Iodo-CCGG

A Single Crystal Structure of A-DNA

Thesis by

Benjamin N. Conner

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1982

(Submitted May 18, 1982)

Acknowledgements

I would like to thank my thesis advisor, Professor Richard E. Dickerson, for his advice and support throughout my graduate years. I would also like to thank Dr. Keiichi Itakura for his part in our productive collaboration, and Dr. Shoji Tanaka for his unconventional and informative introduction into the art of nucleic acid synthesis. I am grateful to Dr. Tsunehiro Takano for his patience, and his willingness to share his complete knowledge of all things crystallographic. My lab partners, Horace Drew and Mary Kopka deserve credit for their help in the realities of experimental science. Finally, I would like to thank Steve Schichman and Perry Norton for their friendship during my years at Caltech.

My graduate program was made possible through the financial support of the National Institute of Health and the California Institute of Technology.

Abstract

Iodo-CCGG was synthesized by the triester method and crystallized in space group $P4_32_12$. Electron density maps were phased using the anomalous scattering from the iodines and refined with the Jack-Levitt constrained refinement program. The final crystallographic R factor was 16.5% for 2 σ data and 19.9% for all data.

Solution of the structure revealed the first fragment of A-DNA to be seen in a single crystal. The helix has a 2.2Å rise per residue, 10.6 base pairs per turn, and an average propellor twist of 17°. A total of 86 water molecules were located in the crystal. Of these, 47 were within 3.5Å of the ^{I}CCGG molecule. A second layer of hydration is formed by water molecules which are within 3.5Å of two or more first layer water molecules. The hydration sites on ^{I}CCGG are ordered based on the relative strengths of DNA to water bonds, as judged by their bond lengths.

^{I}CCGG packs together with the bottom of one helix in the minor groove of another. The packing scheme indicates that the minor groove of A-DNA is a relatively hydrophobic site.This fact may be important in the recognition of A-DNA by proteins and drugs.

A list of refined coordinates of ^{I}CCGG is given in the appendix.

Table of Contents

Chapter 1. Introduction

CHAPTER 1. INTRODUCTION

1.1 Historical Background

When this project began, crystallographic structural information concerning DNA came from the two extreme limits of the diffraction method. The gross structure of nucleic acids, the double helix with anti-parallel strands, was known from the earliest fiber diffraction studies.[1] [2] The fine structure of single bases was known from high resolution single crystal studies.[3]

Fibers of DNA from natural and artificial sources had been studied under various conditions of humidity and salt content. These studies resulted in the classification of DNA fibers into the A,[4] B,[5] C[6] and D[7] [8] families. The limit of this fiber diffraction work was reached with least-squares linked-atom refinement of the A and B structures.[9] However, fiber diffraction studies were not completely satisfactory because the molecules within these fibers were neither rotationally nor translationally aligned. This inherent disorder meant that fine details of the structure, including any structural effects arising from base sequence, were averaged out.

The lack of specificity in these fiber diffraction photos became apparent with the disagreement over the structure of D-DNA. The initial D DNA structure[10] constituted what was later termed[11] "a bizarre left-handed double-helix with unusual furanose

1. J. D. Watson and F. H. C. Crick (1953). *Nature (London)* **171**, 737-738
2. R. E. Franklin and R. G. Gosling (1953). *Nature (London)* **171**, 740-741
3. J. Donohue (1968). *Arch. Biochem. Biophys.* **128**, 591-594
4. R. Langridge, W. E. Seeds, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins and L. D. Hamilton (1957). *J. Biophys. Biochem. Cytol.* **3**, 767-778
5. R. Langridge, D. A. Marvin, W. E. Seeds, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins and L. D. Hamilton (1960). *J. Mol. Biol.* **2**, 38-64
6. D. A. Marvin, M. Spencer, M. H. F. Wilkins and L. D. Hamilton (1961). *J. Mol. Biol.* **3**, 547-565
7. Y. Mitsui, R. Langridge, B. E. Shortle, C. R. Cantor, R. C. Grant, M. Kodama and R. D. Wells (1970). *Nature (London)* **228**, 1166-1169
8. S. Arnott, R. Chandrasekaran, D. W. L. Hukins, P. J. C. Smith and L. Watts (1974). *J. Mol. Biol.* **88**, 523-533
9. S. Arnott and D. W. L. Hukins (1972). *Biochem. Biophys. Res. Com.* **47**, 1504-1509
10. Y. Mitsui *et al.* (1970). ibid

ring shapes", while the alternative structure[12] was praised as "a conformationally unexceptional right-handed double-helix with standard furanose rings". Left-handed structures eventually were proposed as alternatives for both A and B families of DNA.[13] [14] These cracks in the supposedly infallible crystallographic method of structure determination highlighted the need for further crystallographic studies on DNA crystals, which had no fiber disorder. What were needed were single-crystal studies of DNA oligomers.

Considerable single-crystal x-ray diffraction work had already been done with nucleosides and nucleotides of RNA as well as with nucleotides that paired to form single steps of a double helix.[15] This body of work provided a detailed look not only at the geometry of the nucleic acid components, but also at alternate base-pairing mechanisms,[16] water structure near the DNA,[17] and interactions of both intercalating[18] and non-intercalating[19] drugs with RNA. These structures showed, in detail, what could be expected from small fragments of RNA.

However, only one DNA dimer, pTpT, had been studied[20] when the work described in this thesis began, and it was not double helical. Furthermore, studies of short pieces of RNA were not a sufficient base for extrapolation to longer pieces of DNA. A frequently voiced criticism of studies with nucleic acid dimers was that the crystallographer was only "seeing two end effects" in his crystals. The missing information

---

11. S. Arnott *et al.* (1974).
12. S. Arnott *et al.* (1974).
13. G. Gupta, M. Bansal, V. Sasisekharan (1980). *Proc. Nat. Acad. Sci. USA* **77**, 6486-6490
14. V. Sasisekharan, M. Bansal and G. Gupta (1981). *Biochem. Biophys. Res. Com.* **102**, 1087-1095
15. J. Rosenberg, N. C. Seeman, R. O. Day, and A. Rich (1976). *J. Mol. Biol.* **104**, 109-144
16. K. Hoogsteen (1959). *Acta Cryst.* **12**, 822-823
17. B. Hingerty, E. Subramanian, S. D. Stellman, T. Sato, S. B. Broyde and R. Langridge (1976). *Acta Cryst.* **B 32**, 2998-3013
18. H. M. Sobell, C. Tsai, S. C. Jain and S. G. Gilbert (1977). *J. Mol. Biol.* **114**, 333-365
19. D. M. L. Goodgame, I. Jeeves, F. L. Phillips and A. C. Skapski (1975). *Biochem. Biophys. Acta* **378**, 153-157
20. N. Camerman, J. K. Fawcett and A. Camerman (1976). *J. Mol. Biol.* **107**, 601-621

would be found in structures of intermediate chain length. These would have sufficient resolution to see the details of DNA structure and yet would also be long enough to show some of the structural constraints on DNA of biological size.

1.2 Definition of the Problem

As a first step in this direction, work was begun on a four-base *segment*, a tetramer, of DNA. This length was chosen because it was thought that a smaller piece would not provide enough information, and that a larger piece would be harder to crystallize. The latter point subsequently proved to be incorrect. In fact, the tetramer length is so short that, in terms of packing within the crystal, its behavior is as much that of a sphere as that of a long cylinder. In at least one case, the tetramer proved to be disordered[21] while a hexamer of the same sequence crystallized well.[22]

The sequence CpCpGpG was chosen for several reasons. The sequence is self complementary so that only one strand need be synthesized but in solution two molecules will pair to form a double helix. The sequence is CG rich, so that even for this short piece the melting temperature, T_m, of the double helix is not too low. (T_m has been measured by NMR as 42°C at 20 mM [CCGG], in 0.1 M phosphate, 0.01 M ethylenediamine-tetracetic acid (EDTA) and neutral pH.)[23] CCGG was also known to be the target site for the restriction endonuclease *Hpa* II from *Hemophilus parainfluenza* and its associated methylase. For *Hpa* II, methylation takes place on the 5 position of the ring of the second cytosine [24] and cutting takes place on the 3' side of the first cytosine.[25]

An important reason for choosing this particular sequence of self complementary

---

21. J. L. Crawford, F. J. Kolpack, A. H.-J. Wang, G. J. Quigley, J. H. van Boom, G. van der Marel and A. Rich (1980). *Proc. Nat. Acad. Sci. USA* **77**, 4016-4020
22. A. H.-J. Wang, G. J. Quigley, F. J. Kolpack, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich (1979). *Nature (London)* **282**, 680-686
23. D. J. Patel (1977). *Biopolymers* **16**, 1635-1656
24. M. B. Mann and H. O. Smith (1977). *Nuc. Acid Res.* **4**, 4211-4221
25. D. E. Garfin and H. M. Goodman (1974). *Biochem. Biophys. Res. Com.* **59**, 108-116

CG-rich DNA was that it contained the GpG sequence. This sequence was of interest because it was known to be the most active site for binding of the most commonly used inorganic cancer chemotheraputic agent cis-dichlorodiaminoplatinum(II) (cisplatin).[26] Since an alternating sequence, CGCG, was already being used in the Dickerson group on a separate project, it could also be used in the study of cisplatin. One part of the project was to be a comparison of cisplatin + CCGG and cisplatin + CGCG to study the binding mechanism of this drug.

Along with the cisplatin + DNA experiments we also planned to try crystallizing the sequence without any drugs. Not only would the structure of CCGG itself be interesting in its own right for reasons detailed previously, but it would also be useful for comparisons with the cisplatin studies.

1.3 Results

As often happens, the aspect of the experiment which was planned to be the major part turned out to be the minor part, and vice versa. Co-crystallization of DNA with cisplatin proved to be very difficult. Although a very large number of conditions were tried, suitable crystals were never obtained. On the other hand, the native crystallizations were successful and provided the first detailed look at a member of the A-DNA family. This single-crystal ^{I}CCGG structure gives a more detailed picture of A-DNA than the fiber structures, and points out sequence specific changes in the helix geometry. It also allows us to see the water structure around the DNA and to compare it with the recent study of water structure in B-DNA.[27] Subsequent single-crystal x-ray analyses of GGCCGGCC[28] and GGTATACC[29] have allowed us to study A-DNA structure as a function of both length and sequence. This body of

26. P. J. Stone, A. D. Kelman, F. M. Sinex, M. M. Bhargava and H. O. Halvorson (1976). *J. Mol. Biol.* **104**, 793-801

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

28. A. H.-J. Wang, S. Fujii, J. van Boom and A. Rich (1982). submitted - *Proc. Nat. Acad. Sci. USA*

29. Z. Shakked, D. Rabinovich, W. B. T. Cruse, E. Egert, O. Kennard, G. Sala, S. A. Salisbury and M. A. Viswamitra (1981). *Proc. R. Soc. Lond.* **B 213**, 479-487

information represents a large advance in our understanding of DNA structure.

Chapter 2. Synthesis of the Tetranucleotides

CHAPTER 2. SYNTHESIS OF THE TETRANUCLEOTIDES

2.1 General Considerations

Tetranucleotides were not commercially available when this project was begun, so compounds for study had to either be isolated from biological sources or synthesized by biochemical or organic means. Isolation of a given sequence from biological sources in sufficient quantity for x-ray crystallization involves enormous problems in purification.[30] For a piece of DNA as small as a tetramer, these problems are probably insurmountable. By manipulating specific DNA polymerases, biochemical synthesis had in the past provided homopolymers, alternating sequences, and occasionally even repeating triplet sequences[31] of DNA suitable for x-ray fiber work. However, the choice of sequence is very limited, and does not include CCGG. Therefore, the remaining method was organic synthesis. The native tetramer, CCGG, was generously provided by Dr. Keiichi Itakura's laboratory at the City of Hope Medical Research Center.

In order to solve the crystallographic problems, one or more heavy-atom derivatives of the native compound are required. One method, that of isomorphous replacement, involves soaking a heavy metal into the crystal and using its position and the resulting change in intensity as a basis for solving the structure. This approach had been tried with CGCG by a fellow graduate student, Horace Drew, without success. He found that soaking heavy metals into a polyanion of regular repeating structure resulted in multiple substitution, and uninterpretable results. Another possibility is to modify the DNA after synthesis but before crystallization. Since DNA has so many chemically reactive groups, I felt that this approach would lead to modification of varying degree at multiple sites. For example, it would be very difficult to prevent modification of both cytosines in CCGG instead of just one.

30. O. B. Kallai, J. M. Rosenberg, M. L. Kopka, T. Takano, R. E. Dickerson, J. Kan and A. D. Riggs (1980). *Bioch. Bioph. Acta* **601**, 113-124

31. E. Selsing, S. Arnott and R. L. Ratliff (1975). *J. Mol. Biol.* **98**, 243-248

Such multiple modification increases the crystallographic problems tremendously. Therefore, it seemed best to synthesize the required derivatives starting from commercially available modified nucleosides, following the procedures of the triester method of synthesis.

2.2 Choice of Heavy Atom Derivative

For the sequence CCGG, the two heavy atoms, iodine and bromine, were chosen for both synthetic and crystallographic reasons. These two halogens were available as 5-iodo-2'-deoxycytidine from P. L. Biochemicals and 5-bromo-2'-deoxycytidine from Sigma Chemical Company.

Bromine was chosen because its atomic number of 35 makes it a heavy but acceptable choice for the isomorphous phasing technique. In order for the isomorphous phasing method to be effective the change in diffraction intensity must be significantly above errors in measurement.[32] The change in average intensity of a CCGG tetramer because of a single bromine atom can be estimated[33] as 130%, far above the required minimum.

Iodine was chosen for use in heavy-atom phasing. As a rule of thumb for heavy-atom derivatives,[34]

$$\frac{\sum_{\text{cell}} Z_{\text{light}}^{\,2}}{\sum_{\text{cell}} Z_{\text{heavy}}^{\,2}} \approx 1.0 \qquad (2.1)$$

where Z_{light} and Z_{heavy} refer to the atomic numbers of all light and heavy atoms in the cell. For iodine attached to a CCGG tetramer, this ratio is 0.66. Thus iodine's 53 electrons make it a light but acceptable choice for heavy-atom phasing. Four different tetramers, ^{I}CCGG, C^{I}CGG, ^{Br}CCGG, and C^{Br}CGG, were synthesized by the

32. T. L. Blundell and L. N. Johnson (1976). Protein Crystallography, Academic Press, New York, p161

33. F. H. C. Crick and B. S. Magdoff (1956). *Acta Cryst.* **9**, 901-908

34. G. H. Stout and L. H. Jensen (1968). X-ray Structure Determination, The Macmillan Co., New York, p278

triester method, in the hope that at least two of the five sequences (including the original CCGG) would crystallize as an isomorphous pair.

2.3 Overview of the Triester Method

In the hands of an expert, the triester method of DNA synthesis allows organic synthesis of DNA segments up to around 25 base pairs in length. For smaller pieces, the yields at each coupling step are often greater than 90%. However, in order to crystallize DNA extremely high purity is required. Stringent purification at each step produces yields that are lower, but are still above 50%. The procedure is relatively rapid, and bases can usually be added at a rate of one per day. As originally proposed,[35] [36] the triester method was used only to make poly dT DNA. However, development since then, including that done by Dr. Itakura in the laboratory of S. A. Narang,[37] allows any sequence to be synthesized.

An overview of the triester synthesis is shown in Figure 1. Basically, the method is the following:[38]

1. Block all reactive groups on the DNA monomers
2. Deblock only the reactive sites (*ie.* the 5' end of one sugar and the 3' end of another)
3. Couple the free 3' end to the free 5' end
4. Purify the reaction mixture with silica gel chromatography
5. Start again with step 2 and repeat cycle until desired chain length is reached
6. After desired chain length is reached, remove all blocking groups from the DNA

---

35. A. M. Michelson and A. R. Todd (1955). *J. Chem. Soc.* 2632-2638
36. R. L. Letsinger and K. K. Ogilvie (1967). *J. Am. Chem. Soc.* **89**, 4801-4803
37. K. Itakura, N. Katagiri, C. P. Ball, R. H. Wroghtman and S. A. Narang (1975). *J. Am. Chem. Soc.* **97**, 7327-7332
38. K. Itakura and A. D. Riggs (1970). *Science* **209**, 1401-1405

7. Purify the deblocked DNA by chromatography.

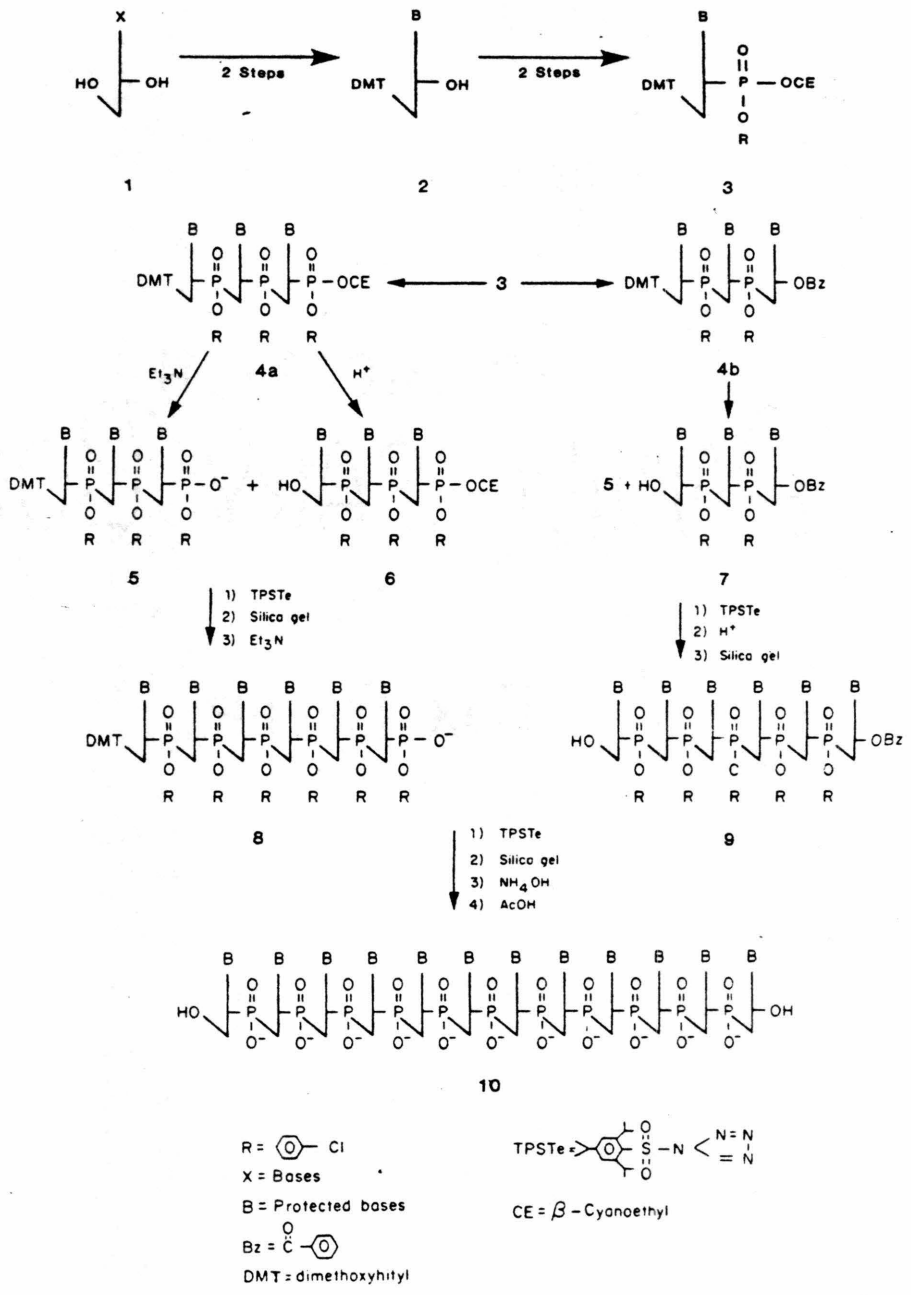

Figure 1. Overview of DNA synthesis by triester method[9]

2.4 Materials Supplied by Collaborators

The syntheses were run in collaboration with Dr. Itakura's laboratory at the City of Hope and under the guidance of Dr. Shoji Tanaka from that lab. They supplied not only the necessary experience but also some of the materials used. In particular, the fully protected monomers G, Gp and Cp were generously donated. Horace Drew was kind enough to donate fully protected 5-bromo-2'-deoxycytidine.

For the CCGG derivatives, the only nucleotides involved are guanine, cytosine and the halogenated cytosines. The O6 site on guanine and the O2 site on cytosine are not reactive enough under subsequent conditions to require protection. However, the base amino groups, N2 for guanine and N4 for cytosine, must be blocked. The N2 of guanine is protected by condensation with isobutyric acid. The N4 of cytosine is blocked by condensation with benzoic anhydride. (see section 2.5.1)

The sugar phosphate backbone contains several different blocking groups because the coupling sequence requires that a protecting group on a specific reactive site of the DNA be removed or replaced without affecting the other protecting groups. The 5' hydroxy of the deoxyribose is blocked with a 4,4'-dimethoxytrityl group (DMT in Figure 1). For all residues except the 3' terminal residue, the 3' hydroxy of the deoxyribose is phosphorylated in preparation for the coupling reaction. As the name of the synthesis suggests, the phosphate is protected as a triester. It is esterified to 3' hydroxy of the sugar, to a p-chlorophenol, and to a β-cyanoethyl group. If the deoxyribose 3' hydroxy is not to be phosphorylated, it is blocked by either anisole or benzoyl. Anisole was used for the two brominated tetramers. However, when the iodinated tetramers were made, benzoyl was recommended because of problems encountered by Dr. Itakura's group in deblocking anisole derivatives. All of the protecting groups are left in place throughout the synthesis with the exception of the β-cyanoethyl on the phosphate and the 4,4'-dimethoxytrityl on the 5' hydroxy of the deoxyribose. These two reactions are

discussed in greater detail in the section 2.5. The fully protected forms of these nucleotides received from Dr. Itakura's lab are shown in the following figure.

