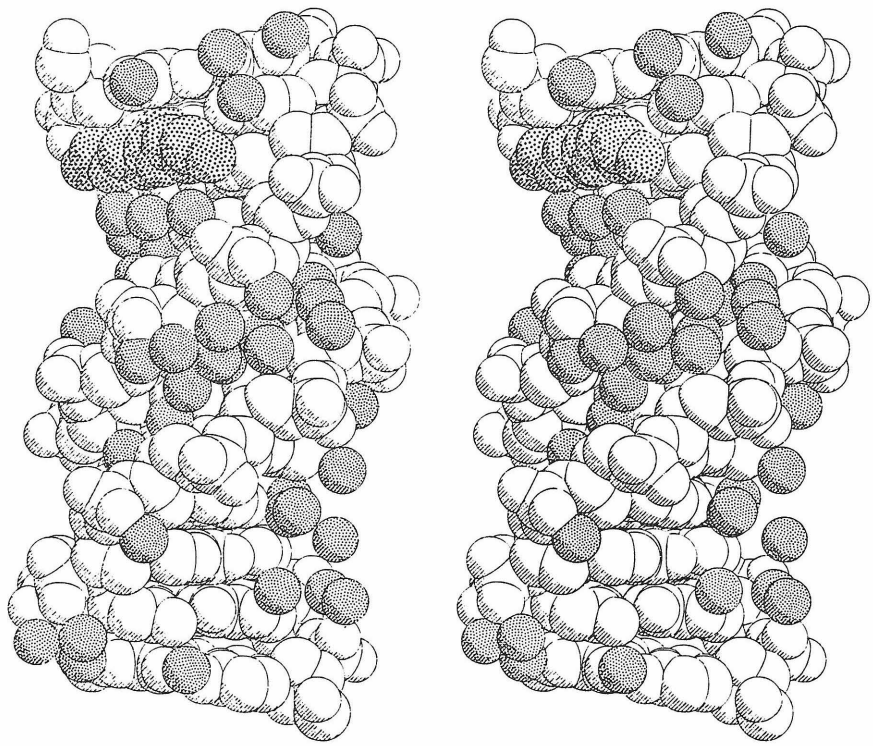


Figure 3a

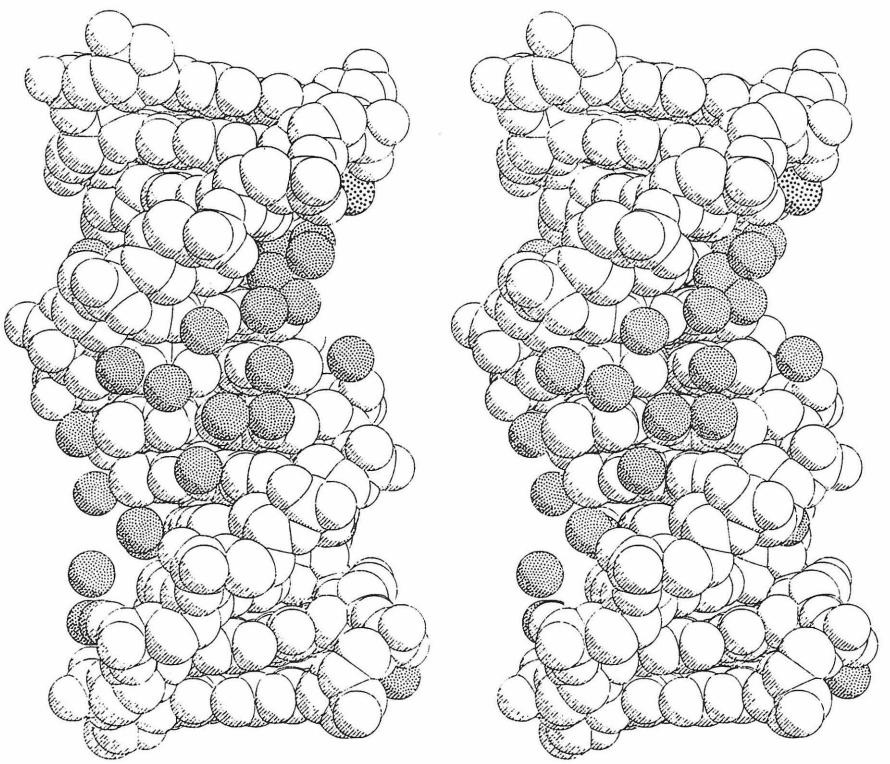


Figure 3b

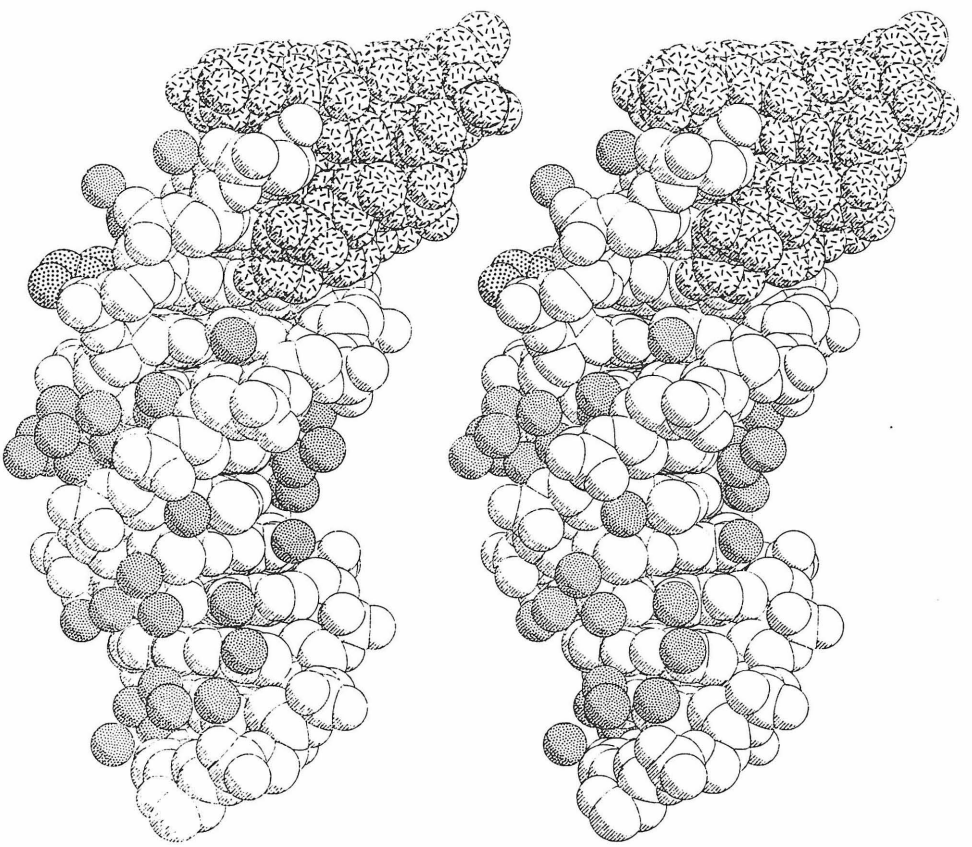


Figure 4a

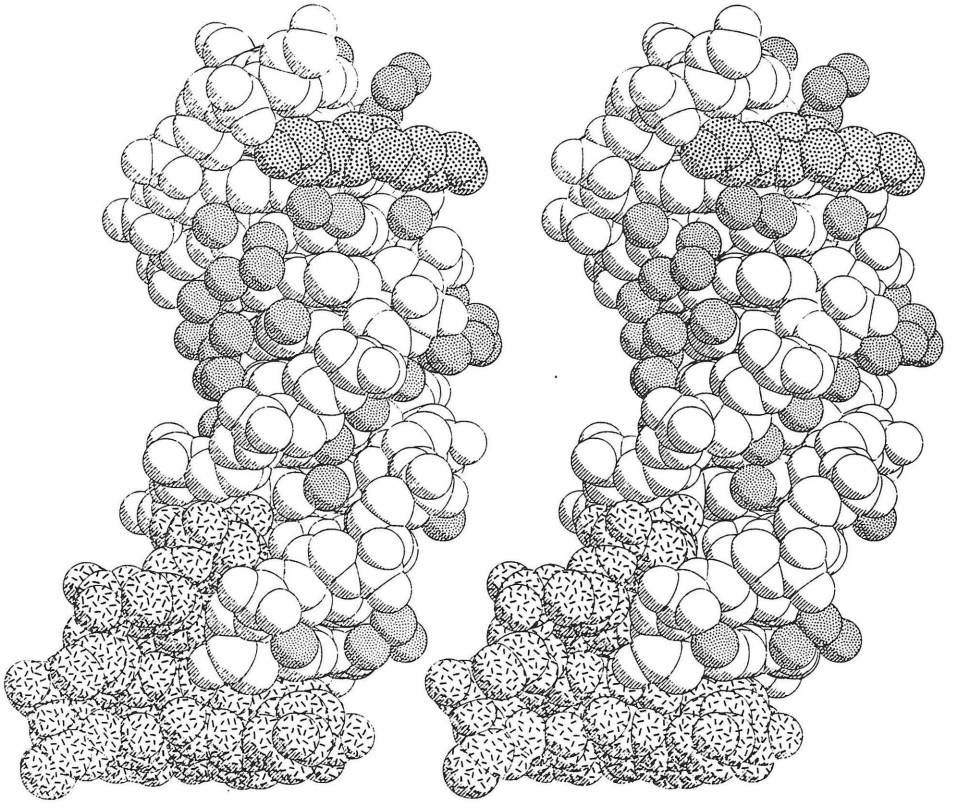


Figure 4b