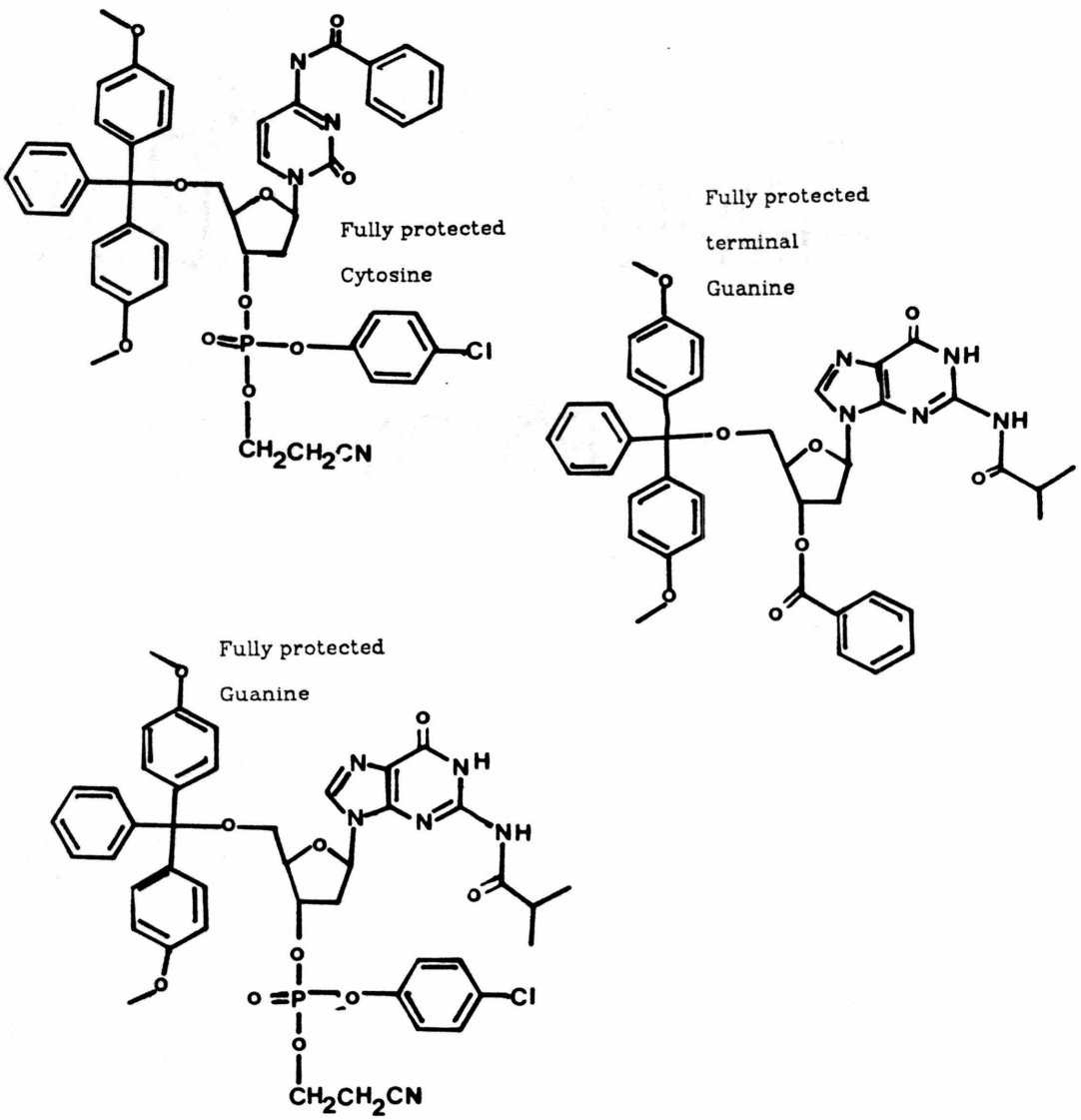

Figure 2. Fully protected groups not synthesized

2.5 Synthetic Reactions Performed

The reactions I performed dealt with the protection and phosphorylation of 5-iodo-2'-deoxycytidine and the actual chemical linking of the monomers together. This latter part involved synthesis of the coupling agent, removal of protection groups in preparation for coupling and the coupling reaction itself.

2.5.1 Protection of ^{I}C: For cytosine and its iodine and bromine derivatives, the amine is the only reactive group on the base itself. It is protected by refluxing 5-iodo-2'-deoxycytidine with a 2 molar excess of benzoic anhydride in dry ethanol for 3 hours while constantly stirring.[39] Progress of the reaction was checked hourly by chromatography on silica gel 60 F254 plates run in 5/1 $CHCl_3/CH_3OH$. Disappearance of the low mobility iodocytidine marked the end of the reaction. After the reaction was completed, the mixture was put in an ice bath and the product precipitated in the form of light yellow crystals. These were filtered out and washed with diethyl ether.

A sample of this N-benzoylated 5-iodo-2'-deoxycytidine was submitted to the Caltech analytical lab for elemental analysis of H, C, N, and Iodine. The reported value of the iodine was within 2% of the calculated value, indicating that the iodine was not destroyed during the protection reaction.

The 5' hydroxy of the deoxyribose was protected by condensation with a 50% molar excess of 4,4'-dimethoxytrityl chloride in 5 ml dry pyridine/ mM N-benzoyl cytidine. Progress of the reaction was checked every 30 minutes by chromatography on silica gel 60 F-254 plates in 9/1 $CHCl_3/CH_3OH$. After chromatography, these plates were sprayed with perchloric acid and heated on a hot plate. Under these conditions the dimethoxytrityl group turns bright orange and the deoxyribose becomes dark brown. Thus the tritylated sugar can be identified as an orange spot

39. Synthetic Procedures in Nucleic Acid Chemistry Vol. I pg 285 ***et seq.*** Zorbach and Tipson editors (1968) Interscience Publishers.

which turned brown upon heating. The reaction was stopped with 5 ml/mM dimethoxytrityl of $NaHCO_3$. and the product extracted in $CHCl_3$. The $CHCl_3$ was washed twice with water and evaporated to dryness. Purification was done on a silica gel column of the type described in section 2.6.

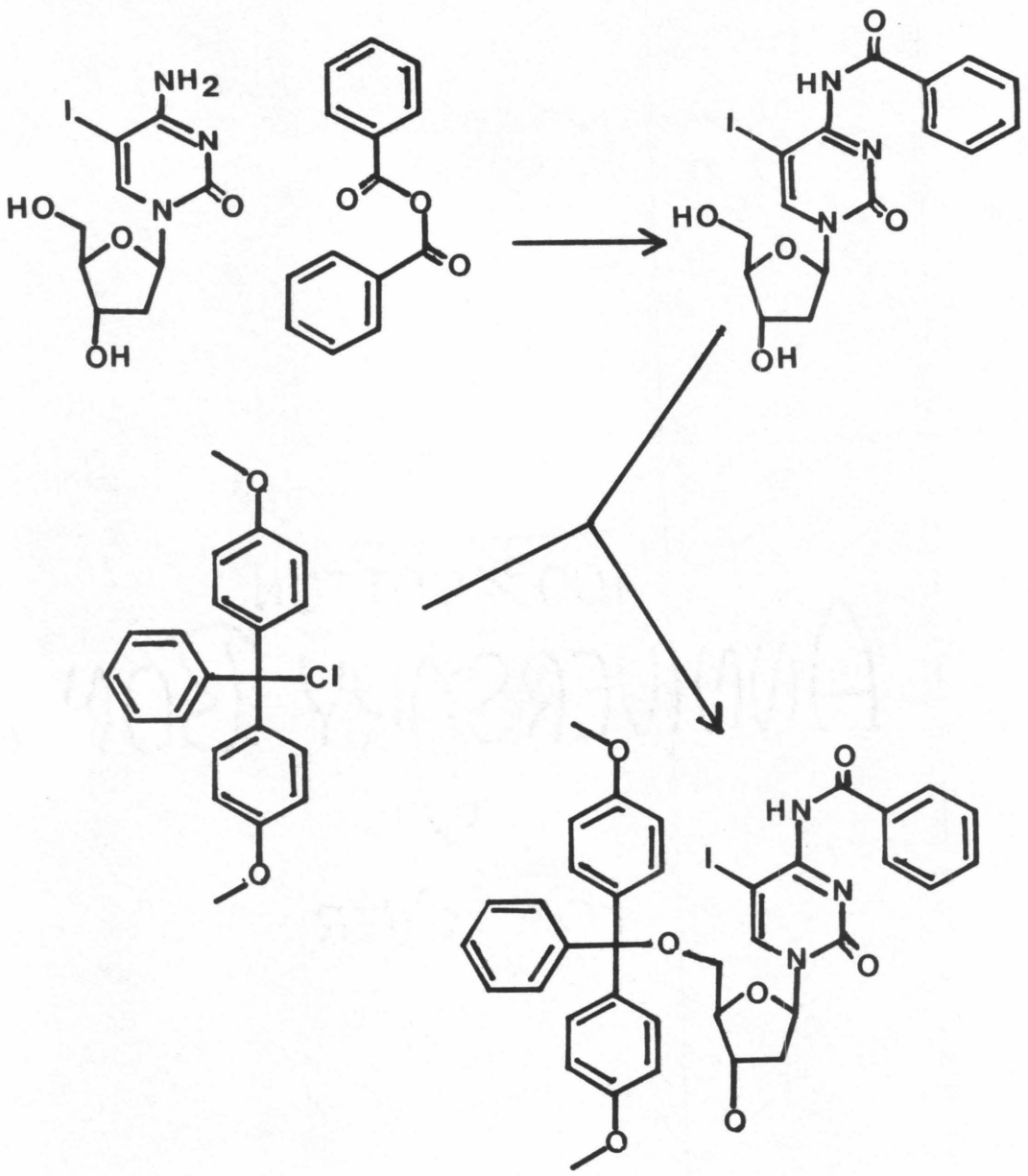

Figure 3. Protection of ^{I}C

2.5.2 Phosphorylation of ^{I}C: The iodocytidine is now protected and ready for phosphorylation of the 3' hydroxy end. This reaction is run in two steps as shown in Figure 4. First, to prepare the activated reagent, the phosphorylating reagent (50% molar excess) is added to 1,2,4-triazole (twofold molar excess) in dry pyridine (5ml/mM phosphorylating reagent). Triethylamine (3.3 molar excess) in dry dioxane (twofold molar excess over triethylamine) is added dropwise to the mixture while it is stirring in an ice bath. After addition, the mixture is stirred 15 minutes in its ice bath and then left at room temperature for 1 hour. The sticky white triethylamine salt is filtered out with a glass frit funnel and the liquid reduced to a volume of 6 ml. This reduced volume is added to the protected monomer and allowed to react at room temperature for 4 hours. Every hour, aliquots of the reaction mixture are treated with $(CH_3CH_2)_3N$/pyridine/water and chromatographed on silica gel plates in 15/1 $CHCl_3/CH_3OH$. In this system the charged product shows low mobility with respect to the uncharged starting material. The reaction is stopped by addition of 1 M triethylamine bicarbonate pH=7.5, evaporated to a small volume and the product extracted with chloroform.

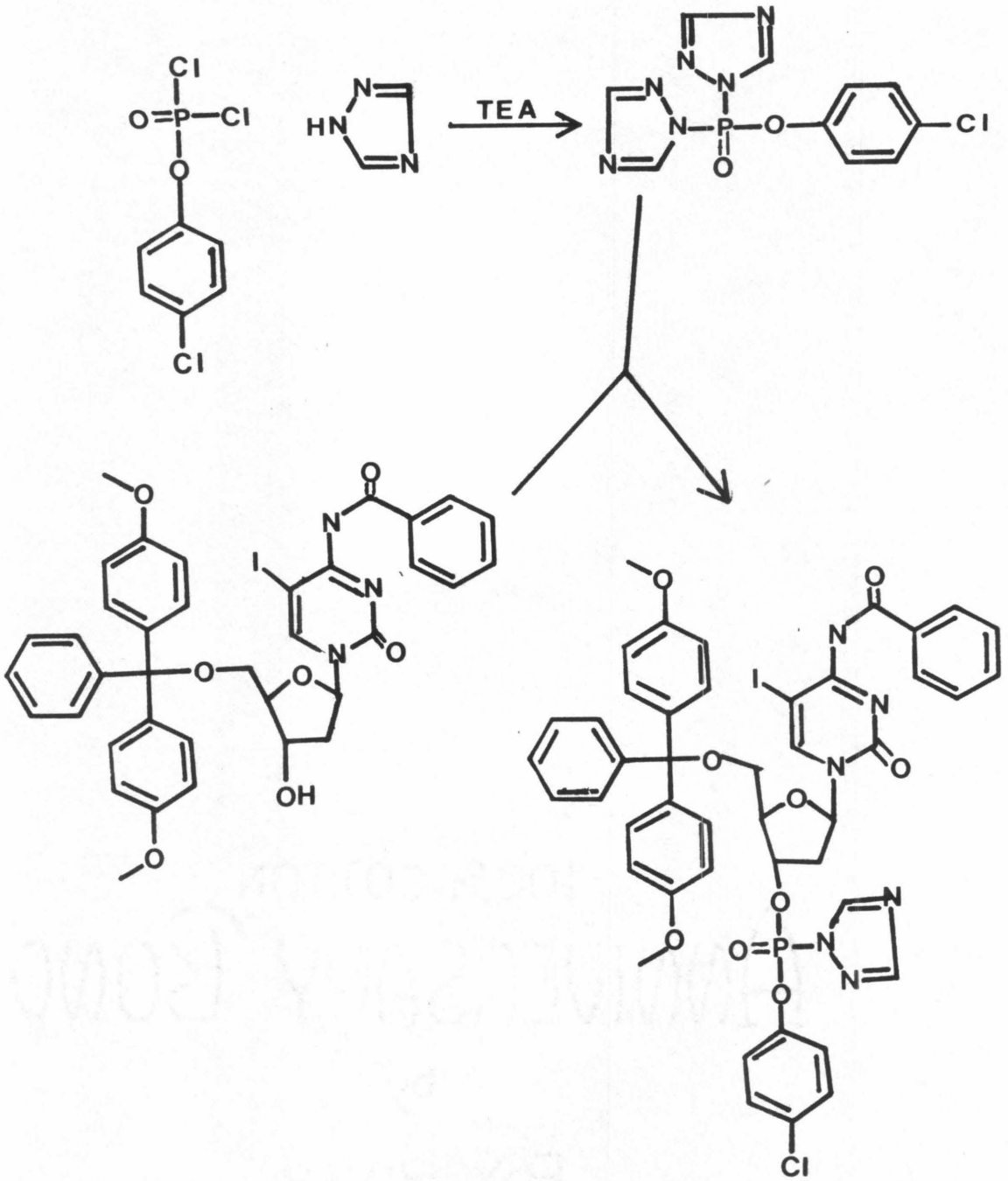

Figure 4. Phosphorylation of ^{I}C

2.5.3 Preparation of Coupling Reagent: The actual coupling of protected monomers is performed by TPS-tetrazole. The synthesis of this reagent is shown in Figure 5. Triisopropylphenylsulfonyl chloride, TPS-Cl, is mixed with an equimolar amount of tetrazole in a dry dioxane (1.5 ml/mM TPS-Cl). An equimolar amount of triethylamine is added dropwise while the mixture was stirring rapidly in an ice bath. After two hours, the triethylamine chloride salt is filtered out and the reaction mixture reduced to a small volume. The product is crystallized overnight at 4°C from benzene/n-pentane.

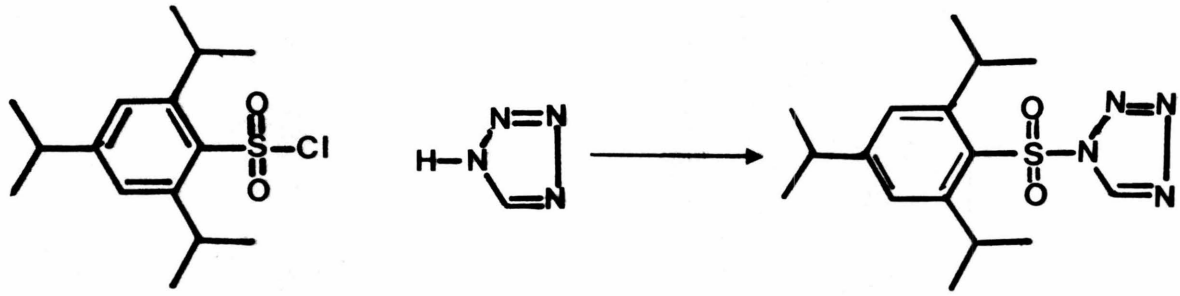

Figure 5. TPS-tetrazole synthesis

2.5.4 Coupling Reaction: As shown in Figure 1., the actual linking of two residues, X-p and Y, together to form the 3' to 5' phosphate linkage X-p-Y is done in four steps:

1. Remove β-cyanoethyl protecting group from the phosphate group of X-p. This is done by treatment of the monomer with 3/1/1 pyridine/water/$(CH_3CH_2)_3N$ at

room temperature for 20 minutes.

2. Remove dimethoxytrityl group from 5' hydroxy of deoxyribose of Y. This is done in 2% benzenesulfonic acid for 20 minutes at 0°C.

3. Add a 50% molar excess of X to Y and dry thoroughly. In a small volume of pyridine add the two monomers and the coupling reagent, triisopropylsulfonyl tetrazole. The reaction runs about 1 hour at room temperature and is checked every 30 minutes by chromatography on a silica gel plate in 9/1 $CHCl_3/CH_3OH$. The reaction is stopped by addition of water and the mixture evaporated to dryness. The final traces of pyridine are removed by evaporation of toluene and the product is purified by chromatography on a silica gel column as described in section 2.6.

2.6 Purification During Synthesis

Because of the high purity required for crystallization, extra care must be taken during the synthesis to purify intermediate products. This is usually done by chromatography on a silica gel column. A typical column is 10 cm. in diameter and consists of 5.5 cm. bed of Silica Gel 60 H (MCB Manufacturing Chemists Inc.) in chloroform. The material to be purified is loaded on the column in chloroform and then chromatographed with a 0% to 5% gradient of methanol in chloroform. Fractions are collected and spotted on Silica Gel F-254 thin layer plates from E. M. Laboratories Inc. Good fractions are combined and concentrated. Careful chromatography during synthesis produces much cleaner reactions later and a purer final product. This purification is a critical step for DNA intended for crystallization.

2.7 Tests of 5-iodo-2'-deoxycytidine Stability

Three experiments were performed in order to check the stability of iodo-deoxycytidine under deblocking conditions.

1. 10% isopropylamine in methanol for 23 hours at room temperature

2. 2/1 conc. NH_4OH/pyridine for 12 hours at room temperature

3. 2/1 conc. NH_4OH/pyridine for 4 hours at 50°C

When judged by chromatography on thin layer plates, all reactions showed deblocking with no major side products. I used the last method for deblocking the tetramers because it was the quickest and most commonly used.

2.8 Deblocking

After synthesis is completed, the compounds must have all of their blocking agents removed. The blocked DNA is dissolved in a small amount of pyridine and added to 80 ml of 2/1 conc. NH_4OH/pyridine in a round bottom flask. The flask is then sealed with a glass stopper, wrapped with tape for protection is case of explosion, and heated to 50°C in a water bath for 12 hours. This procedure removes all base-labile protecting groups, *ie.* all protecting groups except the dimethoxytrityl on the 5' end of the DNA. After cooling to room temperature and evaporating the mixture to dryness, 100 ml of galcial acetic acid are added for 30 minutes at room temperature. The short reaction time and moderate temperature prevent depurination of the DNA. The mixture is evaporated to dryness again, the blocking agents are extracted with ether and the solution then is made ready for chromatography on DEAE cellulose.

The sequence of deblocking given here, base before acid, is important in order to prevent the amination reaction shown in Figure 6. Therefore, the dimethoxytrityl group is left on until after the basic deblocking step is completed.

2.9 Purification After Deblocking

After deblocking, each tetramer had to be purified to remove blocking agents, partially blocked tetramers, unreacted materials and products of side reactions. Two separate portions of the native compound, CCGG, were received from Dr.

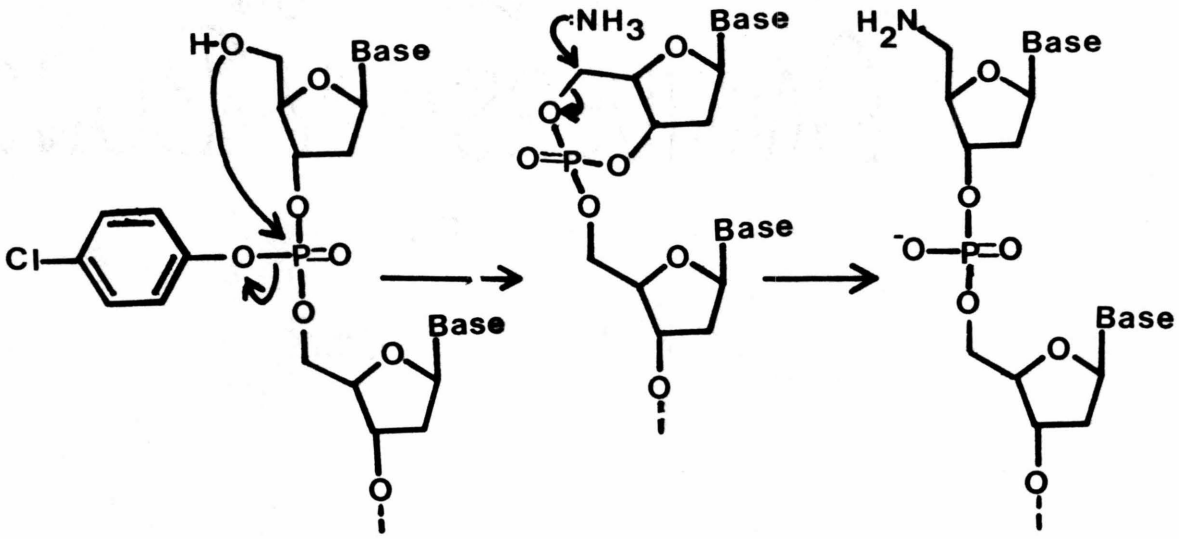

Figure 6. Amination side reaction of deblocking step.

Itakura's laboratory. Purification was done on DEAE cellulose in a triethylammonium bicarbonate pH=7.5, 15% ethanol buffer system. A gradient from 0.1M to 0.7M TEAB was used to chromatograph the tetramer. Of these two preparations of the same compound, only one would crystallize, although they had identical UV absorption and appeared identical by high performance liquid chromatography, HPLC. This situation prompted an increase in concern about purity so a new method of purification was adopted for the halogenated tetramers.

The new method of DNA purification was one used for longer pieces of DNA.[40] It involved chromatography on DEAE cellulose in 7 M urea, 20 mM tris pH=7.5 with a NaCl gradient from 30 mM to 60 mM. DNA from the desired peak was trapped on another DEAE column, washed free of urea, and taken off with TEAB buffer with 15% ethanol. This chromatography system was far from ideal and usually resulted in the

40. E. Ohtsuka, A. Kumar and H. G. Khorana (1972). *J.Mol. Biol.* **72**, 309-327

product peak appearing as a relatively broad peak close to impurities.

A decision about which fractions to pool was made on the basis of HPLC results from selected fractions. This allowed separation of impurities which could not be differentiated by their UV absorbance. Samples were run on a Spectra Physics HPLC equipped with AAX ion exchange column and using a gradient from 0.005 M to 0.05M KH_2PO_4, pH 4.4 as a buffer system. This system served very well in distinguishing compounds which could not be cleanly separated by DEAE chromatography. The final products of these purification efforts were always single peaks with less than 1% contamination as judged by HPLC.

All four of halogenated tetramers were purified in this manner and three of them were subsequently crystallized. The best that can be said for this expensive, time consuming and messy purification procedure is that it had the desired result. This fact is mainly due to the ability of the HPLC to discriminate between product and impurity cleanly when the DEAE column could not.

2.10 Enzymatic Digestions

A final test of chemical composition was performed using snake venom phosphodiesterase which cleaves the tetramers on the 3' side of the phosphate. The monomers were then chromatographed on thin layer plates of Cellulose F. In order to prevent distortion of chromatography by the phosphodiesterase, the material was applied to a cellulose F plate and chromatographed in water for a short distance. The migrating monomers were scraped off the plate, eluted into water, concentrated by lyophilization, applied to another cellulose F plate and chromatographed against standards in 6/3/1 iospropanol/water/conc. NH_4OH This procedure eliminated the distortion of the faster migrating cytosine residues. After chromatography, the separated components were scraped off the plate, eluted with water and their UV spectra taken. The spectra identified the compound and served to determine the

relative amounts of each monomer present. From these spectra the compositions of the tetramers were confirmed. In the case of ^{I}CCGG, this technique also verified the fact that the iodine was still intact after purification, at a time when it could not be located crystallographically.

Chapter 3. Anomalous Phasing

CHAPTER 3. ANOMALOUS PHASING

3.1 The Path from CCGG to ^{I}CCGG

The project began as a study of the compound CpCpGpG. To solve the crystallographic phasing problem, we intended to use the technique of isomorphous replacement. This involves either making a heavy atom derivative of the molecule while it is in the crystal by soaking in heavy ions, or modifying the molecule before crystallization and then crystallizing it in the same space group with the same unit cell packing as the unmodified form. With this goal in mind, the two bromine derivatives described in Chapter 2 were synthesized and crystallization trials were begun. Only one of these compounds, ^{Br}CCGG, could be crystallized, and it was useless as an isomorphous derivative because it was not in the same space group as the native tetramer.

Since ^{Br}CCGG could not be used as an isomorphous derivative, could the crystal structure be solved by itself, using bromine as a heavy atom? The rule of thumb for heavy atom derivatives expressed in equation (2.1) gives a value of 0.29 for bromine on the CCGG tetramer. For a good derivative this ratio should be $\approx$ 1.0. Therefore, bromine is not heavy enough by itself to phase these crystals.

At this point, the phasing problem was reevaluated and two iodine derivatives of CCGG were synthesized as described in Chapter 2, for use in heavy atom phasing. Both of these iodine derivatives were crystallized successfully. One, C^{I}CGG, formed disordered crystals. The other, ^{I}CCGG, formed crystals suitable for x-ray diffraction study. A survey of all crystals grown is given in Tables 1 and 2.

Unit cell dimensions and density measurements of ^{I}CCGG indicated two molecules in the asymmetric unit. This meant that two iodines must be located before phasing could begin. The fact that this ^{I}CCGG crystal is in a different space group from both CCGG and ^{Br}CCGG meant that isomorphous phasing could still not be used. There-

DNA	Conditions	Space Group	Notes
CCGG	1.49 mM CCGG 2.94 mM $MgCl_2$ isopropanol, 2°C	$P2_1$	thin laths 1.0x0.2x0.03mm
^{Br}CCGG	2.23 mM ^{Br}CCGG 4.33 mM $MgCl_2$ isopropanol, 2°C	$P2_12_12$	thin plates 0.5x0.2x0.03mm
C^{Br}CGG	no xtal	---	can't win them all
^{I}CCGG	5.2mM ^{I}CCGG 0.4 mM spermine Cl 30 mM Na Cacodylate pH=7.5 isopropanol, 2°C	$P4_32_12$	0.5x0.5x0.3mm
C^{I}CGG	4.4 mM C^{I}CGG 7.3 mM $MgCl_2$ ethanol, 2°C	tetragonal	0.3x0.2x0.2mm blocks

TABLE 1. Summary of crystals grown of CCGG and its derivatives.

fore the heavy atom approach was begun.

Diffraction data were collected from ^{I}CCGG crystals, and Patterson maps were studied in an attempt to locate the heavy atoms. Because of the inherent symmetry of the DNA molecule, many large peaks appeared on the Patterson maps which were not caused by the iodines, making interpretation of the maps difficult. Several solutions to the Patterson maps were tried but none produced interpretable electron density maps. Harker sections of this map are shown in Figure 1.

DNA	Space Group	Unit Cell Dimensions	Asymmetric Unit
CCGG	$P2_1$	a=22.2Å b=25.6Å c=30.4Å	two single strands
^{Br}CCGG	$P2_12_12$	a=39.7Å b=31.8Å c=13.7Å	two single strands
C^{Br}CGG	no xtal	---	---
^{I}CCGG	$P4_32_12$	a=41.1Å b=41.1Å c=26.7Å	two single strands
C^{I}CGG	tetragonal	a=56.Å b=56.Å c=26.Å	four single strands

TABLE 2. Crystallographic parameters of CCGG and its derivatives.

At this point the method of anomalous dispersion was considered for phasing. It had not been tried previously because the anomalous effect is usually too small to be considered when dealing with molecules of this size containing only H, C, N, O, and P. However, investigation of the anomalous method showed that the iodine attached to CCGG had twice as much phasing power as the chlorine used to calculate phases for ephedrine hydrochloride.[41] Therefore data collection of Friedel pairs began in preparation for anomalous phasing.

3.2 Data Collection for Anomalous Phasing

Since the anomalous difference to be measured is small, great care must be taken in data collection. The Friedel pairs, $\mathbf{F}_{hkl}$ and $\mathbf{F}_{\overline{hkl}}$, must be collected within a short time of each other in order to prevent crystal decay from affecting the anomalous difference. For ^{I}CCGG, Friedel pairs were collected within 30 minutes of each other. The decay in intensity during this period of time was less than 0.1%. The Friedel pairs must also be collected in such a manner that absorption differences due to crystal orientation are negligible. Two methods suitable for our

41. G. N. Ramachandran and S. Raman (1956). *Curr. Sci.* **11**, 348-351

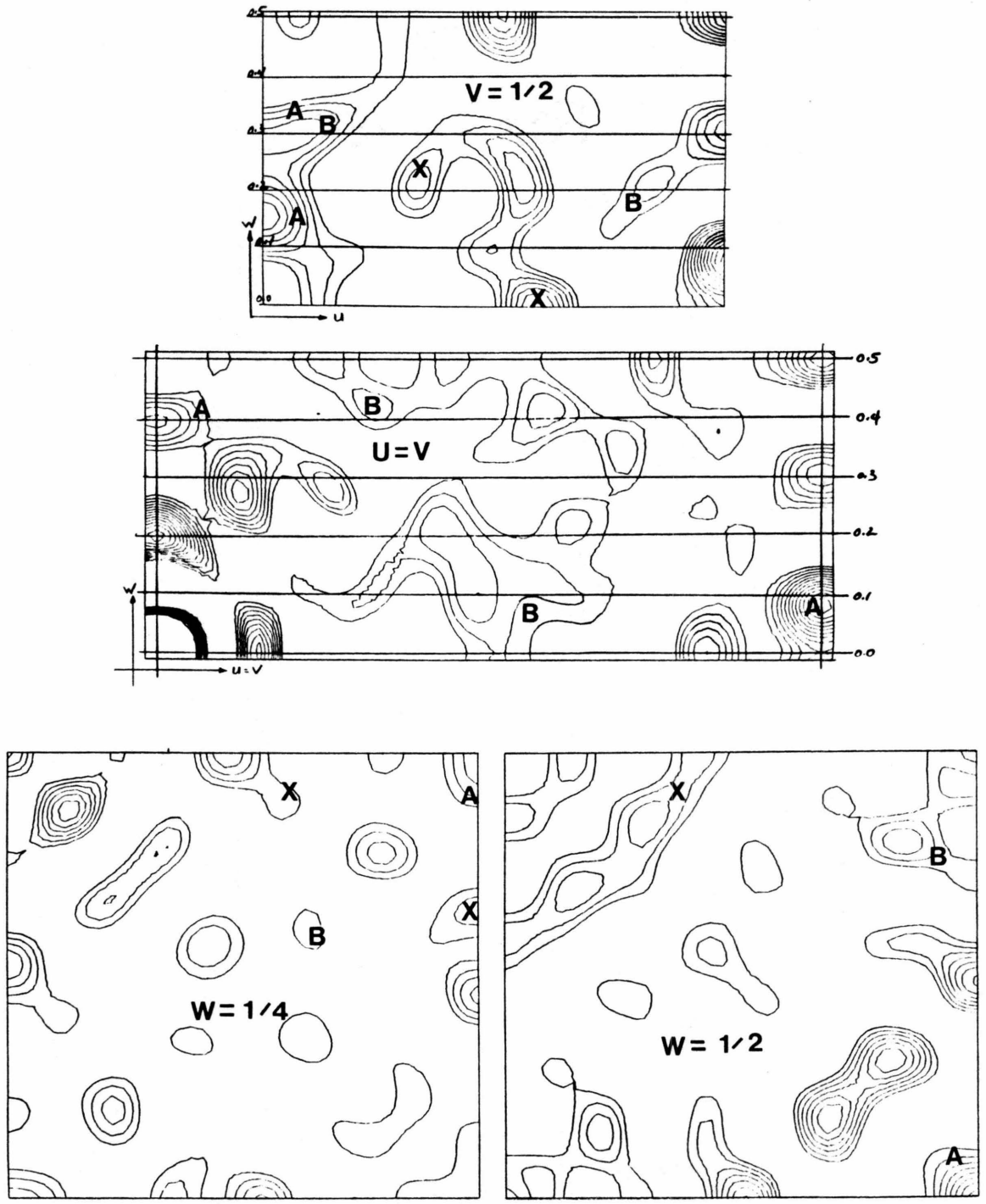

Figure 1. Harker sections from 4/mmm averaged data

four circle diffractometer were tried.

The first method is to collect I_{hkl} at 2Θ, ω, φ, and χ and $I_{\bar{h}\bar{k}\bar{l}}$ at 2Θ, ω, $\varphi+180°$, and $-\chi$. This is the method supplied in the software for our Syntex $P\bar{1}$ diffractometer. The advantage of this method is that the reflections are collected 180° apart in φ. The absorption correction[1] applied to these diffraction data assumes that x-ray absorption is equal when the crystal is at φ and φ + 180°. Experimentally, there is only a small difference so this assumption is usually good. Thus, the same absorption correction is applied to both reflections and the anomalous difference between I_{hkl} and $I_{\bar{h}\bar{k}\bar{l}}$ is not obscured by errors in absorption.

The other method was suggested by Dr. Tsunehiro Takano. ICCGG crystals were usually mounted with b^* along the goniometer head axis. Since in $P4_32_12$, $I_{hkl} = I_{\bar{h}k\bar{l}}$, one can collect I_{hkl} at 2Θ, ω, φ, and χ and $I_{h\bar{k}l}$ at 2Θ, ω, φ, and $-\chi$. The advantage of this method is that both reflections are collected at exactly the same φ value so the absorption correction is exactly the same for both reflections. The $\overline{P1}$ collects data with the l index incremented most rapidly. Therefore, Friedel pairs could be collected close together in time if the crystal were reindexed with the old b^* axis as the new c^* axis. The new indexing was done as follows.

$$h_{new} = h_{old}$$

$$k_{new} = l_{old}$$

$$l_{new} = \bar{k}_{old}$$

In order for this method to collect data efficiently, the diffractometer must be stopped and reset every time the l index reaches its limit. This method of data collection requires constant operator intervention.

---

1. A. C. T. North, D. C. Phillips and F. S. Matthews (1968). *Acta Cryst.* **A 24**, 351-359

Three angstrom data were collected by both methods on a single crystal. In order to compare the two methods, an R factor for each data set was calculated from the two measurements of each centric reflection. Both data sets had the same centric R factor of 5% in F, indicating equal quality data by both methods. Hence, the rest of the data collected (to 2.1 Å resolution) used the method of φ and φ + 180°, because it required less operator intervention.

3.3 Normal vs. Anomalous X-ray Scattering

To understand the cause of anomalous scattering of x-rays by iodine, a brief review of normal x-ray scattering is in order. In a single crystal x-ray diffraction experiment, x-rays are scattered by the electrons of the molecule of the crystal. The three-dimensional periodicity of the crystal lattice causes interference between the scattered x-rays and results in the formation of a three dimensional lattice of diffraction spots of varying intensity. From the value and position of these intensities the electron density throughout the crystal, and thus the structure of the molecule, is deduced.

In the classical derivation of x-ray scattering by an atom, the electron is considered to be weakly coupled to its nucleus to form a simple harmonic oscillator. The normal oscillation frequency of this electron, ω_o, is the frequency with which this electron oscillates when it is displaced from equilibrium. In the diffraction experiment, this atom is placed in an incident beam of x-rays whose frequency, ω, is large compared to ω_o. Under such conditions, the electron is driven by the sinusoidal electric field of the incoming x-ray. The solution to the problem of a damped simple harmonic oscillator driven by a sinusoidal force is well known.[2] An oscillator driven at frequency ω reaches a steady state in which its frequency is equal to the driving frequency, but its amplitude of oscillation and the phase difference between the driving force and the particle are both functions of ω and ω_o.

2. J. B. Marion (1970). *Classical Dynamics of Particles and Systems*, Academic Press, p118

The amplitude is given by

$$A = \frac{F}{\sqrt{(\omega_o^2 - \omega^2)^2 + 4\omega^2\beta^2}} \cos(\omega t - \delta) \qquad (3.1)$$

where

- F is the force on the oscillator
- β is the viscous damping force of the oscillator and
- δ, the phase difference between the driving force and the oscillating particle, is defined by

$$\delta = \tan^{-1}\left\{\frac{2\omega\beta}{\omega_o^2 - \omega^2}\right\} \qquad (3.2)$$

It is this variation of amplitude and phase difference with ω and ω_o that is at the heart of anomalous scattering. Both of these functions change rapidly when $\omega \approx \omega_o$ but are relatively constant otherwise. In the case of normal scattering, $\omega >> \omega_o$ and scattering occurs with the scattered wave 180° out of phase with the incident wave.[3] However, if the scattering atom happens to have an x-ray absorption edge near the frequency of the incident radiation, then $\omega \approx \omega_o$ and both A and δ undergo drastic changes. For example, δ changes from 180° when $\omega >> \omega_o$ to 90° when $\omega = \omega_o$.

It is important to remember that different atoms usually have absorption edges at different frequencies. Since the amount of anomalous scattering depends on how close ω is to ω_o, when one atom is scattering with a large anomalous component, the other atoms in the structure probably will not be. In the specific case of ^{I}CCGG, only iodine has a significant anomalous scattering component for the CuK_α whose wavelength is 1.5418Å. Therefore, the phase and amplitude of a wave scattered from an iodine would be different from those scattered from a carbon in the same position. These differences in phase and amplitude must be taken into account when

3. R. W. James (1967). *The Optical Principles of the Diffraction of X-rays* G. Bell and Sons Ltd. (London) p135

the electron density of the cell is calculated.

This correction to the electron density map is done by placing two mathematically different types of atoms at the site of the anomalous scatter. (It is important to note here that in order to make this correction the position of all of the anomalous scatters in the unit cell must be known.) One of these atoms scatters as would a normal atom, such as a carbon, but the other atom scatters 90° out of phase from a normal atom. The mathematics of this correction involves changing the scattering factor of the atom from a real number, f, to a complex number, $f + f' + i\Delta f''$, having both real and imaginary parts. Now, instead of an atom's contribution to a scattered wave being

$$f\,[\cos 2\pi(hx + ky + lz) + i\sin 2\pi(hx + ky + lz)] \tag{3.3}$$

it becomes

$$(f + \Delta f')\,[\cos 2\pi(hx + ky + lz) + i\sin 2\pi(hx + ky + lz)] \tag{3.4}$$

$$+ \Delta f''\,[-\sin 2\pi(hx + ky + lz) + i\cos 2\pi(hx + ky + lz)]$$

Now the real term has a sine component (which is 90° out of phase from its normal cosine component) and the imaginary term has a cosine component (which is 90° out of phase with its normal sine component). That is to say, at the position (x,y,z) there appear to be two atoms. One scatters as would a normal atom while the other scatters 90° out of phase from a normal atom. By varying the number of electrons on these two atoms, the scattered wave from the anomalous scatter can be reproduced and the electron density of the crystal calculated. Usually the number of electrons on the "imaginary" atom is much less than that on the other normal atom. In the case of iodine in CuK_α radiation the "imaginary" atom has 7.2 electrons while the normal has 52 electrons. Values of f, $\Delta f'$, and $\Delta f''$ for all atoms in the structure are given in Table 3.

	H	C	N	O	P	I
f	1.0	6.0	7.0	8.0	15.0	53.0
$\Delta f'$	0.0	0.0	0.0	0.0	0.2	-1.1
$\Delta f''$	0.0	0.0	0.0	0.1	0.5	7.2

TABLE 3. Corrections for anomalous scattering for all atoms in ^{I}CCGG at $\sin\Theta / \lambda = 0$ with CuK_α radiation

One of the manifestations of anomalous scattering is that Friedel's law, which says that the intensity of the hkl and $\overline{hkl}$ reflections must be equal, is no longer valid. The reason for this breakdown can be seen by examining the structure factor equation. For normal x-ray scattering,

$$\mathbf{F}_{hkl} = \sum_j f_j[\cos 2\pi(hx_j + ky_j + lz_j) + i\sin 2\pi(hx_j + ky_j + lz_j)] \tag{3.5}$$

$$= A_{hkl} + iB_{hkl}$$

where the sum runs over the j atoms in the unit cell which have positions (x_j,y_j,z_j), and scattering factor f_j. For the Friedel mate of the above reflection,

$$\mathbf{F}_{\overline{hkl}} = \sum_j f_j[\cos 2\pi(-hx_j - ky_j - lz_j) + i\sin 2\pi(-hx_j - ky_j - lz_j)] \tag{3.6}$$

$$= \sum_j f_j[\cos 2\pi(hx_j + ky_j + lz_j) - i\sin 2\pi(hx_j + ky_j + lz_j)]$$

$$= A_{hkl} - iB_{hkl}$$

Since the intensity of a given reflection, I_{hkl}, is proportional to

$$|\mathbf{F}_{hkl}|^2 = \mathbf{F}_{hkl} \bullet \mathbf{F}^*_{hkl} \tag{3.7}$$

$$= (A_{hkl} \bullet A_{hkl}) + (iB_{hkl} \bullet -iB_{hkl})$$

$$= A^2_{hkl}+B^2_{hkl}$$

the change in sign of the imaginary part of the structure factor equation cancels out. Therefore, $|\mathbf{F}_{hkl}|^2 = |\mathbf{F}_{\overline{hkl}}|^2$ and $I_{hkl} = I_{\overline{hkl}}$.

However, we have seen that in the case of anomalous scattering, the scattering factor, f_j, becomes the complex function $f_j + \Delta f_j' + i\Delta f_j''$. The imaginary component of the scattering factor adds cross terms of the type $\Delta f_j''\sin 2\pi(hx_j + ky_j + lz_j)$ to the structure factor equation. These cross terms are not the same for $|\mathbf{F}_{hkl}|^2$ and $|\mathbf{F}_{\overline{hkl}}|^2$ so Friedel's law breaks down. It is this difference in I_{hkl} and $I_{\overline{hkl}}$, the Bijvoet difference, which is used to measure the anomalous effect and provides the basis for anomalous phasing techniques.

Although this anomalous effect usually is negligible, in the case of ^{I}CCGG the average difference between $|\mathbf{F}_{hkl}|$ and $|\mathbf{F}_{\overline{hkl}}|$ was 12%. This was easily measured over a noise level of 5%.

Reflections from an anomalous dispersion data set can be divided into two groups, centric and acentric reflections. In $P4_32_12$, the structure factor equation for centric reflections can be reduced to:

$$\mathbf{F}_{hkl} = \sum_j f_j \cos 2\pi(hx_j + ky_j + lz_j) \quad \text{or} \quad \mathbf{F}_{hkl} = \sum_j f_j\, i\sin 2\pi(hx_j + ky_j + lz_j) \quad (3.8)$$

The first form contains only a cosine term which is invariant when changing hkl to $\overline{hkl}$. The second form contains only a sine term which does becomes negative when moving from hkl to $\overline{hkl}$. However, this change is not seen in the intensity because it is masked by the conversion from structure factor to intensity in the same manner as equation (3.7). Therefore, centric reflections do not show Bijvoet differences.

3.4 Summary of Anomalous Scattering Cause and Effect

The causes and effects of anomalous dispersion are summarized here.

- Occurs when the x-ray wavelength is near an absorption edge of the anomalous scatter
- Caused by shift in phase and amplitude of the anomalously scattered wave relative to that of a normally scattered wave.
- Usually a negligible effect but may be significant (*ie.* > 10%)
- Affects the intensities of only acentric, not centric reflections
- May be used in locating anomalous scattering atom in Patterson map.
- May be used to calculate the phases necessary to create an electron density map.

3.5 Location of the Anomalous Scatterers

As pointed out earlier, the location of all anomalous scatterers in the cell must be known in order to use the Bijvoet differences to phase a reflection. In the case of isomorphous replacement techniques, this is done using difference Patterson maps or by using phases calculated from one derivative to phase a second derivative's Fourier map. In the case of ^{I}CCGG however, there is only one data set so no previous phases are available. Therefore, the iodines must be located by Patterson techniques.

A Patterson map is made using the measured intensities of a crystal without the need for phase information. Peaks in this map correspond to vectors between atoms in the crystal. The heavier the peak, the greater electron density of the two atoms which make up that vector. The advantage of this map is that the phase of a given reflection need not be known. However, interpretation of such a map in terms of coordinates for individual atoms is often not straightforward. Peaks in a Patterson map are broader than those in the corresponding electron density map. For this reason the Patterson map becomes more cluttered as the number of atoms

increases, because the peaks begin to overlap one another. This overlap problem may become particularly misleading in cases where the molecule itself possesses high symmetry, as with DNA. Such symmetry increases the odds that several peaks will overlap one another and create the impression of a single peak of higher weight. For example, many small carbon-to-carbon peaks may pile up to simulate one iodine-to-iodine peak.

Once again, anomalous differences can be of some help. Rossmann has demonstrated that a Patterson summation with $(|\mathbf{F}_{hkl}| - |\mathbf{F}_{\overline{hkl}}|)^2$ coefficients gives peaks at the end of vectors relating the anomalous scatters in the structure.[4] Theoretically, the major peaks in the anomalous Patterson map should correspond to iodine to iodine vectors since the other atoms do not scatter anomalously to an appreciable extent. These anomalous Patterson maps were made for ^{I}CCGG. Most useful were the Harker sections. For $P4_32_12$, the relevant Harker sections are the sections $u = v$, $u = \frac{1}{2}$, $w = \frac{1}{4}$, and $w = \frac{1}{2}$. These sections are shown in Figure 2.

4. M. G. Rossmann (1961). *Acta Cryst.* **14**, 383-388

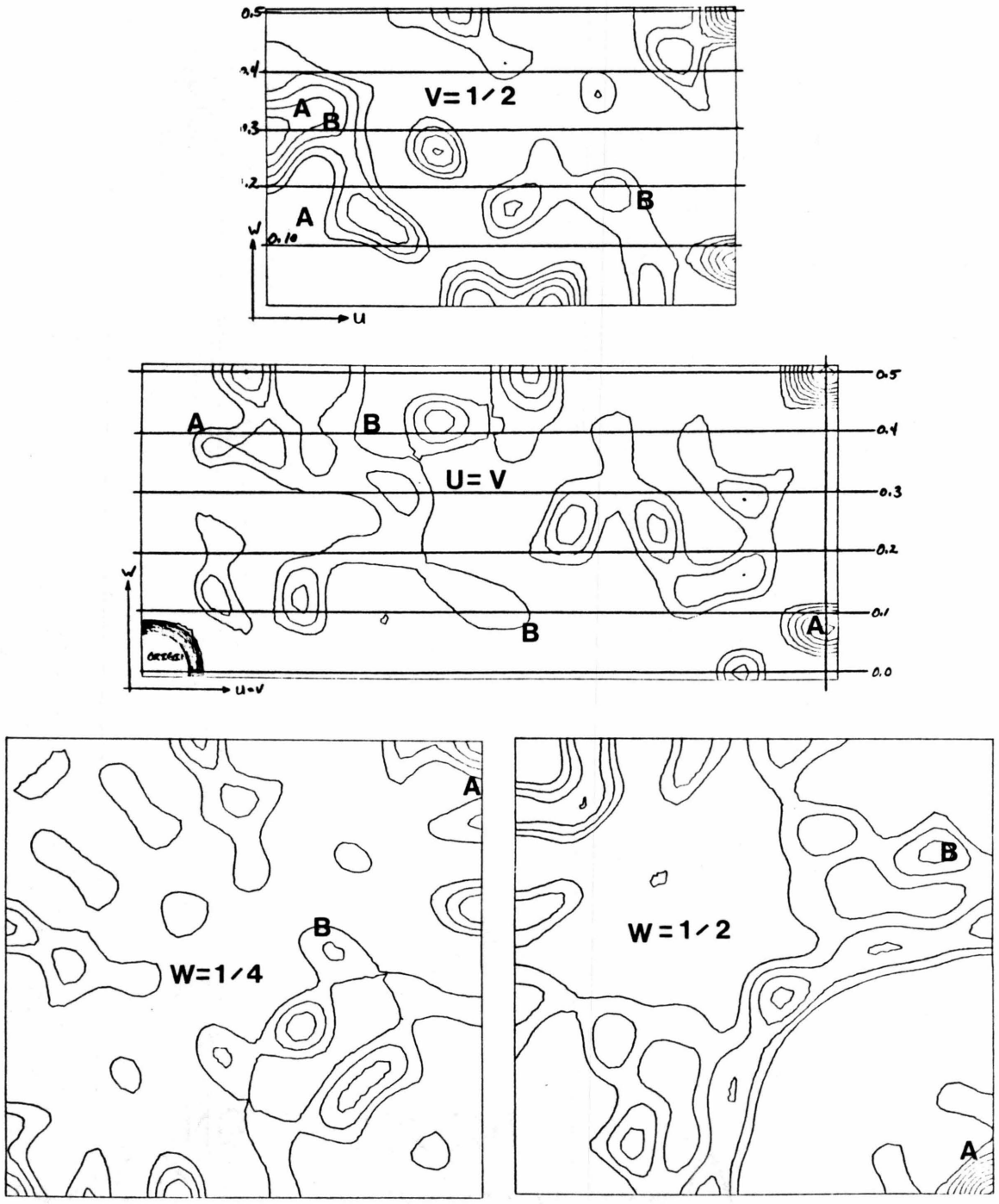

Figure 2. Anomalous Harker sections

Since there are two iodines in the asymmetric unit, each one of these sections

should have at least four peaks on it. Possible solutions for iodines A and B were tested by two methods. To test a single iodine position, all of the self vectors, *eg.* A to A vectors, should appear on the Harker sections. When two possibilities were found, they were tested with each other by looking in the three dimensional Patterson map for appropriate cross vectors, *eg.* A to B vectors.