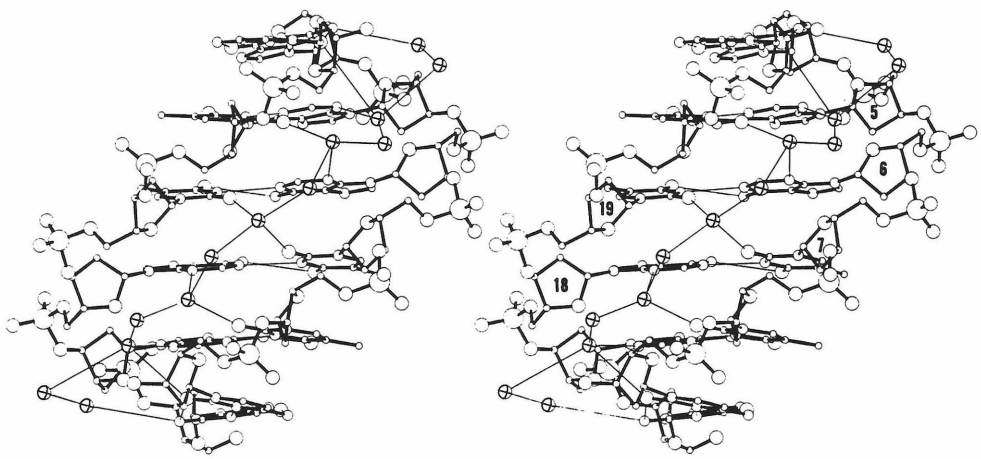


Figure 5

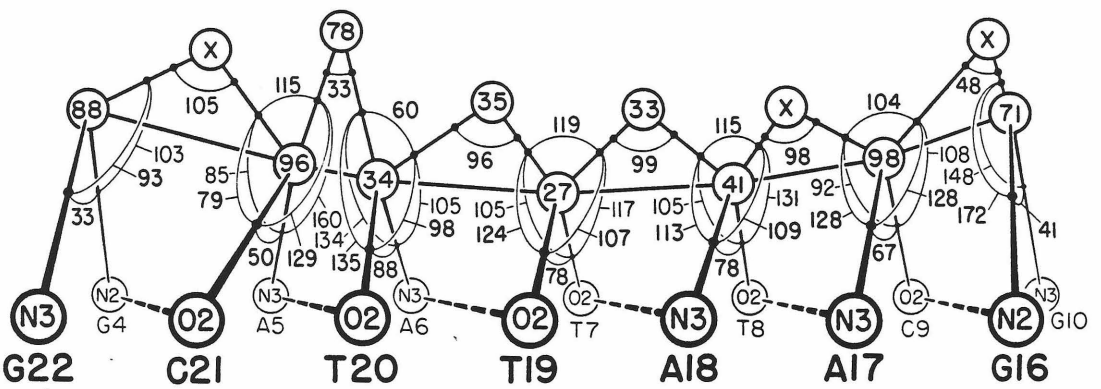

X
78
88
105
115
33
60
35
119
33
115
X
104
48
71
103
93
85
79
96
34
96
27
99
41
98
98
108
148
92
128
172
41
33
160
134
105
98
105
124
117
107
105
113
131