The positions of the correct I-I peaks are marked on the Patterson as A or B. The fact that these positions are only rarely in the center of the large peaks resulted in a good deal of uncertainty in their assignment. Supposedly, the anomalous Patterson helps peak assignment by enhancing those peaks that result from anomalously scattering atoms. That is to say, if a peak is stronger in the anomalous Patterson than in the normal Patterson, it should be due to the anomalous scatterer, iodine. Conversly, if the peak is weaker in the anomalous Patterson than in the regular Patterson, it must be due to a pileup of vectors from atoms that scatter normally. These guidelines turned out to be false.

The problem is that when the $(|\mathbf{F}_{hkl}| - |\mathbf{F}_{\overline{hkl}}|)^2$ coefficients are formed for an anomalous Patterson, the centric data are effectively omitted, since they have no anomalous differences contributions. In a space group with many centric zones, low resolution data may contain a large percentage of centric reflections. In $P4_32_12$, the 3 Å data set contains 36% centric reflections, and severe distortions occur in the anomalous Patterson map as a result of this omission.

This loss of centric data may not be a problem unless the anomalous scatters are located near crystallographic symmetry elements of the cell. However, in the case of ICCGG, the two iodines are positioned close to twofold axes in such a way that A, B, and X vectors (vectors between positions A and B) often point along cell axes. In a normal Patterson map the centric reflections are important in defining these positions. Since the centric reflections are missing from an anomalous Patterson, these peaks are not as well defined as they should be. Therefore, there were some cases

where the I-I peaks was actually diminished in the anomalous Patterson map rather than enhanced. To distinguish such peaks, a total of three Patterson maps were used.

- A normal Patterson map (*ie.* $|\mathbf{F}_{hkl}|^2$ as coefficients)
- A normal Patterson map containing only acentric data (*ie. only* acentric $|\mathbf{F}_{hkl}|^2$ as coefficients)
- An anomalous Patterson map (*ie.* $(|\mathbf{F}_{hkl}| - |\mathbf{F}_{\overline{hkl}}|)^2$ as coefficients)

These Patterson maps are shown in Figures 1, 2, and 3.

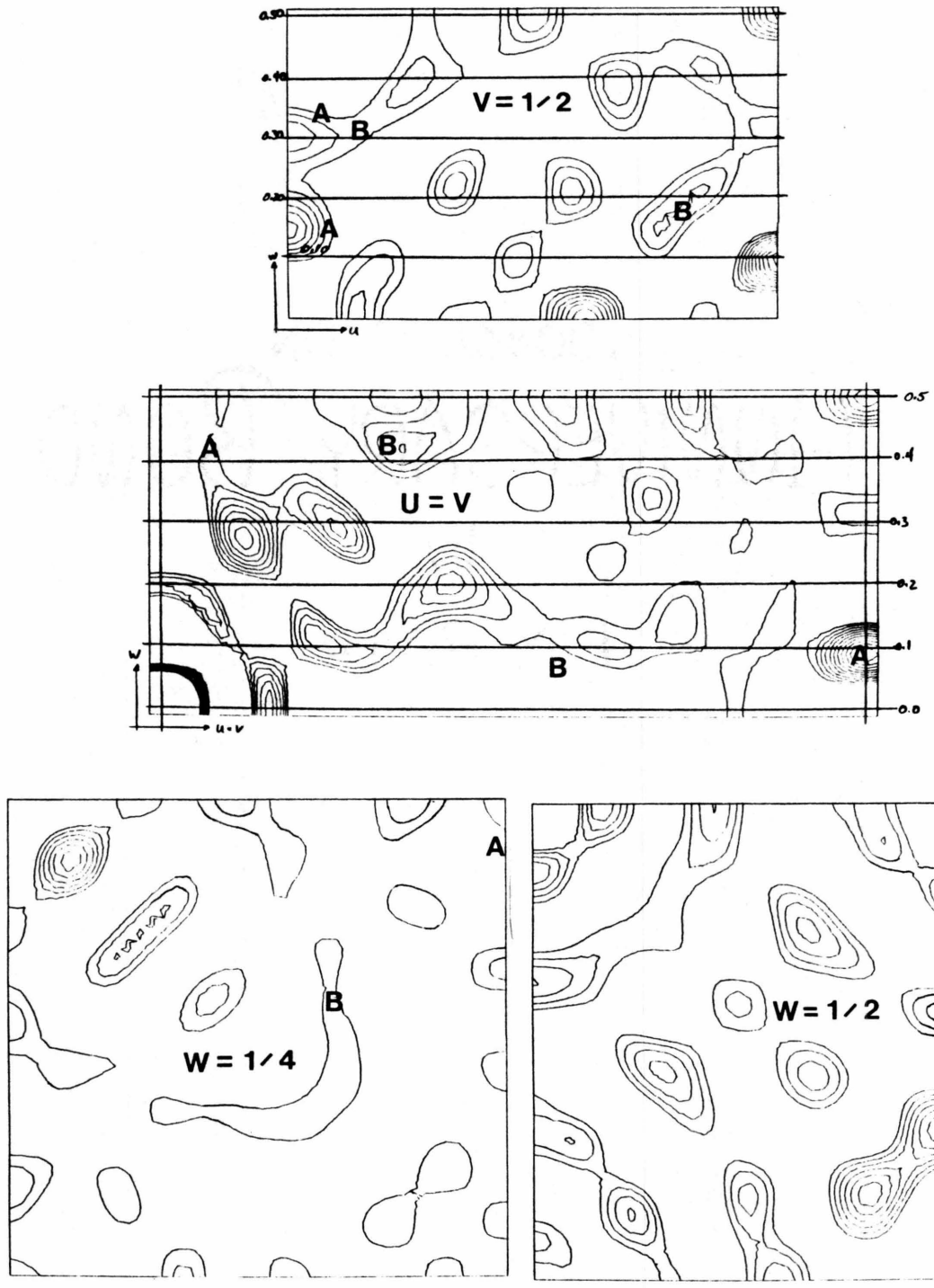

Figure 3. Harker sections from 4/mmm averaged **acentric** data

The following algorithm was used to sort peaks from the Patterson maps:

1. If a peak is higher in the anomalous map than in the normal map, it is considered a possible I-I vector.

2. If a peak is lower in the anomalous map than in the normal map, its height must be checked in the acentric map.

   A. If the peak is relatively the same height in normal and acentric maps, then its diminution in going from normal to anomalous maps indicates it is not an I-I vector.

   B. If the peak decreases in height in going from normal to acentric maps, then the diminution in going from normal to anomalous maps may only be caused by the loss of centric data. *Therefore, this peak must also be considered a possible I-I vector.*

Using the vectors added by this procedure, a self consistent set of positions for the two iodines was found.

One method used to check these two iodine positions was to use one position as a heavy atom in phasing the electron density map, and look for the other iodine position to appear. This was not a good test. Even when the correct positions were used, maps phased with only the first iodine did not show the second iodine well above the noise level. The positions of the iodines was uncertain until the very end, when the correct anomalously phased electron density map was calculated.

3.6 Use of Anomalous Differences in Phasing

The difference in intensity between I_{hkl} and $I_{\overline{hkl}}$, the Bijvoet difference, can help determine the phase of the hkl reflection. In macromolecular crystallography, several different methods are used. These techniques may use the anomalous differences by themselves or combine them with other phase information. As will be discussed later, a single anomalously scattering atom can reduce the phase determination problem for an acentric reflection to a choice between two possibilities. To

make this choice, several other sources of information can be used.

In one example,[5] data were collected on a single crystal using two different x-ray wavelengths near the absorption edge of the anomalous scatterer. The way in which amplitude and phase of the anomalously scattered wave change with input wavelength allows phases to be calculated for the reflections. This method is particularly effective when a tunable x-ray source, such as a synchrotron, is used, so that any desired wavelength can be chosen.[6]

Anomalous phasing is often combined with the method of isomorphous replacement to provide phasing information. Isomorphous replacement involves making a heavy atom derivative of a given crystal, usually by soaking the crystal in a solution containing the heavy atom. By studying the change in intensity between the native crystal and the derivative, phases for the reflections can be found. In practice, suitable derivative crystals are difficult to make and it is best to extract the most information from each one.

When only a single isomorphous derivative exists, the phase problem can be reduced to choice between two possibilities. This choice is usually resolved by making a second isomorphous derivative. However, since these derivatives usually contain heavy atoms such as samarium, uranium or iodine, they often have significant anomalous scattering components. Bijvoet differences collected from these derivatives can resolve the twofold choice in single isomorphous phasing.[7] In effect, each isomorphous derivative becomes twice as useful.[8]

If multiple isomorphous derivatives are used, the phase problem can be solved using only isomorphous data, without considering anomalous scattering. However,

---

5. W. Hoppe and U. Jakubowski (1975). in *Anomalous Scattering*, edited by S. Ramaseshan and S. C. Abrahams, Copenhagen:Munksgaard, 437-461
6. J. C. Phillips, A. Wlodawer, J. M. Gooodfellow, K. D. Watenpaugh, L. C. Seeker, L. H. Jensen and K. O. Hodgson (1977). *Acta Cryst.* **A 33**, 445-455
7. A. C. T. North (1965). *Acta Cryst.* **18**, 212-216
8. J. R. Knox, J. A. Kelly, P. C. Moews and N. S. Murthy (1976). *J. Mol. Biol.* **104**, 865-875

addition of anomalous data allows a more precise determination of the phase angle by adding more information, which narrows the range of errors. This additional information improves the quality of the electron density map and allows interpretation to proceed with greater ease and confidence.[9]

3.7 Solution of the Phase Problem

^{I}CCGG could not be solved by the isomorphous method because no two of the available crystals belonged to the same space group. A multiple wavelength analysis similar to that of Hoppe and Jakubowski[10] was briefly considered. However, the expense of a new x-ray tube, the difficulty in aligning the diffractometer, and the need to collect twice as much data all argued against this technique.

However, the iodine is almost heavy enough to phase the reflections itself using standard heavy atom techniques. It was believed that the combination in the proper weighting scheme of heavy atom information and anomalous dispersion information would be sufficient to solve the structure. In fact, this method worked very well. The initial 3 Å Fourier map showed the DNA molecules clearly separated from the solvent background. Within the molecule itself, the base planes were resolved and the sugar phosphate backbone could be easily traced. The clarity of the initial map allowed the elimination of various left handed [11] [12] and other[13] models which have proposed in recent years.

A method to phase reflections based on Bijvoet pairs from a single crystal has already been worked out.[14] Once the positions of all the anomalous scatterers in the cell are known, the phase of a particular reflection can be calculated by the equa-

---

9. J. R. Herriot, L. C. Sieker, L. H. Jensen and W. Lovenberg (1970). *J. Mol. Biol.* **50**, 391-406
10. W. Hoppe and U. Jakubowski (1975). ibid
11. V. Sasisekharan, M. Bansal and G. Gupta (1981). *Biochem. Biophys. Res. Com.* **102**, 1087-1095
12. G. Gupta, M. Bansal, V. Sasisekharan (1980). *Proc. Nat. Acad. Sci. USA* **77**, 6486-6490
13. R. C. Hopkins (1981). *Science* **211**, 289-291
14. G. N. Ramachandran and S. Raman (1956). ibid

tion: (Please refer to Figure 4 for identification of components.)

$$\alpha_{hkl} = \frac{\pi}{2} + \alpha_A \pm \cos^{-1} \Theta \tag{3.9}$$

where

$$\Theta = \left(\frac{F_{hkl} - F_{\overline{hkl}}}{2\, F''_{hkl}} \right)$$

and

- α_{hkl} is the phase for the scattering from all atoms in the cell and includes the real scattering component from iodine.
- α_A is the phase of the scattering due to the normal dispersion from the iodine.
- F''_{hkl} is the magnitude of the anomalous scattering.

Once α_{hkl} is known, it can be used in a Fourier synthesis with F_{hkl}, the amplitude of scattering due only to normal dispersion, to calculate the electron density map for the crystal. An example of phasing from ^{I}CCGG, $\mathbf{F}_{3,1,1}$ is shown in Figure 4. The two possible choices for phase occur where the two circles intersect.

In order to minimize the mean square error in the electron density map, each of the two phases is included in the map with a weight proportional to the probability of its being the correct choice.[15] Different weighting schemes were used for the two types of data, centric and acentric. Since the centric data contained no anomalous differences, they were weighted by the heavy atom procedure of G. A. Sim[16] which assigns to the phase calculated by the heavy atom method a weight which depends on the probable error between the heavy atom phase and the true phase. The acentric data are phased by the Bijvoet differences, so they are weighted by G. A. Sim's method for anomalous phasing.[17] This scheme weights each of the two phase angles

15. M. M. Woolfson (1956). *Acta Cryst.* **9**, 804-810
16. G. A. Sim (1969). *Acta Cryst.* **13**, 511-512
17. G. A. Sim (1964). *Acta Cryst.* **17**, 1072-1073

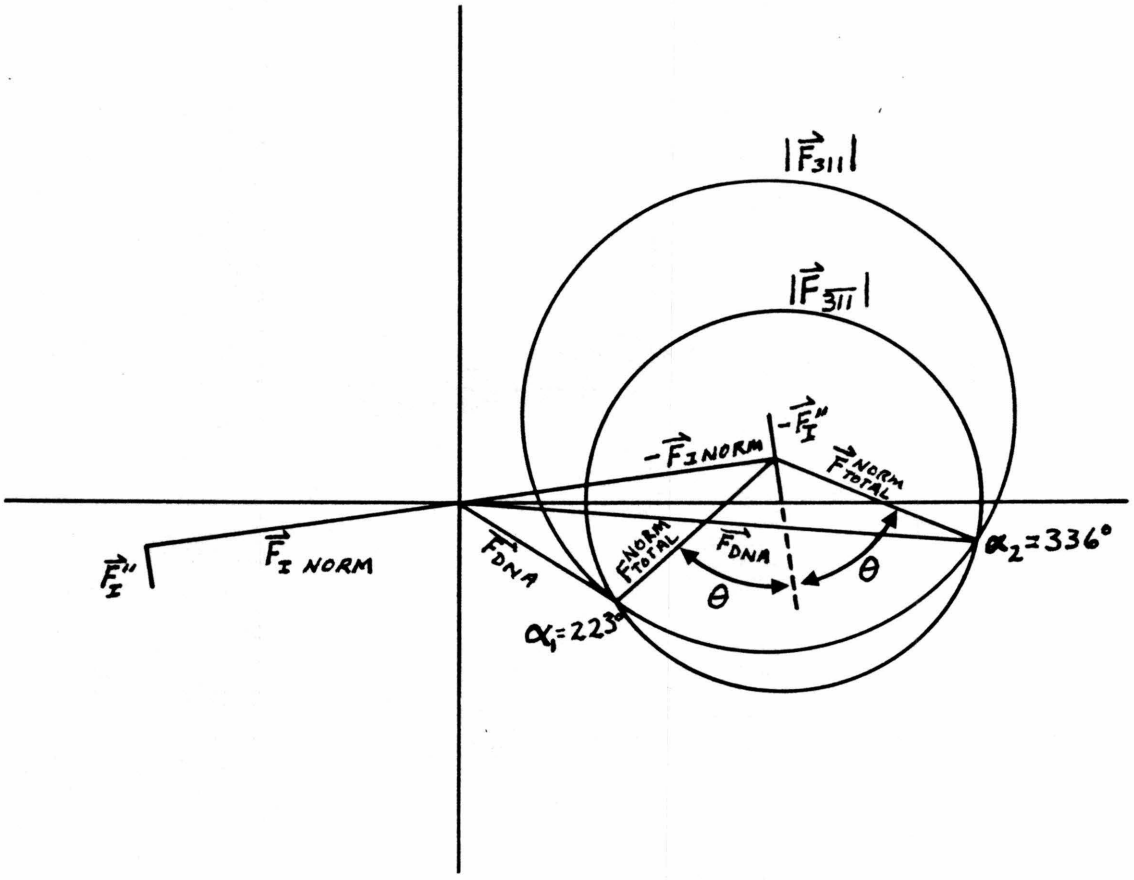

Figure 4. Reflection 3,1,1 phased by iodines at (x, y, z) in $P4_32_12$

based on the following factors.

- The phase angle closest to the heavy atom phase angle receives the most weight.
- Reflections with larger heavy atom contributions receive heavier weight.
- Reflections with larger total intensity receive more weight.

Figure 5 shows an example of phasing from ^{I}CCGG where the two phase choices are close together. In this situation, Sim's weighting method assigns 50% probability to

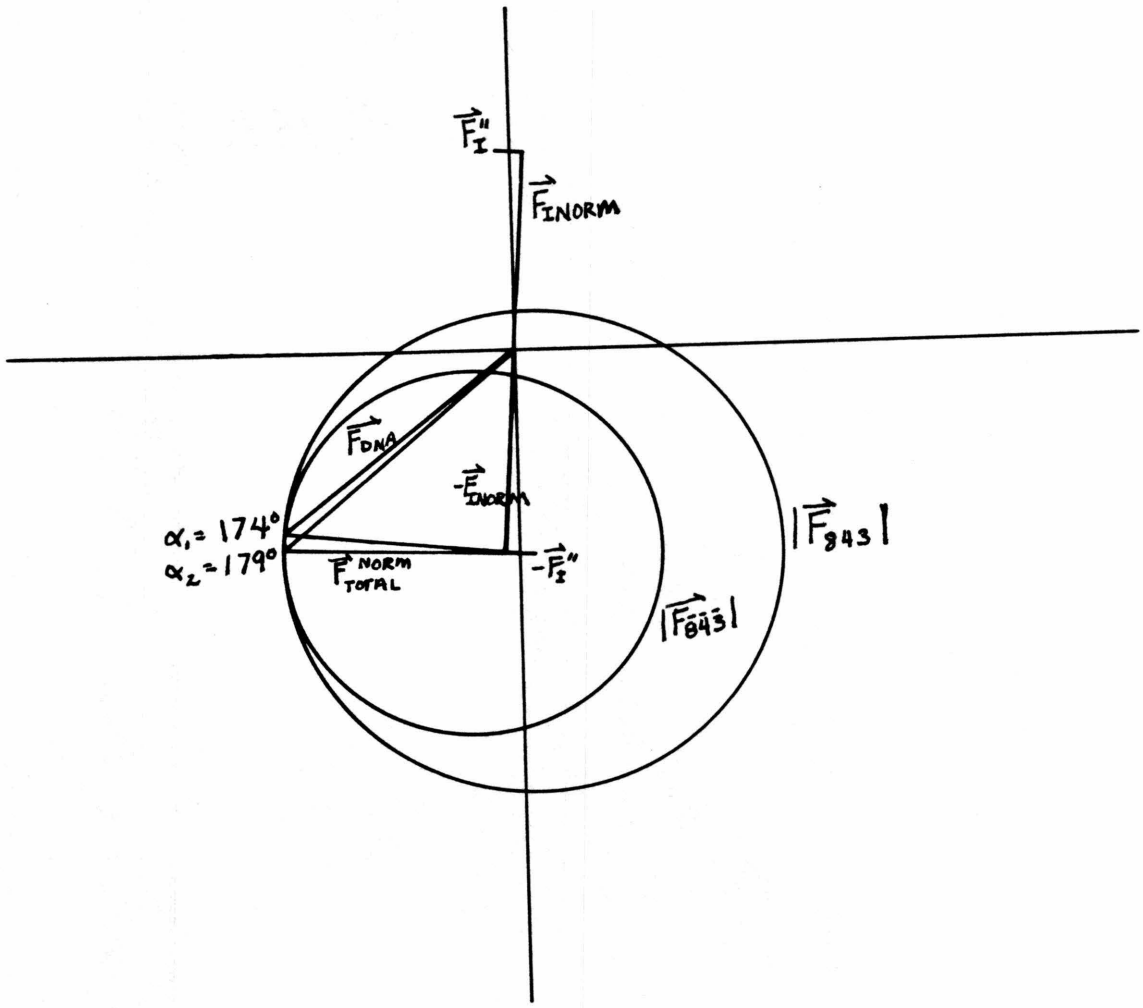

Figure 5. Phasing diagram with Θ approx 0 °; F_{843}

each solution and includes them both. Figure 6 shows an example of phasing from ICCGG where the two phases are relatively far apart and the ambiguity of the phase is at its greatest. In this situation, Sim's weighting method assigns the phase closest to the heavy atom phase the greatest weight, 82%.

3.8 The Final Choice

There is one final choice which must be made when using a single crystal for anomalous phasing. Although a self consistent set of iodine positions was found in

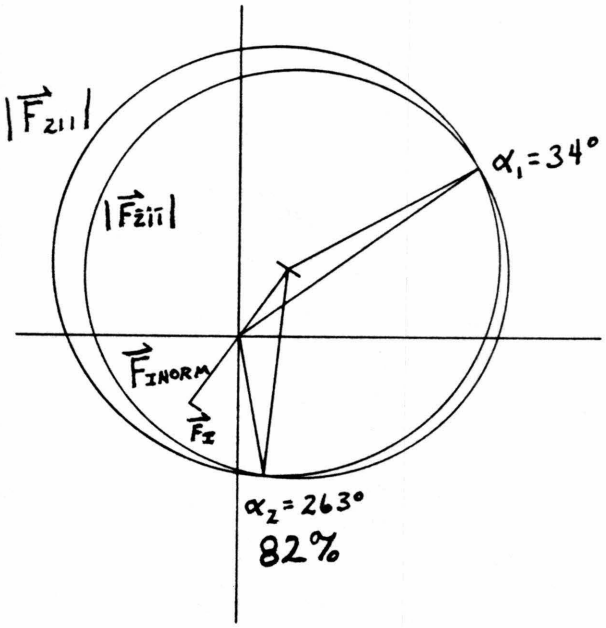

Figure 6. Phasing diagram with Θ approx 90 °; F_{211}

the Patterson maps, there actually are two possible solutions. An atom at (x,y,z) in space group $P4_32_12$ produces a Patterson map identical to an atom at ($\bar{x}$, $\bar{y}$, $\bar{z}$) in $P4_12_12$. When anomalous phasing methods are used, this becomes an important problem. Figure 4 illustrates the calculation of the phase with what later turned out to be the correct choice, space group $P4_32_12$. Figure 7 shows the same reflection phased by iodines at ($\bar{x}$, $\bar{y}$, $\bar{z}$) in $P4_12_12$. The phase is considerably different. The only way to make this choice is to actually construct separate electron density maps for each of the possible solutions. The correct solution, $P4_32_12$, gave a very clear

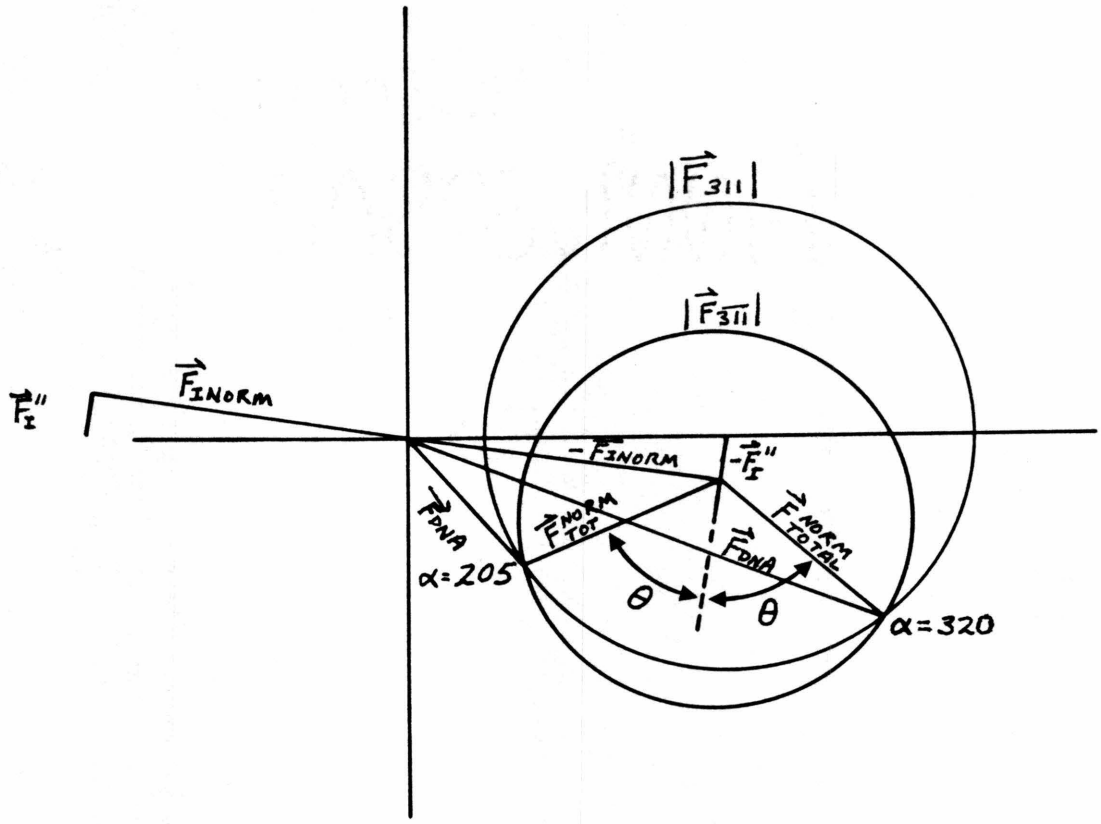

Figure 7. Reflection 3,1,1 phased by iodines at ($\bar{x}$, $\bar{y}$, $\bar{z}$) in $P4_12_12$

picture of the tetramer in which even the base pairs could be resolved from one another. The other solution, space group $P4_12_12$, produced a map with no intelligible features. The decision for $P4_32_12$ was unambigious.

Chapter 4. Intermediate Results

CHAPTER 4. INTERMEDIATE RESULTS

Although only a few water molecules had been located after 28 cycles of refinement, most of the DNA had settled into place. At this point a structural analysis was performed and the results published in *Nature*.[18] ^{I}CCGG was compared with the fiber structure of A-DNA and some preliminary information on the hydration structure was presented. The remainder of this chapter is that paper.

One set of useful illustrations which were not included in the paper are the stereo drawings in Figure 1. These show the sequence CCGG as determined by fiber diffraction.[19] This figure was not included in the paper because of space limitations. However, it is very useful for comparison with ^{I}CCGG because it is drawn from the same viewpoint as Figure 1 of the *Nature* paper. The change in propellor twist can be clearly seen by comparing the side views of the two tetramers.

18. B. N. Conner, T. Takano, S. Tanaka, K. Itakura, and R. E. Dickerson (1982). *Nature* **295**, 294-299

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

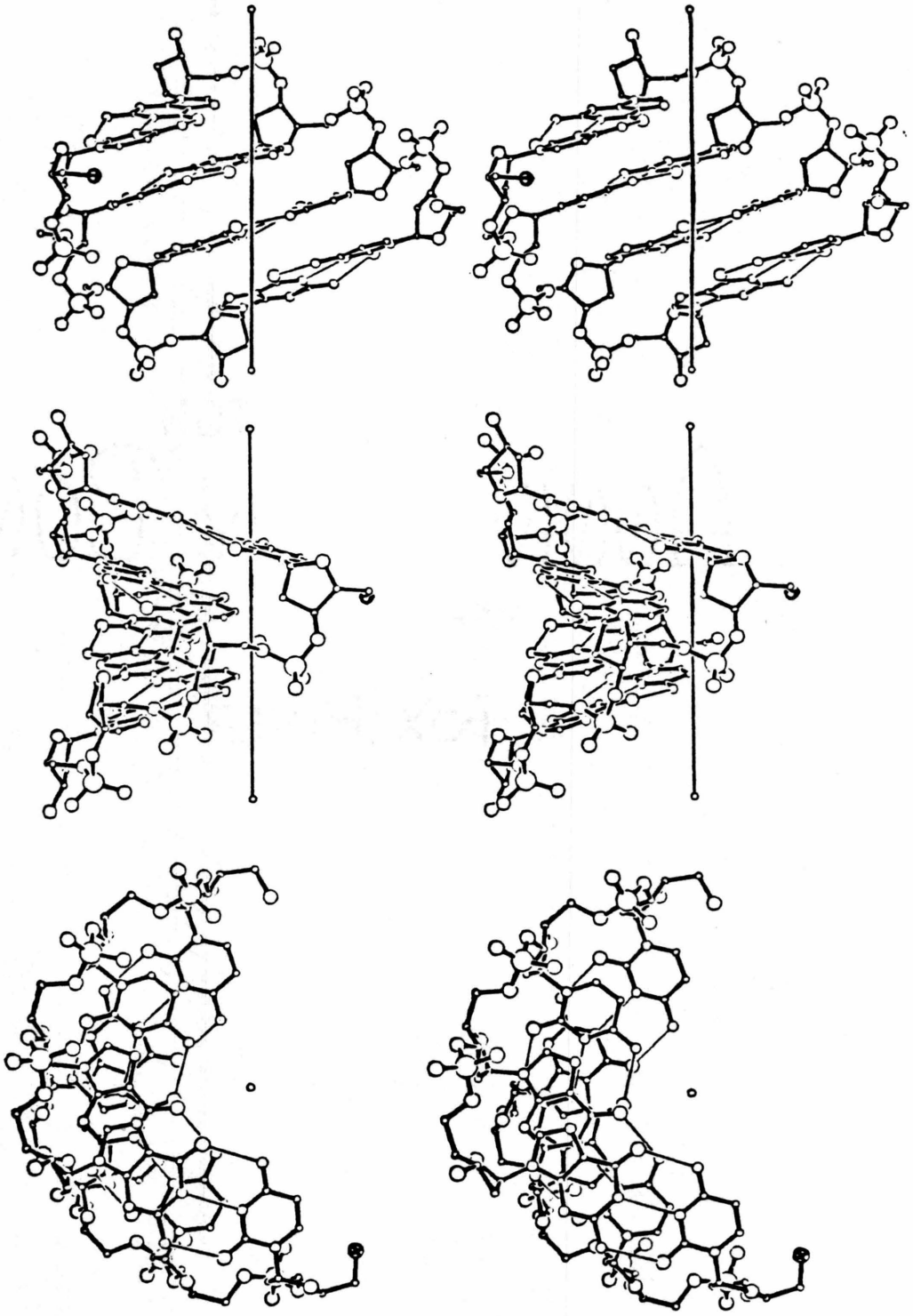

Figure 1. Fiber A-DNA from Arnott and Hukins (1972). *Biochem. Biophys. Res. Com.* **47**, 1504-1509

294 Nature Vol. 295 28 January 1982

The molecular structure of d(ICpCpGpG), a fragment of right-handed double helical A-DNA

Benjamin N. Conner*, Tsunehiro Takano*, Shoji Tanaka*, Keiichi Itakura† & Richard E. Dickerson*

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

The DNA tetramer d(ICpCpGpG) or ICCGG crystallizes as a double-stranded 4-base pair (bp) segment of an A helix. Two such tetramer helices are packed together in the crystal with local helix axes nearly coincident, simulating an 8-bp helix, and four such octamers make up the tetragonal unit cell. Restrained energy and reciprocal space refinement has led to an R factor of 20.5% at 2.1 Å resolution. The ICCGG helix has a twist corresponding to 10.7 bp per turn, a 19° base tilt and a 2.3 Å rise per base pair along the helix axis. The mean propeller twist of 18° is comparable with, and has the same rotational sense as that observed in the B-DNA dodecamer CGCGAATTCGCG at similarly high alcohol concentration. Backbone phosphate groups in A-DNA are extensively hydrated, including a network across the opening of the major groove, whereas base edge N and O groups in major and minor grooves are less hydrated than in B-DNA. The minor groove spine of hydration observed in B-DNA is totally absent. These observations of relative hydration confirm and extend the model for the B-to-A helix transition proposed earlier on the basis of the B helix structure.

THE first double-helical DNA structures were proposed on the evidence of X-ray diffraction data from drawn fibres obtained from natural sources, supplemented by spectroscopic and chemical information, and model building from the known stereochemistry of component parts[1-7]. Two major families of helix were proposed, A and B, differentiated by sugar puckering, base tilt relative to the helix axis, rise per residue along the axis, distance of base pairs from the centre of the helix and relative dimensions of major and minor grooves (Table 1). A-DNA is characterized by a deep and narrow major groove, a shallow minor groove and base pairs tilted by roughly 19° to the helix axis. In various members of the B family, the bases are more nearly perpendicular to the helix axis and the grooves are of comparable depth, although the major groove is wider.

Fibre diffraction patterns can give the overall average helix parameters for DNA of known composition or repeating sequence, but cannot reveal local variations in helix structure that might be produced by that sequence. This kind of information can only come from single-crystal structure analyses of discrete, homogeneous DNA oligomers. The development of triester and related methods has made the synthesis of such oligomers possible[8]. Structure analyses have been carried out for both a 12-base pair (bp) B helix[9,10] and a totally unexpected new family, the left-handed Z helix[11-13]. We describe here the first single-crystal structure analysis of a member of the A-DNA family, the self-complementary tetramer d(iodo-CpCpGpG) or ICCGG. It is also the first example of a nucleic acid structure analysis carried out without isomorphous heavy atom derivatives, using initial phasing derived from anomalous scattering as was done for the proteins haemerythrin and crambin (ref. 14 and W. A. Hendrickson and J. Smith, personal communication). (In another structure analysis of an RNA dimer, r(ApU), complexed with 9-aminoacridine, anomalous scattering was used to locate positions of phosphorus atoms, which then became the starting point for non-anomalous phase analysis[15].)

Structure analysis

To find an isomorphous pair of crystal forms, the native CCGG tetramer and four variants with 5-bromocytosine or 5-idocytosine in the first or second position along the chain were synthesized independently by the triester method. Four of these compounds were crystallized, but no two adopted the same space group. Having discussed with W. Hendrickson his experiences with haemerythrin and crambin, we decided to try to solve the better crystal form of the two iodo derivatives, ICCGG, using anomalous scattering from iodine atoms.

Crystals of the 1-iodo derivative were grown at 2 °C by vapour diffusion against isopropanol, from concentrated aqueous solution of ICCGG containing spermine hydrochloride buffered at *p*H 7.5 with sodium cacodylate. The cacodylate buffer was later found to be unnecessary for crystal growth. Crystals began appearing at around 40% isopropanol, and crystal growth was continued to 80% alcohol. The space group of ICCGG crystals is $P4_32_12$, with cell dimensions $a = b = 41.1$ Å, $c = 26.7$ Å and two tetramer strands (one double helix) per asymmetric unit. A measured density of 1.50 g cm^{-3} and a calculated molecular weight of 1,298 per strand indicate that the crystals are 51.0% DNA by weight.

Data were collected on a Syntex P$\bar{1}$ automatic diffractometer with graphite monochromator, and corrected for X-ray decay and absorption[16]. The pattern was quite strong to 3.5 Å resolution, but declined rapidly beyond that point. One anomalous data set to 2.1 Å resolution contained 1,486 Friedel pairs. Multiple data sets were collected on different crystals for accuracy, so that each reflection within the 3.0 Å sphere was measured at least three times. Mean R factors between independent data sets were ~5% in F.

Patterson and anomalous difference Patterson maps[17] were employed to locate the two iodine positions per asymmetric unit (16 per cell), using the 472 pairs of reflections above the 2σ confidence level within the 3.0 Å sphere. Approximately 36% of these reflections were centric and hence gave no anomalous contribution, but the iodine anomalous scattering factor of $\Delta f'' = 7.2$ electrons for CuKα radiation produced an average difference between Friedel pairs of 11% in F among acentric reflections. The location of iodine atoms was complicated by two factors: the iodine of cytosine C1 was accidentally situated

* Present addresses: Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90024, USA (B.N.C., R.E.D.); Institute of Protein Research, Osaka University, Osaka, Japan (T.T.); Laboratory of Microbiology, Suntory Institute for Biomedical Research, Mishima-gun, Osaka 618, Japan (S.T.).

0028-0836/82/040294—06$01.00

Nature Vol. 295 28 January 1982

Table 1 Structural parameters in double-helical DNA

	[1]CCGG crystal structure	Fibre diffraction results					
		A	B family			Z family	
			B	C	D	Z	Z'
Helix sense	Right	Right	←	Right	→	Left	Left
Sugar pucker	C-3'-*endo*	C-3'-*endo*	←	C-2'-*endo*	→	Alternating	
Glycosyl angle	Anti	Anti	←	Anti	→	Syn(G)/anti(C)	
Twist per base pair	33.6°	32.7°	36°	38.6°	45°	−60°/2	−60°/2
Bases per turn	10.7	11	10	$9\frac{1}{3}$	8	12	12
Rise per base pair (Å)	2.3	2.6	3.4	3.3	3.0	3.7	3.8
Base tilt	19°	19°	−6°	−8°	−16°	−7°	−9°
Groove width (Å)							
Minor	8.8	11.0	5.7	4.8	1.3		
Major	2.0	2.7	11.7	10.5	8.9		
Groove depth (Å)							
Minor	3.7	2.8	7.5	7.9	6.7		
Major	13.8	13.5	8.5	7.5	5.8		
Refs	This work	4	4	4,5	4,5	11	13

Groove width is the perpendicular separation of ideal helix strands drawn through the phosphate groups, decreased by 5.8 Å to represent van der Waals radii of two phosphate groups. Groove depth is defined precisely in ref. 4, but again is based on van der Waals radii, as though measured from a space-filling model.

directly above a 2-fold axis, with coordinates close to $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, and that on cytosine C5 of the other strand of the double helix was nearly just one-quarter of the cell away from it along the *a* axis. Each of these two coincidences introduced false centrosymmetry that decreased the iodine anomalous contribution to many of the acentric reflections. Thus the simple $(\Delta F_{ano})^2 = (F_{hkl} - F_{\overline{hkl}})^2$ anomalous difference Patterson map was not interpretable. Iodine positions were ultimately found by comparing F^2 Patterson maps using both centric and acentric reflections, with F^2 Patterson maps using only acentric data. As we suspected that our difficulties arose because the iodines sat on or near symmetry elements, those Patterson peaks which were deleted by the removal of centric data were reinforced in the simple anomalous difference Patterson map, and the composite map was then interpretable in terms of a self-consistent set of iodine–iodine vectors.

Once the iodines were located, electron density maps were calculated using Sim-weighted heavy atom phases[18] for centric reflections, and Sim-weighted anomalous phases[19,20] for acentric reflections. Of the two possible enantiomorphic maps, one was uninterpretable, while the other gave a clear and clean image of the tetramer double helix. The helix obtained from this map was subjected to 28 cycles of simultaneous energy and reciprocal space refinement[21], using the same energy constraints as for the B-DNA dodecamer[10]. Refinement progress was monitored periodically by Fourier and difference Fourier electron density maps. During this process, the *R* factor fell from 39% at 3.0 Å resolution (2σ data) to the current 20.5% at 2.1 Å. The model contains 29 solvent molecules at present and refinement is continuing.

Tetramer structure

The complete unit cell contains eight identical double-helical tetramers, one of which is shown in Fig. 1. The tetramer is a fragment of an A-DNA double helix. A helix-generating program provided by J. Rosenberg enabled us to use the relationships between base pairs within one tetramer to define the parameters of an ideal continuous helix (see Fig. 1). The helix twist angle is 33.6° per base, corresponding to 10.7 bp per turn, and the rise per base pair along the the axis is 2.3 Å. Bases are tilted by 19° from perpendicularity to the helix axis. Sugar rings are all in the C-3'-*endo* conformation with the exception of the 3'-terminal sugar of guanine G4, which is C-2'-*endo*. (See upper left of Fig. 1*b*, or lower left of Fig. 1*c*.) However, as the G4 sugar has no chain continuation to force it into a regular geometry, it is not surprising that it adopts an anomalous conformation. Table 1 compares observed CCGG helix parameters with idealized A-DNA values derived from fibre data and with those of other types of helix.

The base pairs have large propeller twists: 19°, 18°, 18° and 23° for C1–G8, C2–G7, G3–C6 and G4–C5 respectively. The large propeller twist was visible early in the analysis of the anomalously phased electron density map, and hence is not a consequence of the refinement process. The last value is perturbed by the same intermolecular crystal packing that produces the tilted C5 plane visible in Fig. 1, but the other three values are quite consistent. They are larger than the 13° average propeller twist observed for B-DNA in the low-alcohol crystals (35% methylpentanediol, MPD) of CGCGAATTCGCG[10], and more like the 19° observed in the higher alcohol (60% MPD) crystals of CGCGAATT^{Br}CGCG (M. L. Kopka, A. V. Fratini, H. R. Drew and R.E.D., in preparation). The sense of the propeller rotation in A-helical CCGG is the same as that in these B helices: clockwise rotation of the nearer base plane when the base pair is viewed along its long axis, defined here as the positive rotational sense for propeller twist. In contrast, the published coordinates for a model A-DNA structure derived from fibre diffraction data[22] yield an opposite rotation, with a propeller twist angle of −12°. (A recent unpublished set of new fibre-derived A-DNA coordinates supplied by S. Arnott has positive propeller twist angles of +7.5°.) Two RNA dimers are known which are fragments of an A helix: r(ApU) and r(GpC)[23,24]. In these molecules the propeller twist is smaller and of variable sign: +12.5° (average) for r(ApU) and −7.2° for r(GpC). It seems likely that a substantial propeller twist is a consequence of efficient base stacking at successive steps along one strand of a continuous double helix, and hence reasonable that the full twist typical of an infinite helix would not be generated in helices shorter than 3 or 4 bp. Indeed, one unexpected feature apparent from the single-crystal analyses of oligomers of A- and B-DNA has been the large propeller twist angles, which were not anticipated from fibre and dimer studies.

A-helical octamers in the crystal

The most apparent distortion of the tetramer in Fig. 1 is the pushing of the iodine on cytosine C5 towards the centre of the molecule, with a resultant bending of the C5 plane away from that of G4. This is simply explained in terms of the way in which tetramers are packed within the crystal. Two tetramers are stacked with their C1–G8 base planes in contact, their helix axes nearly coincident and a rotation between tetramers that closely approximates a continuous 8-bp A helix (Fig. 2). Four of these octamers are then packed around the 4-fold screw axes of the

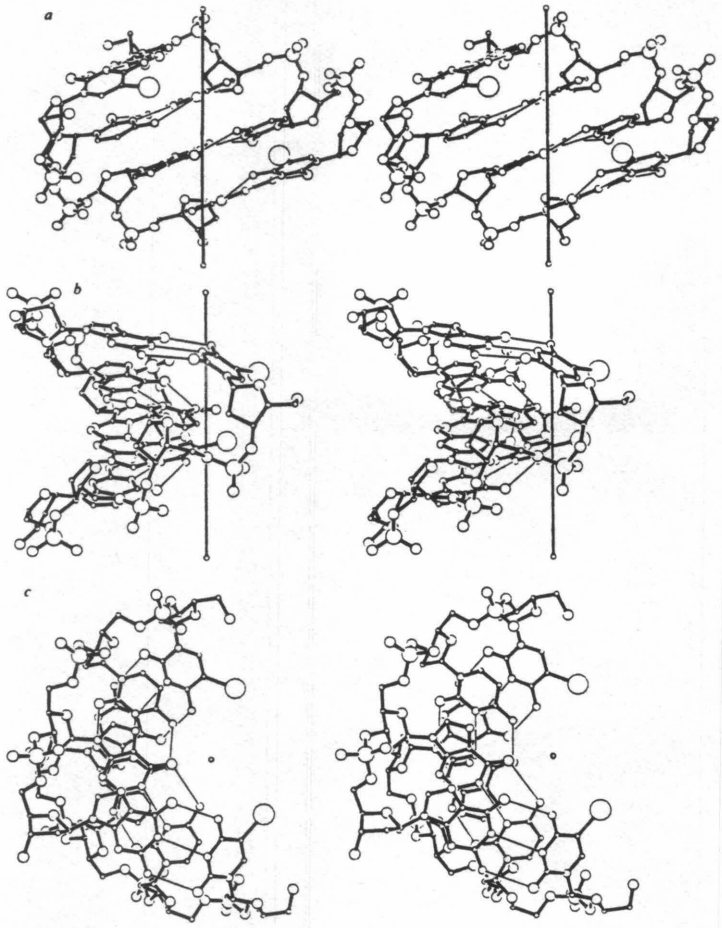

Fig. 1 Stereo views of the ^{1}CCGG double-helical A-DNA molecule. *a*, View into the major groove with helix axis vertical. Bases are numbered: 5′-C1-C2-G3-G4-3′ in one strand and 5′-C5-C6-G7-G8-3′ along the other. The two largest atoms are iodines on cytosines C1 (lower right) and C5 (upper left), and other atoms in order of decreasing size are P, O, N and C. Hydrogen bonds between base pairs are shown as thin lines. Base pair C1–G8 is at the bottom and G4–C5 at the top. *b*, View 90° to the left of *a*, showing the shallow minor groove in profile at left, and the large propeller twist of base planes. *c*, Top view looking directly down the helix axis onto base pair G4–C5. The iodocytosine base planes are tilted away from the planes of their hydrogen-bonded partners, C5 more so than C1, because of intermolecular contacts within the crystal.

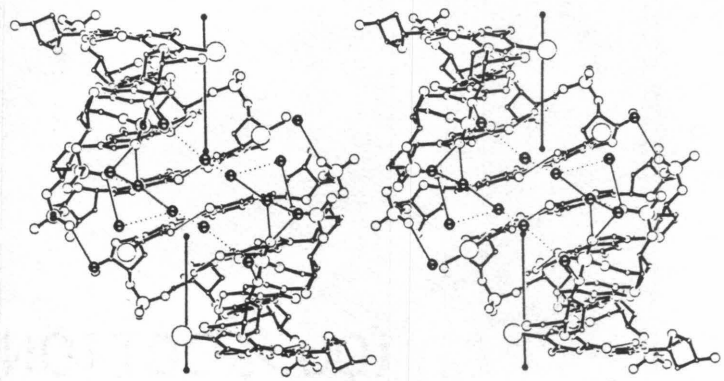

Fig. 2 Two ^{1}CCGG tetramers stacked around a 2-fold axis (normal to the page through the centre of the drawing) as they are observed in the crystal. Their C1–G8 base pairs are in contact, and have very nearly the correct displacement to continue the A helix for 8 bp. Local helix axes, as determined independently for the two tetramers, are shown by vertical lines, and are almost coincident. This view is directly into the major groove, lined by phosphates 6, 7 and 8 (numbered on one strand). After the missing phosphate between tetramers, a continuous helix would use phosphates 2, 3 and 4 around the back of the drawing on the other tetramer. Seven water molecules and their symmetry-related partners are shown spanning the major groove (⊖). Distances of ≤3.4 Å are shown by thin solid lines, and distances up to 3.8 Å by dotted lines. As discussed in the text, in a continuous A-DNA helix the major groove would be narrower than shown here, and the bridging network of hydration would be different in detail although similar in character.

$P4_32_12$ cell. The iodocytosine ring C5 at each end of an octamer (top and bottom of Fig. 2) is packed against the minor groove of a neighbouring octamer, between bases C2 and G3. Although the C5 iodine atom is pushed back towards the centre of the octamer, it still is in tight van der Waals contact with the C-2' sugar atom of cytosine C2 and the C-5' atom of guanine G3 on a neighbouring octamer.