109
128
50
129
135
88
78
67
N2
G4
N3
A5
N3
A6
78
O2
T7
O2
T8
O2
C9
N3
G10
N3
G22
O2
C21
O2
T20
O2
T19
N3
A18
N3
A17
N2
G16

Figure 6a

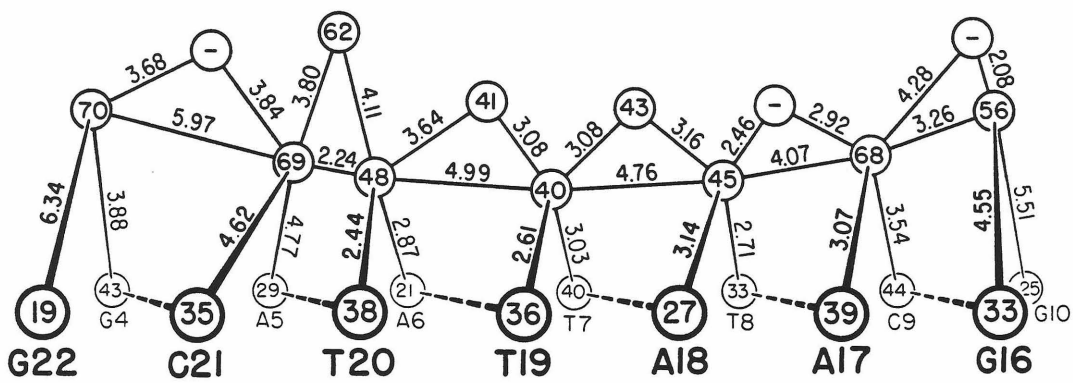

62
-
-
-
70
69
48
41
43
45
68
56
3.68
3.84
3.80
4.11
5.97
2.24
3.64
3.08
3.08
3.16
2.46
2.92
4.28
2.08
3.26
4.07
4.99
4.76
6.34
3.88
4.62
4.77
2.44
2.87
2.61
3.03
3.14
2.71
3.07
3.54
4.55
5.51
19
43
35
29
38
21
36
40
27
33
39
44
33
25
G4
A5
A6
T7
T8
C9
G10
G22
C21
T20
T19
A18
A17
G16