At the centre of each octamer (deep inside the major groove in Fig. 2), the iodine of cytosine C1 on one tetramer is displaced slightly because of a close contact with the five-membered ring of G8 on the other tetramer (the next base step along the helix), but the effect on its cytosine orientation is small. The two local helix axes as viewed in Fig. 2 are displaced sideways by 1.5 Å, which serves to lessen the overlap between each iodine and its facing G8 ring. The extreme top and bottom of the two axes in Fig. 2 are tipped down into the plane of the page so the axes make an angle of 8° to one another, which also lessens the iodine/guanine clash. The absence of phosphate groups to connect the tetramers allows a degree of adjustment that might not be present in a true double-stranded octamer.

Four octamers as shown are packed into the tetragonal unit cell of the crystal in helical arrays around water channels with elliptical cross-sections, parallel to the z axis. Each octamer uses its relatively flat minor groove side to define a long side of one elliptical channel, and its more concave major groove to make the sharp bend at the short side of a different elliptical channel. The individual octamers are hydrated as described below, and networks of localized solvent molecules extend across the channels.

The idealized extended A helix

The octamer in Fig. 2 does not provide a true impression of a continuous A-DNA helix with this helix geometry. Figure 3 shows two views of a full turn of helix, constructed by stacking three identical tetramers on top of one another with their local helix axes coincident. The very deep major groove is particularly striking in stereo (Fig. 3*a*). The groove is narrower in the continuous helix (Fig. 3) than in the crystal structure octamer (Fig. 2) because of the 8° angle between local helix axes in the octamer. If this angle were straightened out to 0°, then the major groove in the octamer would appear as in Fig. 3*a*, with phosphorus–phosphorus separations across the groove of 7.8 Å rather than 10.9 Å, the P6–P6 separation in Fig. 2. This factor will become important in the following discussion of hydration.

The shallow minor groove is hardly a groove at all in comparison with its structure in B-DNA. Figure 3*b* shows the hollow channel down the centre of the helix, a feature peculiar to A-DNA. In a sense the A helix is more a twisted, double-edged ribbon than a cylinder of stacked base pairs. The major groove is only one-quarter as wide as the minor one, but is four times as deep (Table 1).

Hydration and contrast with B-DNA

The A-DNA structure itself has been completely refined;

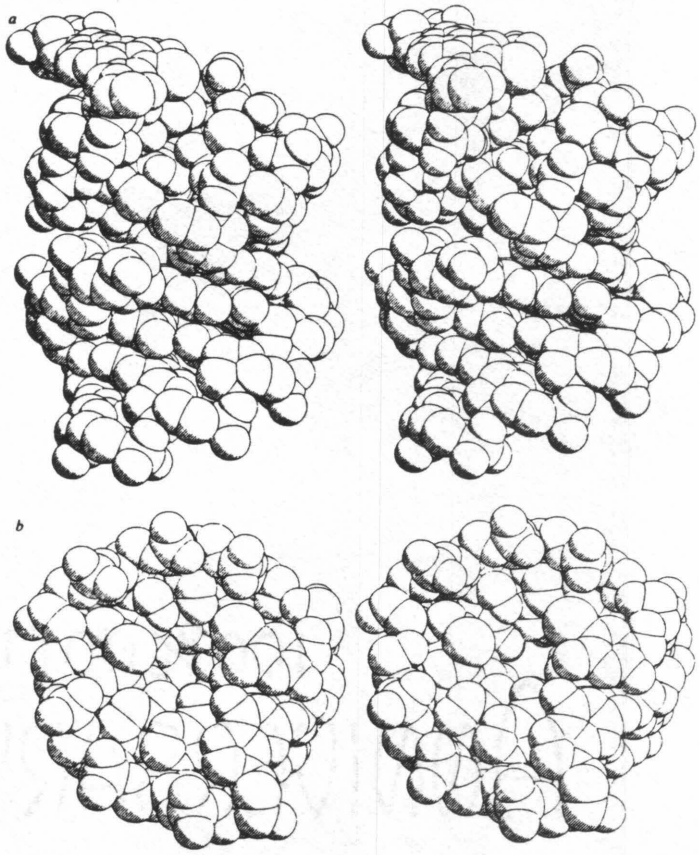

Fig. 3 PLUTO space-filling drawings of three ^{t}CCGG tetramers stacked atop one another along their helix axes to form a complete turn of A-DNA helix. In the actual octamer found in the crystal (Fig. 2), one tetramer is inverted before being stacked on the other, with equivalent base pairs in contact. In contrast, in this simulation of a complete turn of helix, all three tetramers are stacked upright in the sense of Fig. 1, with base pair G4–C5 at the top and C1–G8 at the bottom. No attempt has been made to idealize or regularize individual molecules, or to fill in the missing phosphate groups that would connect tetramers in a continuous helix; each tetramer is exactly as observed in the crystal. *a*, Side view looking into the cavernous major groove at the top, with the shallow minor groove in profile at right. *b*, Top view down the helix axis, showing the 'spiral staircase' arrangement of base pairs and the hollow centre.

remaining refinement will be concerned principally with adding more solvent peaks until the difference electron density map has no significant residual features. The 29 solvent peaks added so far give a good impression of the hydration state of the A helix, since most of the peaks that have yet to be incorporated lie in the open solvent channels between molecules. A detailed analysis of intermolecular water networks, and comparison with other results (refs 25, 26 and M. L. Kopka, A. V. Fratini, H. R. Drew and R.E.D., in preparation) will be described elsewhere (B.N.C. and R.E.D., in preparation). The helical octamers are long enough to establish true major and minor groove solvent environments, even though the chemically unique unit is only four bases long.

The hydration state of A-DNA in CCGG is significantly different from that of B-DNA in CGCGAATTCGCG. In the 35% MPD crystals of B-DNA[26], all the N and O groups on the edges of base pairs in the major groove are hydrated by a monolayer of solvent molecules, but the bulk water within the major groove appears disordered. Similarly, except for three examples where a water molecule is trapped in a clathrate-like manner between a phosphate group and a thymine methyl, hydration of B-DNA along the phosphate backbone also is unlocalized, without clearly defined peaks. But taking the B-DNA dodecamer to the higher alcohol concentration of 60% MPD produces a more extensive second- and third-shell hydration within the major groove, and a more extensive and ordered hydration along the phosphate backbone (M. L. Kopka, A. V. Fratini, H. R. Drew and R.E.D., in preparation). Almost every free phosphate oxygen has a solvent ligand, and another solvent molecule frequently bridges the two oxygens. The most striking and unvarying aspect of B-DNA hydration, however, is a zigzag spine of water molecules running down the narrow minor groove, bridging purine N-3 and pyrimidine O-2 atoms on adjacent base pair steps, where not interrupted by N-2 amino groups on guanines. This feature is identical in low and high MPD crystals, though slightly more extensive in the latter. It has been proposed that the breakup of this minor groove spine of hydration is the crucial first step in the cooperative transition from the B to the A form[26,27].

With respect to the phosphate backbone, CCGG in 80% isopropanol resembles $CGCGAATT^{Br}CGCG$ in 60% MPD: a chain of localized water molecules follows each tetramer backbone strand, bonding to the two unesterified oxygens on each phosphorus. Eighteen well defined solvent peaks are associated with the phosphate backbone, and many of these first hydration shell ligands are themselves bridged by other solvent molecules (see Fig. 4). In the view of the octamer in Fig. 2, a network of six solvent molecules (and their six symmetry-related partners) arches across the opening of the major groove, connecting phosphates P-6 and P-7 on opposite sides. This octamer is so short that the two edges of the major groove are in opposition only at P-6 and P-7. In a continuous A helix, such as in Fig. 3, one would expect the opening of the major groove to be crossed by a continuous network of solvent molecules, especially as the actual A-DNA groove would be 3 Å narrower than that in the octamer of the crystal structure. This network of solvent molecules stitching together the two edges of the major groove may be a distinctive feature of the A helix.

In contrast, fewer ordered water molecules seem to be located deep in the major groove of A-DNA, or in the very shallow minor groove, even though the major groove is more accessible in this crystal structure analysis than it would be in continuous A-DNA. Only one-third of the N and O atoms on base edges have clearly defined full-occupancy ligands, five in the major groove and three in the minor one. Another one-quarter has small nearby peaks that may represent partial occupancy sites (three each in major and minor), but hydration within the grooves is less impressive than in either of the grooves of B-DNA, even at comparable alcohol concentrations.

Water structure and the B-to-A transition

In summary, in A-DNA there is no trace of the minor groove spine of hydration that is so prominent in B-DNA, nor does there appear to be as uniform a monolayer of ordered solvent lining the major groove. The extensive solvation of phosphate backbone observed in B-DNA at 60% MPD reappears in A-DNA, where it seems to lace together the two edges of the major groove.

These observations permit the ideas proposed in ref. 26 on the B-to-A helix transition to be carried one step further. The backbone phosphate groups are more polar, and have a greater attraction for water molecules, than do the N and O groups on the edges of base pairs within the grooves. In conditions where the activity coefficient of water is high (conditions favouring B-DNA), all these polar groups should be solvated; but when the water activity is decreased, either by dehydration or by addition of alcohol, then the groove N and O groups should lose their solvation shells before the phosphates do. When the spine of hydration in the minor groove is disrupted, the integrity of the B structure is destroyed and the helix slips into the A form, in which the still hydrated phosphate groups participate in a network of solvation that laces together the two edges of the now quite narrow major groove. The phosphate-associated lacing across the major groove in A-DNA may have a stabilizing role comparable with that of the base-associated spine within the minor groove in B-DNA. When the water activity around an A helix is increased again, rehydration of base edge N and O groups, including re-establishment of the minor groove spine,

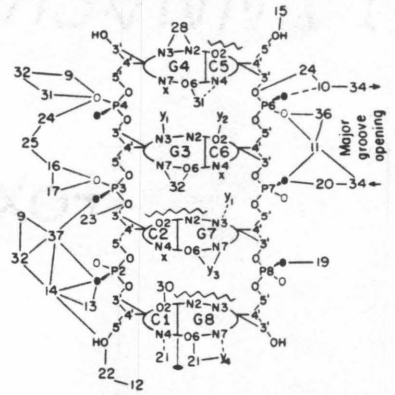

Fig. 4 Schematic drawing of hydration in the immediate vicinity of the CCGG double helix in the crystal. The four central ovals represent purine–pyrimidine base pairs identified by base numbers, as 'G4/C5' at the top. Base ring nitrogen and oxygen atoms are shown by N3, N2, O2 and so on around the perimeter of each oval. The upper edge of each oval faces the minor groove, and the lower edge faces the major. The two sugar–phosphate chains run vertically at either side of the base pair ovals, with C-3', C-4' and C-5' atoms indicated merely by 3', 4' and 5'. (Atoms C-1', C-2' and O-1' are omitted for clarity.) Phosphates are numbered P2–P4 and P6–P8 according to the base to which they are linked via C-5' atoms. Solid black oxygen atoms on phosphates are those facing the major groove. Large numbers identify the most prominent solvent molecules in the immediate vicinity of the double helix. x Indicates that a peak was specifically looked for in the electron density map but not found. y_1–y_4 Are lesser peaks not included in the original hydration search, but found in the map when specifically examined for base edge coordination. ~ Indicates that base edge atoms are blocked by neighbouring DNA substituents, and hence unavailable for hydration. The region of the major groove opening depicted in Fig. 2 appears at the upper right edge. A chain of five solvent molecules (24-10-34-11-36) runs diagonally downwards to the right from phosphate P-6 in Fig. 2, with the parallel symmetry-related chain (36-11-34-10-24) just below it. Solvent molecules 19 and 20 in two symmetry-related pairs lie just below and above these two strands in Fig. 2. The solvent network shown at the left of the diagram is necessarily incomplete and distorted because a three-dimensional network cannot easily be flattened into a planar diagram. Solvent atom 32 at the lower left, for example, is part of a network connecting phosphate P-2 with phosphate P-4 on a different double helix (upper left). It also is coordinated directly to major groove atoms N-7 and O-6 on base G3 of the same helix. However, the figure does provide an accurate depiction of the first hydration shell of the helix.

shifts the equilibrium in favour of B-DNA, because the number of molecules solvating the phosphate backbone is not altered, but only their disposition.

Since submission of this manuscript, O. Kennard and co-workers at Cambridge have solved the structure of another A-DNA helix, of sequence: GGTATACC (personal communication).

Received 27 July; accepted 27 October 1981.

1. Watson, J. D. & Crick, F. H. C. *Nature* **171,** 737–738 (1953).
2. Langridge, R., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. *J. molec. Biol.* **2,** 19–37 (1960).
3. Fuller, W., Wilkins, M. H. F., Wilson, H. R., Hamilton, L. D. & Arnott, S. *J. molec. Biol.* **12,** 60–80 (1965).
4. Arnott, S. in *Organization and Expression of Chromosomes* (eds Allfrey, V. G., Bautz, E. K. F., McCarthy, B. J., Schimke, R. T. & Tissieres, A.) 209–222 (Dahlem Konferenzen, Berlin, 1976).
5. Arnott, S. & Selsing, E. *J. molec. Biol.* **98,** 265–269 (1975).
6. Marvin, D. A., Spencer, M., Wilkins, M. H. F. & Hamilton, L. D. *J. molec. Biol.* **3,** 547–565 (1961).
7. Davies, D. R. & Baldwin, R. L. *J. molec. Biol.* **6,** 251–255 (1963).
8. Itakura, K. & Riggs, A. D. *Science* **209,** 1401–1405 (1980).
9. Wing, R. M. *et al. Nature* **287,** 755–758 (1980).
10. Drew, H. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 2179–2183 (1981).
11. Wang, A. H.-J. *et al. Nature* **282,** 680–686 (1979).
12. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. & Ratliff, R. L. *Nature* **283,** 743–745 (1980).

We thank Al Fratini, Mary Kopka and Horace Drew for permission to quote their dodecamer results in 60% MPD before publication, and Lillian Casler for assistance with the figures. This work was supported by NIH grant GM-12121 and NSF grant PCM79-13959. B.N.C. was the recipient of an NIH predoctoral traineeship. This is contribution no. 6482 from the Division of Chemistry and Chemical Engineering.

13. Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. *Nature* **286,** 567–573 (1980).
14. Hendrickson, W. A. & Teeter, M. M. *Nature* **290,** 107–113 (1981).
15. Seeman, N. C., Day, R. O. & Rich, A. *Nature* **253,** 324–326 (1975).
16. North, A. C. T., Phillips, D. C. & Mathews, F. S. *Acta crystallogr. A24,* 351–359 (1968).
17. Rossman, M. G. *Acta crystallogr.* **14,** 383–388 (1961).
18. Sim, G. A. *Acta crystallogr.* **13,** 511–512 (1960).
19. Sim, G. A. *Acta crystallogr.* **17,** 1072–1073 (1964).
20. Ramachandran, G. N & Raman, S. *Curr. Sci.* **11,** 348–351 (1956).
21. Jack, A. & Levitt, M. *Acta crystallogr.* A**34,** 931–935 (1976).
22. Arnott, S. & Hukins, D. W. L. *Biochem. biophys. Res. Commun.* **47,** 1504–1509 (1972).
23. Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P. & Rich, A. *J. molec. Biol.* **104,** 109–144 (1976).
24. Rosenberg, J. M., Seeman, N. C., Day, R. O. & Rich, A. *J. molec. Biol.* **104,** 145–167 (1976).
25. Neidel, S., Berman, H. M. & Shieh, H. S. *Nature* **288,** 129–133 (1980).
26. Drew, H. R. & Dickerson, R. E. *J. molec. Biol.* **151,** 535–556 (1981).
27. Dickerson, R. E., Drew, H. R. & Conner, B. N. in *Biomolecular Stereodynamics* Vol. 1 (ed. Sarma, R. H.) 1–34 (Adenine, New York, 1981).

Chapter 5. Subsequent Refinement and Hydration

CHAPTER 5. SUBSEQUENT REFINEMENT AND HYDRATION

5.1 Results of Subsequent Refinement

The helix obtained from the anomalously phased electron density map was subjected to 48 cycles of simultaneous energy and reciprocal space refinement,[20] using energy constraints similar to those employed for the B-DNA dodecamer.[21] For the first 45 cycles of refinement, an averaged structure factors defined as, $F_{average} = \frac{1}{2}(F_{hkl}+F_{\overline{hkl}})$ were used for refinement. During the last three cycles, F_{normal}, the sum of the normal scattering from all atoms in the cell,[22] was used for refinement. This quantity may be calculated as

$$F_{normal} = \sqrt{\frac{1}{2}[F^2_{hkl} + F^2_{\overline{hkl}}] - F^2_{anom}} \qquad (5.1)$$

This change in data sets resulted in only a 0.3% change in R factor since the R factor between the $F_{average}$ and the F_{normal} was less than 1%. The path of refinement can be followed on the R factor curves in Figure 1.

During the refinement process the R factor for 2σ data fell from 39% at 3.0 Å resolution to the current 16.5% for reflections between 8Å and 2.1 Å resolution. The R factor for all data between 8Å and 2.1 Å resolution is 19.9%. The asymmetric unit contains two tetramer strands (one double helix) with 86 water molecules positioned in the final map. The RMS error in atomic position, as estimated by the method of Luzzati,[23] is 0.25 Å as shown in Figure 2. Structure factors and atomic coordinates are available from the Brookhaven Protein Data Bank.[24] The atomic coordinates from the final cycle of refinement may be found in the Appendix of this thesis.

---

20. A. Jack and M. Levitt (1976). *Acta Cryst.* **A34**, 931-935
21. H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson (1981). *Proc. Nat. Acad. Sci. U.S.A.* **78**, 2179-2183
22. G. N. Ramachandran and S. Raman (1956). *Curr. Sci.* **11**, 348-351
23. V. Luzzati (1952). *Acta Cryst.* **5**, 802-810
24. F. C. Berstein, Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973

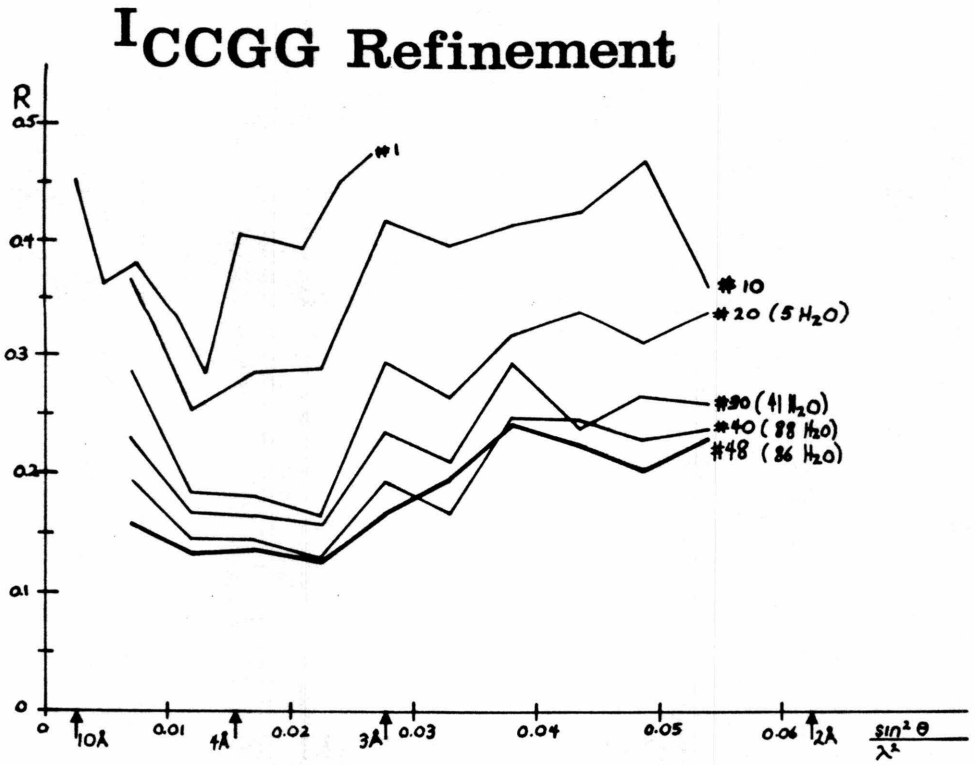

Figure 1. Crystallographic R factor as a function of resolution and refinement cycle.

5.2 The Structure of ^{I}CCGG

The helix parameters of ^{I}CCGG shown in Table 1 were calculated with a program provided by John Rosenberg. In a short piece of DNA there is some ambiguity as to what constitutes the "best" helix. Rosenberg's program calculates helical parameters by determining the best helix to move a given vector from one position to its helix related position. The values in Table 1 were calculated using a vector that pointed from the sugar C1' to the base nitrogen, N9 or N1. The parameters may vary slightly depending on the set of atoms used to define the vector.

The helix parameters of ^{I}CCGG show a regular helix of A-DNA with a rise/residue of 2.2 Å and 10.6 base pairs/turn. All sugars, except the terminal G4 sugar, adopt the C3'-endo conformation. Torsion angles of the refined molecule are shown in Table 2. Although these torsion angle values are clumped together in relatively tight groups, they do differ significantly from the fiber A-DNA values[25] shown at the

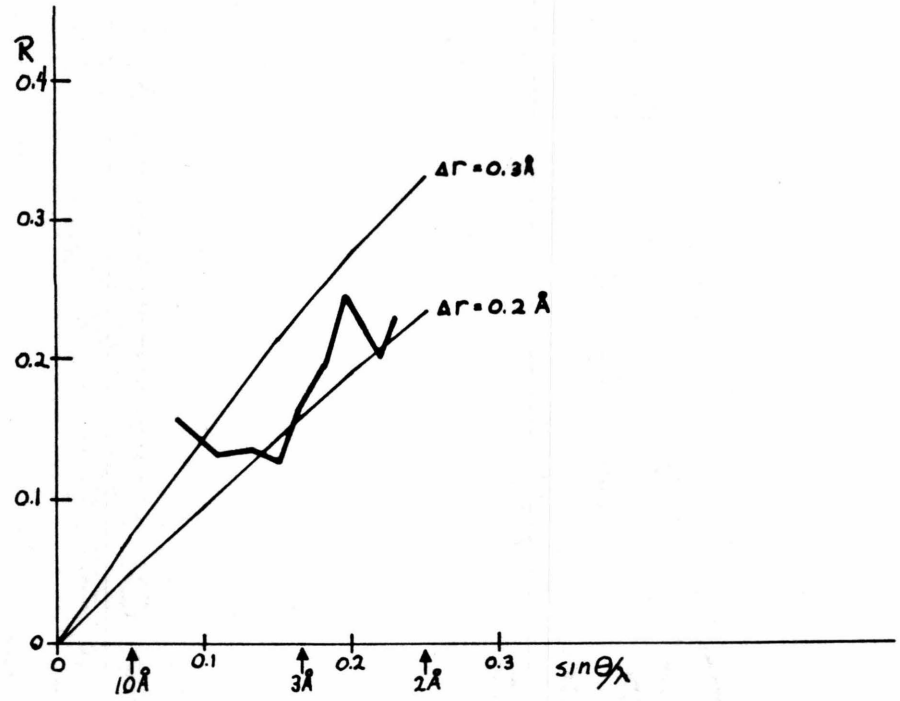

Figure 2. Luzzati estimate of RMS error in atomic position for ^{I}CCGG.

bottom of the table.

As shown in the cell packing diagram of Figure 3, , the ^{I}CCGG tetramer fits into the $P4_32_12$ unit cell with two single strands (one double helix) as the asymmetric unit. The crystallographic 4_3 axis passes near the G4 sugar and causes the bottom of G4/C5 base pair to fit into the flat minor groove of the adjacent helix in a manner similar to that described by Wang *et al.* for GGCCGGCC.[26] A better view of the close contacts caused by packing the bottom and minor groove side of residues G4, C5, and C6 into the minor groove side of C2, G3, G4, G7, and G8 on the 4_3 related molecule is shown in Figure 4. However, none of these contacts seem to be ionic or hydrogen bonding interactions. In this area, the molecules must be held together by van der Waals interactions.

---

25. S. Arnott and D. W. L. Hukins (1972). *Biochem. Biophys. Res. Com.* **47**, 1504-1509
26. A. H.-J. Wang, S. Fujii, J. van Boom and A. Rich (1982). submitted - *Proc. Nat. Acad. Sci. USA*

Helix Parameters of ^{I}CCGG			
Helix Step	bp/turn	deg/turn	rise/res
Single Step			
2 to 1	9.9	36.2°	3.2Å
3 to 2	11.8	30.6°	2.9Å
4 to 3	10.1	35.7°	3.3Å
Ave. Single Step	10.6	34.0°	2.2Å
Double Steps			
3 to 1	10.8	33.3°	3.0Å
4 to 2	11.0	32.7°	3.1Å
Ave. Double Step	10.6	34.1°	2.5Å
First to Last	10.8	33.4°	3.2Å
Octamer Parameters			
Ave. Single Step	10.4	34.6°	2.8Å
Step Btwn Helices	9.9	36.2°	2.7Å

TABLE 1. Helix parameters from ^{I}CCGG

Near the 4_3 axis, the crystal packing is very tight and some distortions of the helix occur. The $\chi-\delta$ plot in Figure 5 shows that while seven of the sugars have the 3'-endo conformation, residue G4, near the 4_3 axis, has the 2'-endo conformation. This is not significant, however, since G4 has no continuation of the helix to keep it in the regular C3'-endo conformation. Also, residue C5 has been shoved underneath residue C6. Figure 6 shows that the helix axis is displaced toward the cytosines.

Torsion Angles of ^{I}CCGG								
Residue	α	β	γ	δ	ε	ζ		χ
1 ^{I}C			60	76	-157	-69		-169
2 C	-70	174	57	77	-151	-65		-162
3 G	-73	166	70	73	-170	-72		-171
4 G	-72	-174	79	150				-127
5 ^{I}C			143	90	-168	-58		-150
6 C	-68	-170	51	80	-162	-68		-159
7 G	-80	-178	72	87	-160	-67		-164
8 G	-75	-176	55	80				-152
A_F	-90	-149	47	83	-175	-45		-154
B_F	-41	136	38	139	-133	-157		-102

TABLE 2. Torsion angles of iodo-CCGG

This is caused by the G4/C5 base pair being pushed underneath the G3/C6 base pair. The movement of C5 underneath C6 causes some shift in the phosphate and thus an increase in the propeller twist of the G3/C6 pair. The base pair propeller twists are 15°, 12°, 22°, and 17° for C1/G8, C2/G7, G3/C6 and G4/C5 respectively.

At the other end of this double helix is a crystallographic twofold axis which is positioned perpendicular to the helix axis in such a way as to stack two tetramers on top of each other to make a continuous octamer of A-DNA. This interaction can be seen in the cell packing diagram of Figure 3, but a clearer view is shown in Figure 7. The twist between base pair C1/G8 and its twofold related base pair is 36.2° with a rise/residue of 2.7 Å. This is comparable with the helix rotations for other steps shown in Table 1. The smoothness of the interface between the two tetramers in the octamer may also be seen by looking at the torsion angles in Table 2. The torsion angles γ of residue C1 and δ of residue G8 near the twofold axis are comparable to those belonging to base pairs C2/G7 and G3/C6. However, the torsion angles δ of G4 and γ of C5 at the other end of the molecule, near the tightly packed 4_3 axis, take values different from the angles of the other base pairs in the tetramer. Of course, this situation emphasizes the distortion of the helix around the 4_3 end of ^{I}CCGG but also shows that the interface at the 2_1 end of the helix is very smooth. In fact, these

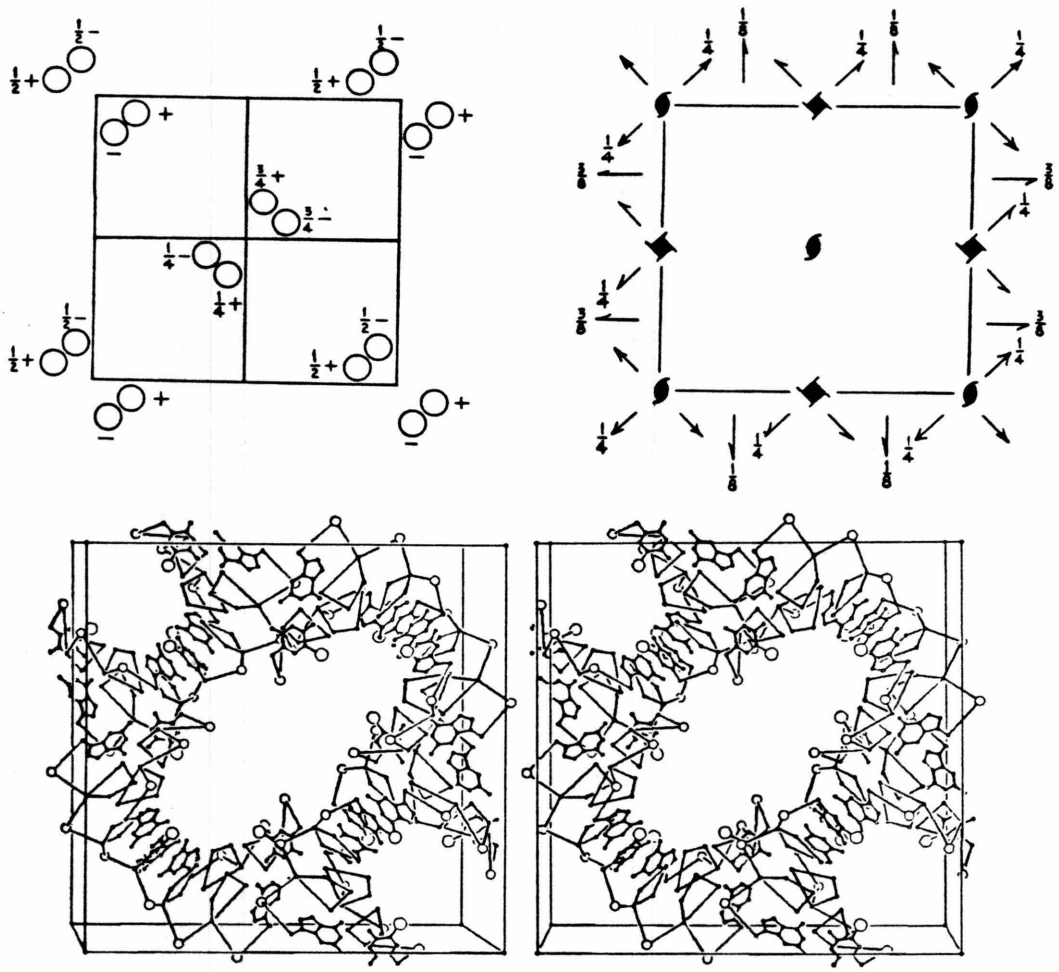

Figure 3. Crystal packing in ^{I}CCGG.

two stacked tetramers may be considered to be one octamer with single phosphate atom per chain missing between them.

5.3 Water Around ^{I}CCGG

The electron density maps show ^{I}CCGG surrounded by water molecules. Figure 8 shows the 47 solvent molecules within 3.5 Å of the tetramer.

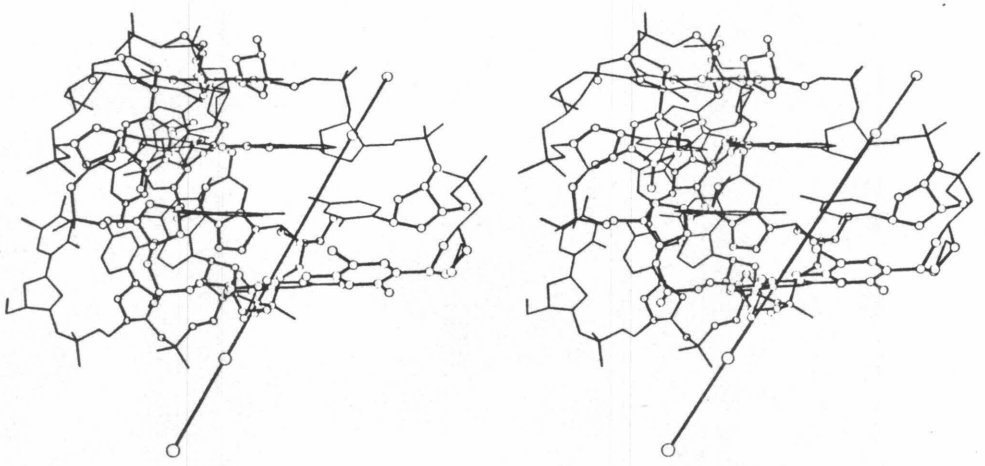

Figure 4. Interactions about the 4_3 axis of ^{I}CCGG. Distances closer than 5Å are marked with enlarged atoms.

Three of these molecules, W19, W31, and W86, appear in the figure in more than one place because of the symmetry of the unit cell. Such waters are marked with an apostrophe the second time they appear, e.g. W19 becomes W19'. First layer water contacts and distances are shown in Tables 3 and 4.

The distance of 3.5Å was chosen for hydration analysis because it represents the outer bound of hydrogen bonding distance. It is probably fair to say that water at distances greater than 3.5 Å represents an interaction but not a true hydrogen bond. Water interactions have been studied in some of the earlier high resolution studies: the RNA sequences ApApA,[27] and GpC,[28] and the DNA sequence pTpT.[29] Since these structures all have resolution limits near 1Å and crystallographic R factors

27. D. Suck, P. C. Manor and W. Saenger (1976). *Acta Cryst.* **B 32**, 1727-1737
28. B. Hingerty, E. Subramanian, S. D. Stellman, T. Sato, S. B. Broyde and R. Landgridge *Acta Cryst.* **B 32**, 2998-3013
29. N. Camerman, J. K. Fawcett and A. Camerman (1976). *J. Mol. Biol.* **107**, 601-621

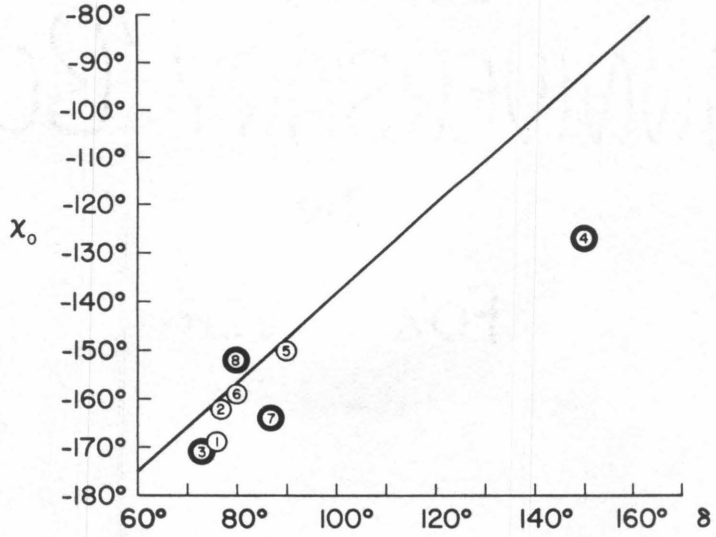

Figure 5. Plot of backbone torsional angles χ vs. δ showing the clustering of sugar values about the C3'-endo conformation.

near or below 10%, their solvent to DNA distances are very accurate. They have water to water and water to phosphate oxygen distances near 2.8Å, but water to O3' or O5' distances of 3.0Å. Taking the number of 3.0Å as the longest average bond distance and adding twice the Luzzati estimated RMS error in atomic distance, gives the 3.5Å limit used in this study.

When comparing the major and minor groove views of the hydrated tetramer in Figure 8, one immediately sees that the major groove and sugar phosphate backbone are heavily hydrated while the minor groove is almost dry. One reason for the dryness of the minor groove is the crystal packing about the 4_3 axis that was described earlier. The bottom base pair of one molecule effectively fills most of the minor groove of its 4_3 related molecule.

Even so, that part of the minor groove which is not covered by the symmetry related molecule still has only a small number of waters in it. All of the waters in the minor groove are attached to guanine 4 or to oxygens in the sugar phosphate backbone. In ^{I}CCGG, cytosines 1 and 6 are not blocked by symmetry related helices, and yet still have no hydration. The minor groove G7 and G8 is blocked by the 4_3 related molecule, but G3 and G4 are free. One water in the minor groove near G3 is pulled away from the base by an interaction with the 4_3 related molecule. Around G4 are several water molecules which attach to N2 and N3 of the base. Several waters also are attached to the sugar phosphate backbone in the minor groove. However, these waters are usually stabilized by attachments to other waters or to a crystallographically related double helix. In general, those parts of the sugar phosphate backbone accessible from the minor groove are not heavily hydrated.

In contrast to the minor groove, the 12 waters in the major groove occupy every possible hydrogen bonding site. This major groove hydration pattern has approximate twofold symmetry reflecting the approximate twofold symmetry of the DNA itself. The overall impression is that of a meshwork of water molecules stretched over the form of the A-DNA to cover the surface of the major groove entirely with water.

The use of 5-iodo-cytosine in place of cytosine in this crystal resulted in two slight alterations of the hydration pattern. First, each iodine has one water attached to it about 3.4Å directly along a line drawn from the base through the iodine. This distance is close to the normal Van der Waals iodine to oxygen distance of 3.55Å. However, the water's alignment with the iodine indicates that some interaction is taking place. This may be a charge transfer bond involving donation of the lone pair electrons of the oxygen to the empty 5d orbitals on the iodine. Such an interaction has been reported in the high resolution crystal structure of 5-iodo-2'-deoxyuridine.[30]

30. N. Camerman and J. Trotter (1965). *Acta Cryst.* **19**, 203-211

Second, the iodine also serves to prevent water from binding to the cytosine 4-amino group. Those cytosines without iodines, cytosine 2 and 6, have waters attached to their 4-amino groups.

Both the N7 and the O6 of guanine are hydrated, although the positioning of the waters is not the same in all cases. One particularly interesting case is that of guanine 4, where both water 31 and water 33 bridge to other moieties in the helix. Water 31 connects O6 of guanine 4 with O6 of guanine 3, above it. Water 33 links together the guanine O6 with one of the phosphate oxygens. It should be emphasized that this guanine is not in the typical A-DNA conformation, but more in the B form. Thus, the positioning of the backbone which allows a single water to link a base with a phosphate oxygen is not found on any of the other guanines.

The backbone is the most heavily hydrated portion of the structure. It contains 30 first layer waters of hydration, in contrast to the 12 waters found in the major groove and the 7 waters of the minor groove. Since the O1' oxygens in the deoxyribose rings are only accessible from the minor groove, they have no major groove waters. However, there are minor groove waters attached to two of them. The O3' and O5' oxygens are accessible from the major groove or from the outside of the helix. Several waters are close to these oxygens, but in most cases this proximity is probably due to simple backbone geometry rather than to hydrogen bonding. One example of this may be seen in water 37, which bridges between the charged phosphate oxygens on cytosine 2 and those on guanine 3. Here the backbone geometry pushes the sugar O5' toward water 37, which is firmly held in place by three hydrogen bonds. Two of these hydrogen bonds are to charged phosphate oxygens and the third is to a guanine bound water. In the middle of the chain, O5' contains no hydrogen bonds except those characterized above as being due to backbone geometry.

On the other hand, there are some waters which do appear to be bound primarily to O3' or O5'. All of the free ends, for example, contain at least one and usually two bound waters. The O3' oxygen often has attached waters which are usually toward the minor groove side of the helix.

Along the major groove side of the backbone of residues 1 through 4 is a chain of waters containing 1O5'-W14-2OM-W37-3OM-W69-W33-4OM. (Phosphate oxygen OM is on the major groove side of the backbone while oxygen Om is on the minor groove side.) Such a chain is not found on the other side of the helix.

It is possible, though probably somewhat meaningless, to continue analysis of the water structure for many levels. However, it is instructive to consider the second layer of hydration. I have defined a "second layer water" as water which touches at least two other first layer waters. This distinguishes them from those which touch only a single first layer water and are therefore probably less tightly bound. The twenty waters involved in this layer are shown in Figure 9 (along with the first layer, of course).

The contacts and distances of these waters are listed in Table 5. Comparison of Figures 8 and 9, which show first and second layers of hydration respectively, illustrates how the second layer begins to knit together some of the more isolated sections of the first layer. A good example is W45 which links together W60 on Guanine 7, W58 on Guanine 8 and W20 attached to a phosphate oxygen of residue 7. A similar case is found in W48 which links together waters from cytosine 5, cytosine 6 and phosphate oxygens from residue 6. These two examples show how the water layers onto the DNA; first by attaching directly to the charged and polar atoms of the DNA and then by filling in between appropriately placed first layer waters.

In summary it can be said that no major water structure, such as the "spine" in the minor groove of AT rich B-DNA,[31] is found in this structure of A-DNA. Instead, water binds in a more haphazard fashion which appears to cover the molecule rather than shape it. All three types of water bridges appear: base to base, phosphate to phosphate, and base to phosphate. However, none of these forms a systematic network. The overall impression of hydration for this fragment of A-DNA is fortuitous hydration, heavy along the backbone and in the major groove, but light in the minor groove.

---

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

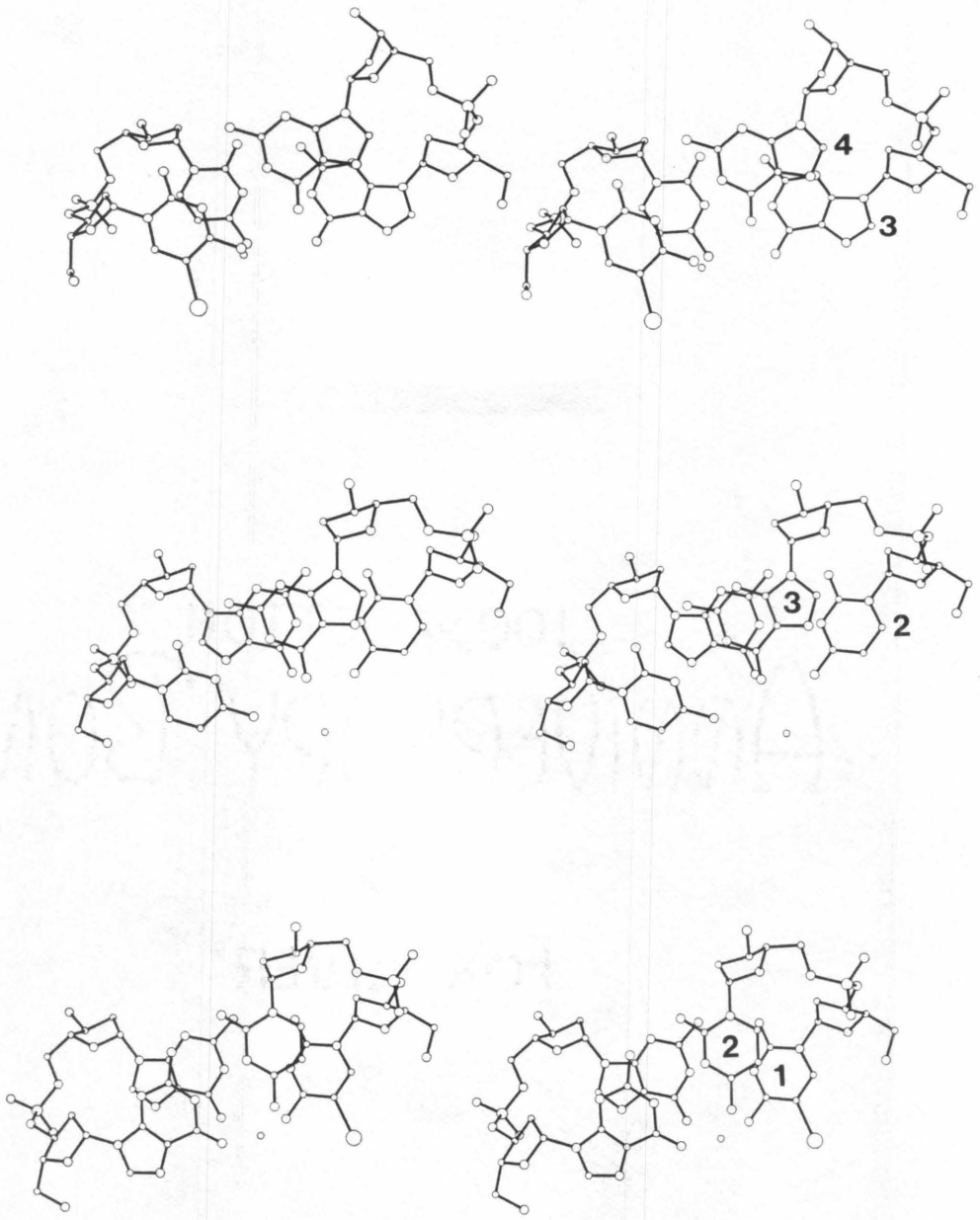

Figure 6. Stereo drawings of each base pair of ^{I}CCGG viewed down the local helix axis.The position of the helix axis is shown as a dot in each drawing.

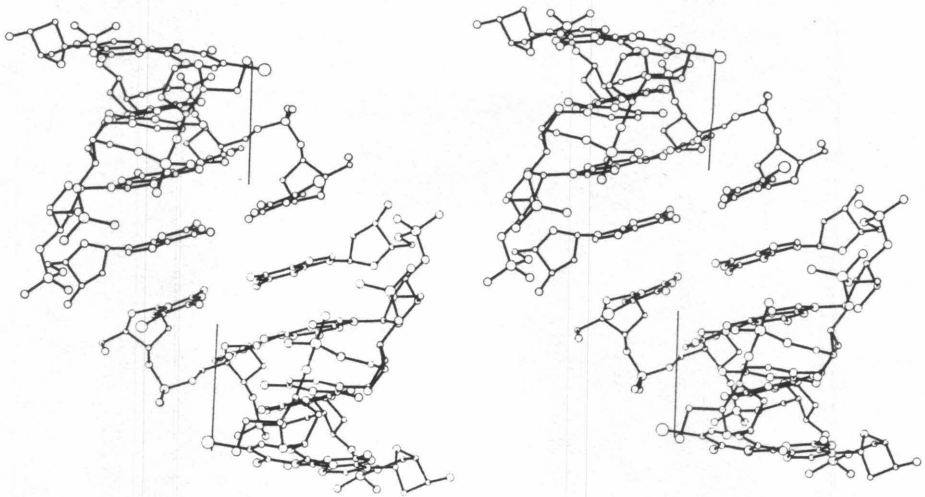

Figure 7. ^{I}CCGG tetramers related by crystallographic twofold.