Figure 6b

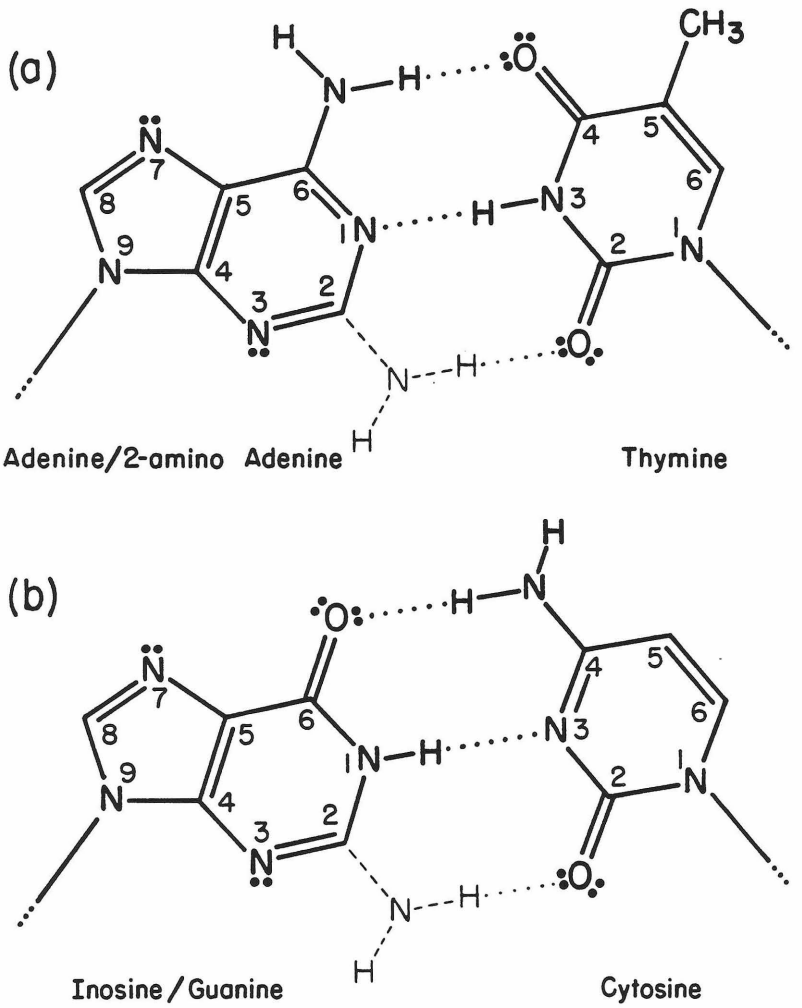

(a)
Adenine/2-amino Adenine
Thymine
(b)
Inosine / Guanine
Cytosine

Figure 7

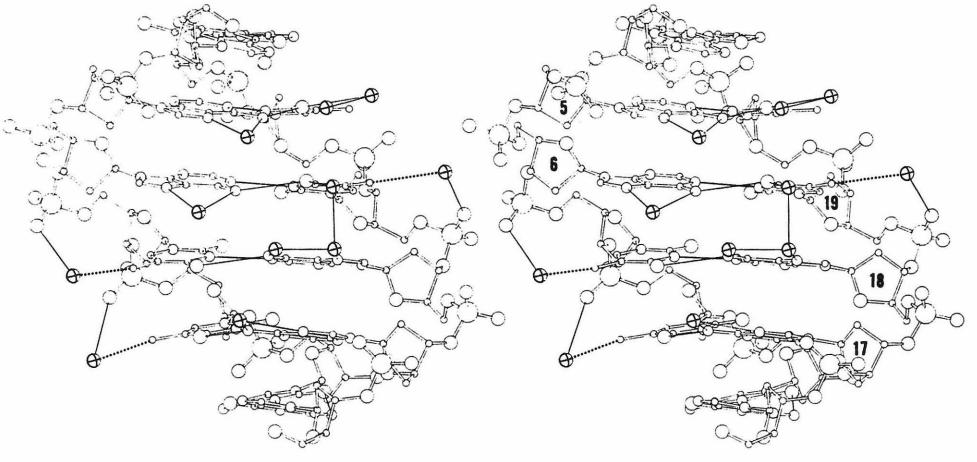


Figure 8

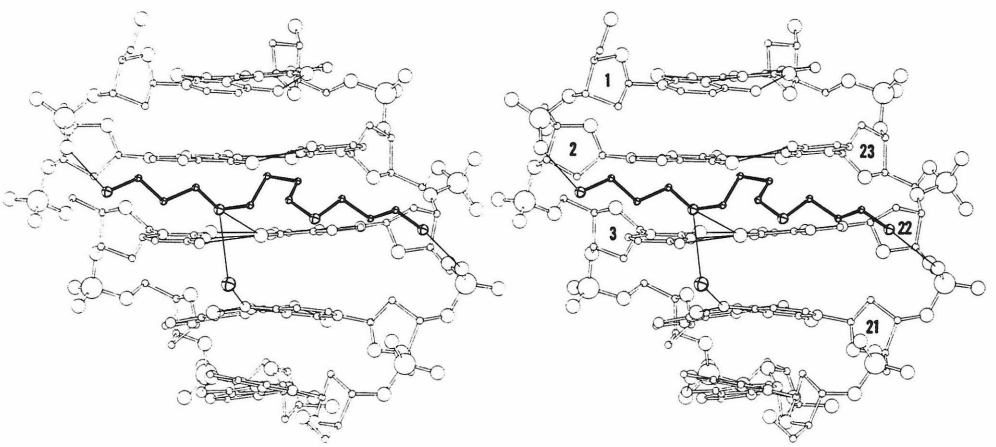


Figure 9

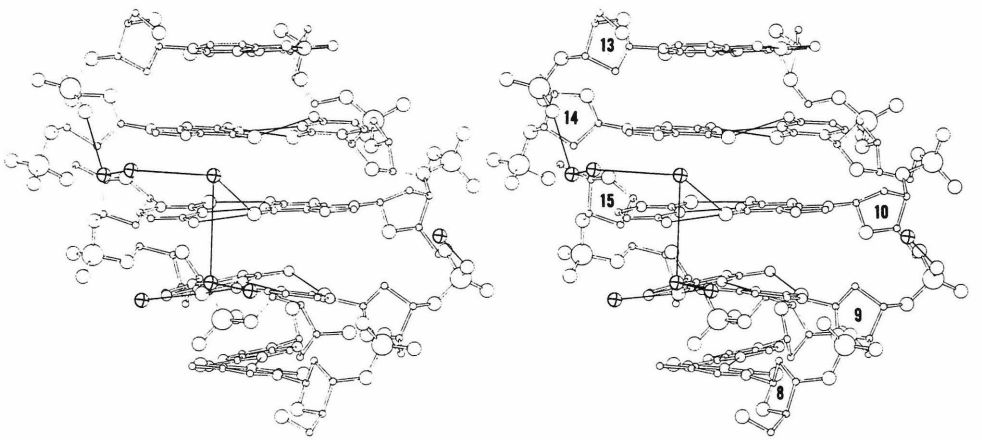


Figure 10

2.6 CONCLUSIONS

This single-crystal x-ray analysis of a DNA dodecamer, CGCGAATTCGCG, has provided the first high-resolution picture of a right-handed B-type double helix. The overall helix axis is isotropically bent into a radius of curvature of 110 Å, but this deformation has little effect on local helix parameters, which are more influenced by the sequence of bases in each chain. Sequence-induced variations from ideal B helix geometry are not large, but may be of importance in some types of recognition processes.

For example, the minor groove intercalator ethidium bromide is known to prefer a pyrimidine-purine site such as CpG or TpG over other sequences such as CpC or TpC, and also to induce a mixed sugar pucker in residues adjacent its binding site: C3'-endo at the pyrimidine, and C2'-endo at the purine. This sequence preference is readily understandable in terms of B helix fine structure. Pyrimidine-purine steps in B-DNA open slightly to the minor groove, where they provide a wedge-like preintercalation binding surface for a flat drug molecule; furthermore, pyrimidine-purine sites are predisposed to adopt a mixed sugar pucker, since pyrimidine sugars are already in a mixed O1'-endo/C1'-exo conformation, midway between C2'-endo and C3'-endo.

The structure of water around a B helix may also affect some of its properties. Ordered water molecules will cluster in the minor groove of any AT-rich sequence, and this may preferentially stabilize the B form over the alternative double-helical conformations A and Z. Certain antibiotic drugs, such as netropsin, distamycin A, and CC-1065,

are known to prefer the minor groove of an AT-rich B helix, and this binding specificity may also be related to solvent structure in a way which is not yet clear. It is at least conceivable that these drugs form hydrogen bonds to an ordered lattice of A/T-specific minor groove water molecules.

Both netropsin and CC-1065 induce similarly large structural changes in crystals of the dodecamer CGCGAATTCGCG, and the precise nature of this solid-state phase transition is under investigation (M. Kopka and A. Fratini, personal communication). The structures of these complexes will be of great interest, not only as models of drug specificity, but also as examples of how subtle changes in DNA conformation might be involved in transcriptional specificity.