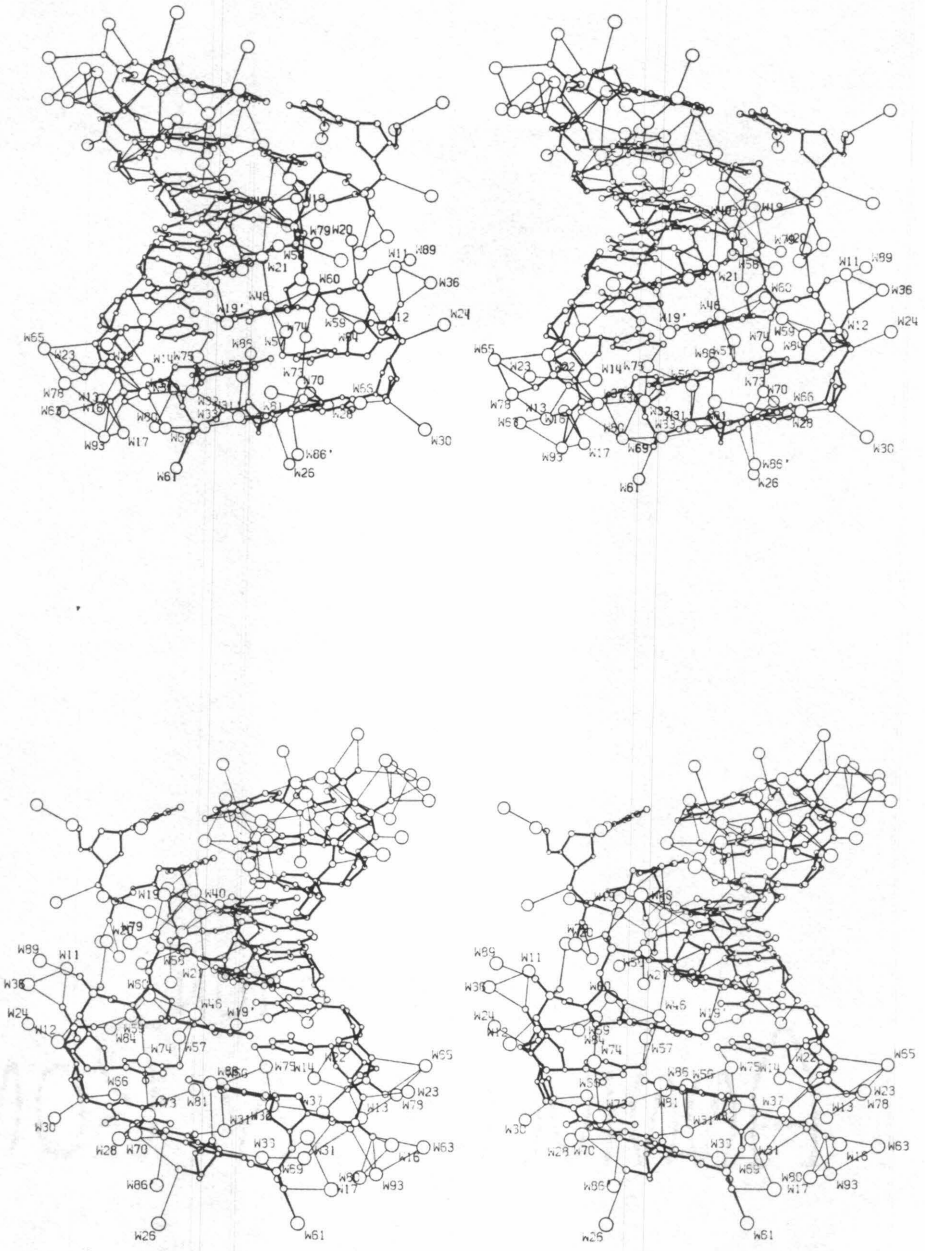

Figure 8. First layer waters within 3.5Å of ^{I}CCGG

First Layer Waters of ^{I}CCGG				
Water Number	DNA Contacts	DNA Distances	Other Contacts	Other Distances
W11	6OL	2.6Å	W36	2.7Å
W12	7O3'	3.4Å		
W13	2OR	3.0Å	W78	2.8Å
W14	1O5' 2OR	3.2Å 2.8Å		
W16	3OL	3.1Å	W17	3.2Å
W17	3OL	2.8Å	W16 W93	3.2Å 2.6Å
W19	8OR	3.1Å		
W19'	I1	3.4Å		
W20	7OR	2.9Å		
W21	8O6	2.8Å	W46	3.4Å
W22	1O5'	3.2Å		
W23	3O3'	3.1Å		
W24	6O3'	3.2Å		
W26	4O3T	3.4Å		
W28	4N2 4N3	3.4Å 3.4Å	W70 W73	3.3Å 2.4Å
W30	5O5'	3.2Å		
W31	3O6 4O6	2.9Å 2.6Å	W32 W56	3.5Å 3.3Å
W31'	4O3'	3.2Å	W80	3.2Å
W32	3N7	2.9Å	W31 W37 W75	3.5Å 3.3Å 2.9Å
W33	4N7 4OR	2.7Å 3.0Å	W69	2.7Å
W36	6OL	2.8Å	W11	2.7Å
W37	2OR 2O5' 3OR	2.8Å 3.3Å 3.2Å	W32	3.3Å
W40	8OL 8O3T	3.4Å 3.5Å		
W46	6N4 7O6 7N7	3.4Å 2.3Å 3.4Å	W21 W57 W60	3.4Å 2.7Å 3.5Å

TABLE 3. Water within 3.5Å of ^{I}CCGG *(First of two parts)*

First Layer Waters of ^{I}CCGG (Continued)				
Water Number	DNA Contacts	DNA Distances	Other Contacts	Other Distances
W56	3O6	3.0Å	W31 W75	3.3Å 3.1Å
W57	6N4	3.0Å	W46	2.7Å
W58	8N7	2.5Å		
W59	6OR	2.6Å	W60 W84	2.6Å 3.4Å
W60	7N7	2.8Å	W46 W59	3.5Å 2.6Å
W61	4OL 4OR	2.7Å 3.0Å		
W63	3OL	3.4Å	W93	3.3Å
W65	2O3' 2OL	3.5Å 3.2Å	W78	3.4Å
W66	5O5'	3.0Å		
W69	3OR 3O5'	3.4Å 3.4Å	W33	2.7Å
W70	4O3T	3.3Å	W28 W73	3.3Å 2.6Å
W73	4N3	3.4Å	W28 W70 W74	2.4Å 2.6Å 3.3Å
W74	7O1'	3.3Å	W73	3.3Å
W75	2N4	2.6Å	W32 W56	2.9Å 3.1Å
W78	2OL	3.0Å	W13 W65	2.8Å 3.4Å
W79	8OL	2.5Å		
W80	4OL	3.5Å	W31	3.2Å
W81	5I	3.2Å		
W84	6OR	2.7Å	W59	3.4Å
W86	4O1'	3.4Å		
W86'	4N3	3.3Å		
W89	7OL	2.9Å		
W93	3OL 3OR	2.6Å 3.2Å	W17 W63	2.6Å 3.3Å

TABLE 4. Water within 3.5Å of ^{I}CCGG *(Second of two parts)*

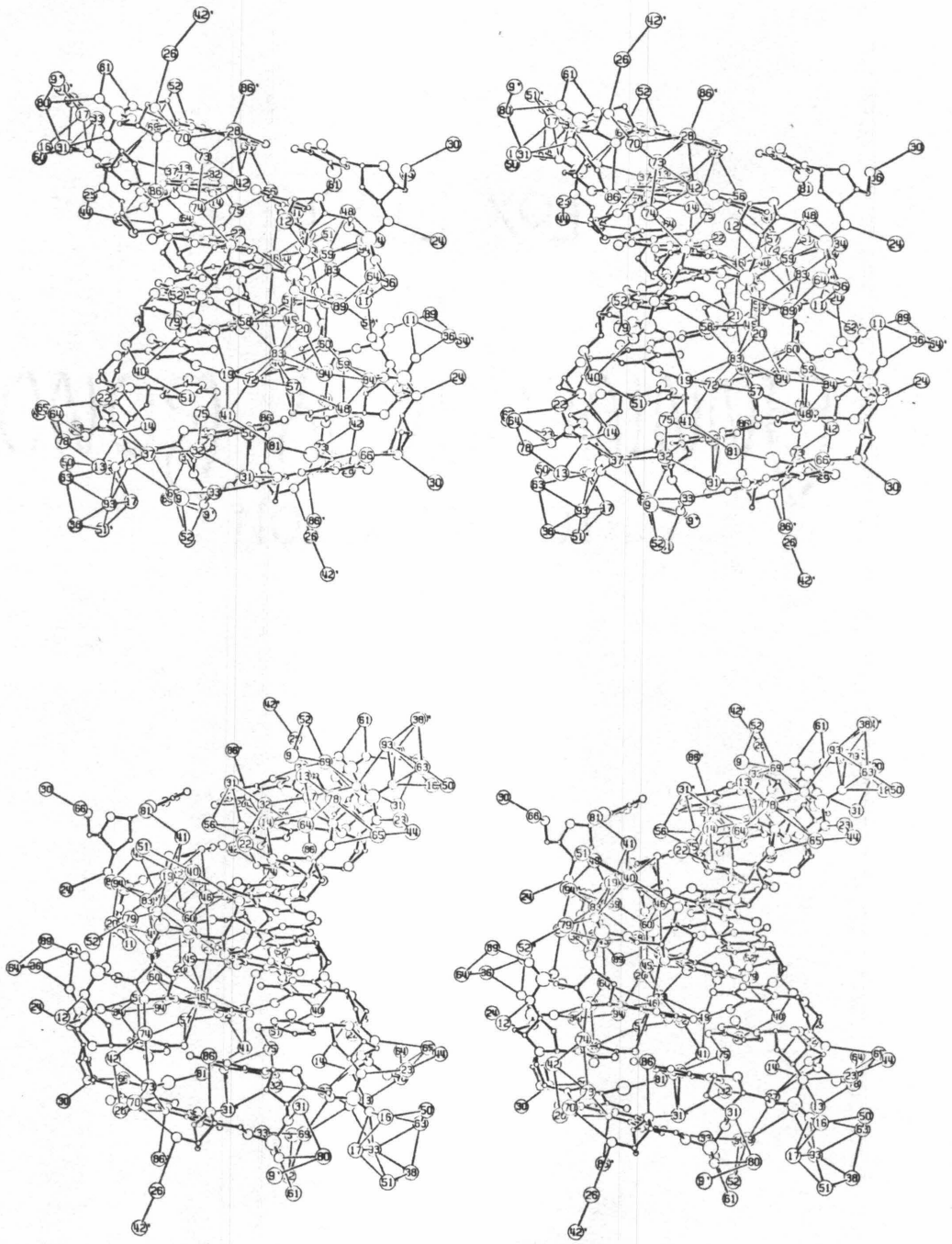

Figure 9. Second layer waters within 3.5Å of two first layer waters

Second Layer Waters of ^{I}CCGG				
Water Number	First Layer Contacts	First Layer Distances	Second Layer Contacts	Second Layer Distances
W9	W69	2.5Å	W52	2.9Å
W9'	W80	3.3Å		
W38	W63 W93	3.1Å 2.7Å	W51	2.7Å
W41	W19 W72 W81	2.8Å 2.8Å 3.4Å		
W42	W28 W73 W74	3.2Å 3.0Å 3.1Å		
W42'	W26	3.0Å		
W44	W23 W65	2.9Å 2.6Å		
W45	W20 W58 W60	3.0Å 2.8Å 2.8Å	W83	2.6Å
W48	W57 W59 W84	3.4Å 3.3Å 2.6Å		
W50	W16 W63	3.1Å 2.5Å		
W51	W40	3.5Å		
W51'	W17 W93	2.7Å 3.1Å	W38	2.7Å
W52	W69	3.0Å	W9	2.9Å
W52'	W79	3.1Å		
W64	W78	3.1Å		
W64'	W36 W89	2.8Å 2.7Å		
W72	W19 W57	2.5Å 3.4Å	W41	2.8Å
W83	W19 W20 W58	3.4Å 2.7Å 3.4Å	W45 W94'	2.6Å 3.1Å
W94	W84	3.2Å		
W94'	W20	3.4Å	W83	3.1Å

TABLE 5. Second layer waters of ^{I}CCGG and their contacts

Chapter 6. CONCLUSIONS

Chapter 6 Conclusions

6.1 Structural Analysis

6.1.1 Comparison of ^{I}CCGG with fiber studies:

All discussions of DNA structure must consider the foundation laid by x-ray diffraction of DNA fibers. For A-DNA, the pertinent structure is the linked atom least square refinement by Arnott and his coworkers.[1] This structure is a regular right-handed helix with a rise per residue of 2.6 angstroms and a twist of 32.7°. The bases are paired in Watson Crick fashion with a propellor twist of -12°. As mentioned earlier, the regularity of the structure is an inevitable consequence of the fiber diffraction method. Variations in the structure are averaged out by sequence heterogeneity, rotational disorder, and translational disorder of the DNA in the fibers. Nevertheless, the fiber patterns from A-DNA are the most crystalline of all DNA fibers.[2] It is not surprising, therefore, to find that the structural parameters of A-DNA derived from the fiber are close to those found by this single crystal study.

We find the base plane inclination in ^{I}CCGG to be increased to 23° from the fiber value of 19°, and the average rise per residue decreased from 2.6Å to 2.2Å. As discussed in Chapter 5, these differences may be due in part to the difficulties associated with trying to extract helix parameters from a short piece of DNA. ^{I}CCGG confirms the 3' endo sugar pucker that was found in the A-DNA fiber structure. The only sugar not in this conformation is residue 4, whose conformation is not constrained by a following residue. The major difference in A-DNA helix parameters found between fiber and crystal is the propellor twist of the GC base pair. Arnott *et al.* report a value of -12° while we find an average value of +17°. The fact that a difference was also found between fiber and single crystal propellor twist values for

---

1. S. Arnott and D. W. L. Hukins (1972). *Biochem. Biophys. Res. Com.* **47**, 1504-1509
2. A. G. W. Leslie, S. Arnott, R. Chandrasekaran and R. L. Ratliff (1980). *J. Mol. Biol* **143**, 49-72

B-DNA[3] indicates that fiber patterns are not very sensitive to this parameter.

The recent solution of the A-DNA octamer, GGTATACC,[4] confirms the positive propellor twist discovered in ^{I}CCGG. The octamer structure was solved by rotation methods using fiber A-DNA coordinates as a model, so the initial propellor twist of base pairs was negative. However, subsequent refinement of the structure revealed that the twist is indeed positive.[5]

6.1.2 Other single crystal A-DNA structures:

Two other single crystal A-DNA structures, GGTATACC[6] and GGCCGGCC,[7] have been solved. GGTATACC packs into a $P6_1$ unit cell with two octamer chains, or one double helix, as the asymmetric unit. GGCCGGCC packs into a $P4_32_12$ unit cell in the same manner as ^{I}CCGG, and so has only one strand, or half a double helix as the asymmetric unit. These structures are both at 2.2Å resolution, comparable to ^{I}CCGG's 2.1Å resolution. Currently, only preliminary information is available for the AT containing sequence of Shakked *et al.*, but some information from a refined structure is available for the GGCCGGCC sequence of Wang *et al.*

6.1.3 Major groove opening at purine pyrimidine sequence:

One of the few structural details reported about the Shakked structure is that the major groove opens up by bending the helix axis at the molecular twofold located at ApT, a purine-pyrimidine sequence. This same effect can be seen about the crystallographic twofold in ^{I}CCGG. as shown in Figure 1. This opening of the major groove at a purine-pyrimidine sequence has also been seen in B-DNA.[8] In

3. R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson (1980). *Nature* **287**, 755-758
4. Z. Shakked, D. Rabinovich, W. B. T. Cruse, E. Egert, O. Kennard, G. Sala, S. A. Salisbury and M. A. Viswamitra (1982). *Proc. R. Soc. Lond.* **B 213**, 479-487
5. O. Kennard - personal communication
6. Z. Shakked *et al.* (1982). - ibid
7. A. H.-J. Wang, S. Fujii, J. van Boom and A. Rich (1982). submitted - *Proc. Nat. Acad. Sci. USA*
8. R. E. Dickerson and H. R. Drew (1981). *Proc. Nat. Acad. Sci. USA* **78**, 7318-7322

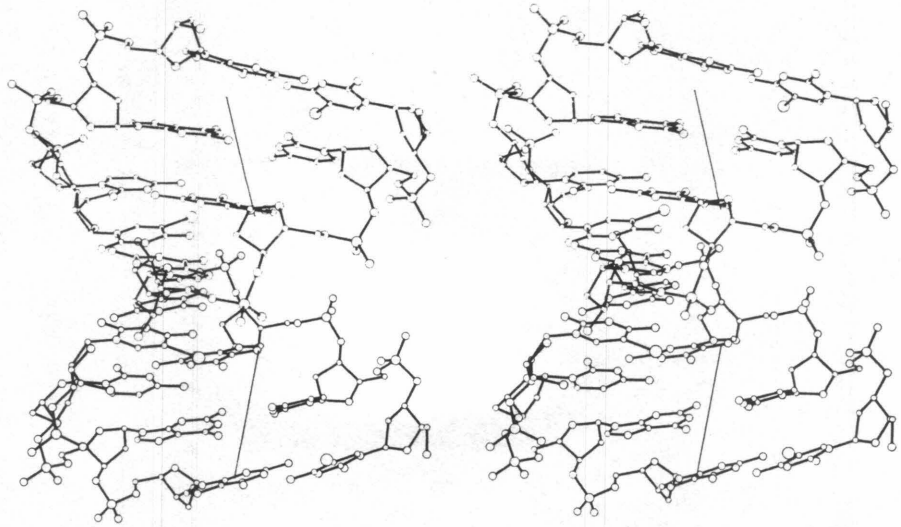

Figure 1. Stereo view of the octamer of ^{I}CCGG

^{I}CCGG the crystallographic twofold runs between a G(p)C sequence where the phosphate is missing. This opening is probably due to phosphate - phosphate repulsion across the narrow major groove of these short A-DNA helices.

6.1.4 Packing of A-DNA:

It is interesting to note that all three of these A-DNA structures have the same packing interactions about their respective screw axes. In each case, the base pair at the end of one helix fits into the dry minor groove of a screw-related helix. as shown in Figure 4 of Chapter 5. This interaction occurs without stabilization from any hydrogen bonds between the two helices. This situation can be contrasted with that of the B-DNA dodecamer, where an interaction between the minor groove of two helices occurs, but is stabilized by hydrogen binding between the involved bases. This A-DNA packing scheme indicates that the minor groove of A-DNA is relatively hydrophobic. Such hydrophobicity may be an important key for molecules, such as proteins or drugs that recognize A-DNA.

Three different types of intermolecular interactions between DNA have been seen in these crystal structures:

1. end-to-end stacking (A, B, and Z DNA)
2. minor groove hydrophobic packing (A DNA)
3. minor groove H-bonds (B DNA)

6.1.5 Regularity of ^{I}CCGG:

The general regularity of ^{I}CCGG can be seen in table of torsion angles in Chapter 5. However, it is better illustrated by Figure 2, a cylindrical polar plot of the tetramer's backbone.

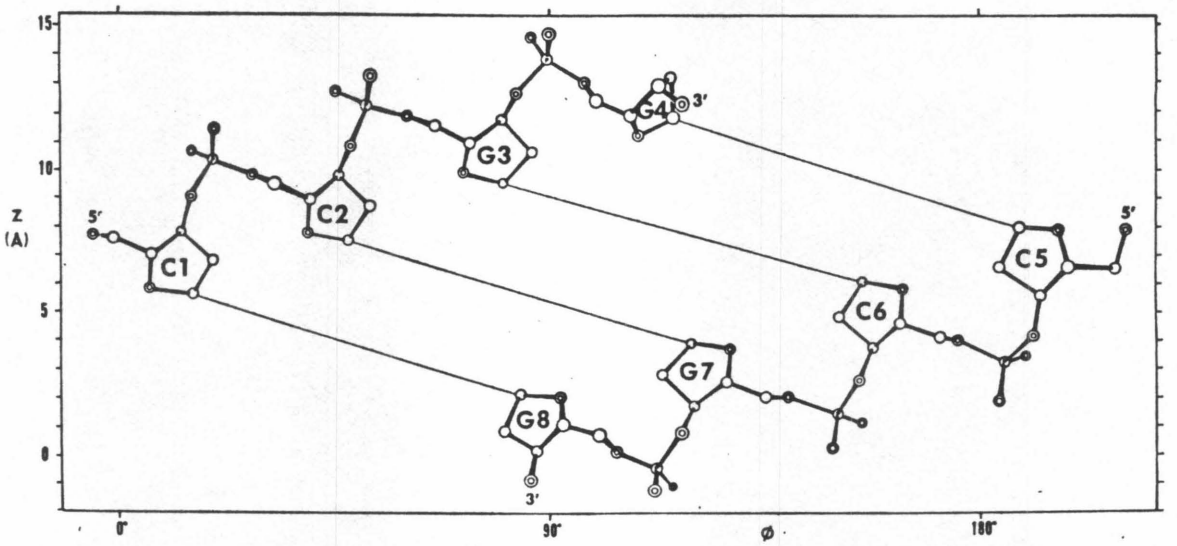

Figure 2. Cylindrical polar plot of ^{I}CCGG

Although there is some small variation along the backbone, as evidenced by changes in phosphate oxygen positions, the overall impression is one of regularity. The distortion due to helix packing is easily seen in this figure. The C1/G8 base pair has its free O3' and O5' in positions similar to the second and third base pair. However, the

G4/C5 base pair has its free hydroxy groups in radically different positions.

Wang's GGCCGGCC was analyzed at both -18°C and -8°C. The -8° analysis shows an A-DNA helix with alternating sugar-phosphate backbone. These alterations in sugar geometry are visible in the χ - δ plot of Figure 3.

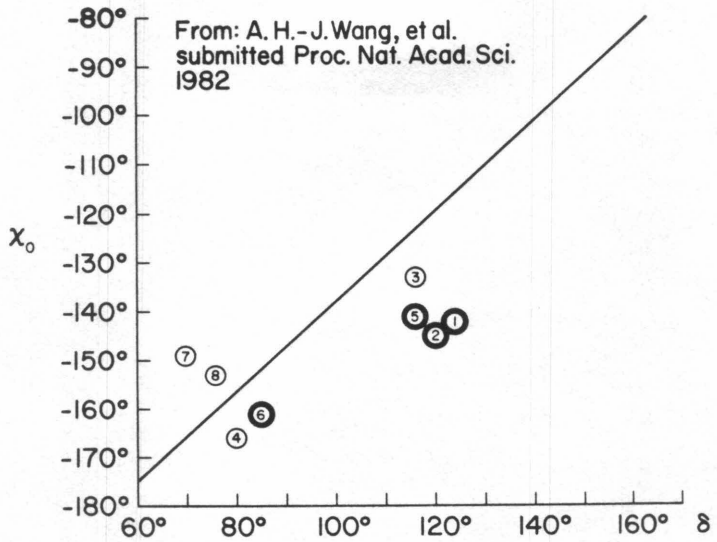

Figure 3. χ vs. δ torsion angles of GGCCGGCC by Wang *et al.*

For comparison, the χ - δ plot of ^{I}CCGG is shown in Figure 4. The sugar puckers of the GGCCGGCC fall into two groups, C3' endo and C1' exo, while the χ and δ angles of ^{I}CCGG all lie in the C3'-endo region. This variation in backbone conformation is only about half the variation seen in the B-DNA dodecamer,[9] but the two groups are distinct enough that this must be considered a real effect. Why is this sequence variation visible in the octamer but not in the tetramer, ^{I}CCGG? The answer, as discussed below, may lie in the lengths of the molecules.

9. H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson (1981). *Proc. Nat. Acad. Sci. USA* **78**, 2179-2183

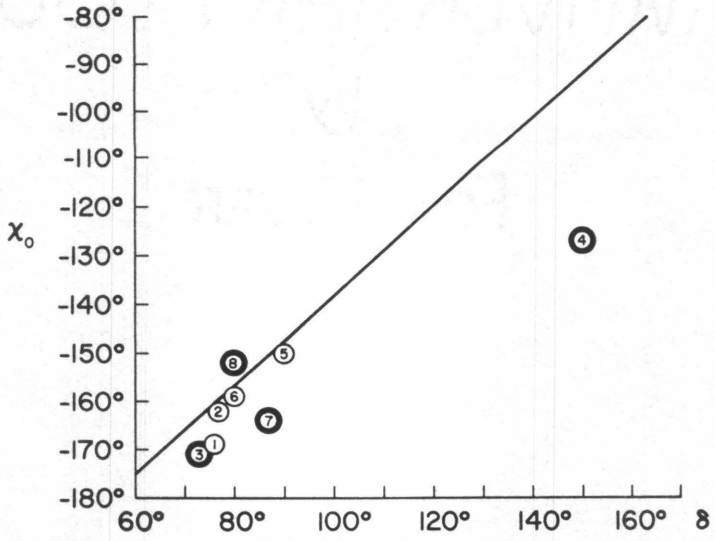

Figure 4. χ vs. δ torsion angles of ^{I}CCGG

6.1.6 Variety in sugar pucker:

A recent bone of contention among nucleic acid crystallographers has been the variability of sugar pucker. Previous small molecule work nearly always showed sugar puckers in either 3'-endo or 2'-endo conformations. In only one instance was another conformation, the O1'-endo pucker, found.[10] This constancy was interpreted as arising from a "rigid nucleotide" which snapped back and forth between the 2'-endo and 3'-endo conformation.[11] However, theoretical calculations of the energy required for the transition from 2'-endo to 3'-endo showed that the energy surface was rather flat.[12] The activation energy for transition between the two states was calculated as 0.6 kcal./mole. Crystal structures of longer pieces of DNA which were done about the same time showed intermediate sugar puckers.[13] As longer

10. J. Konnert, I. L. Karle and J. Karle (1970). *Acta Cryst.* **B 26,** 770-778
11. C. Altone and M. Sundralingam (1972). *J. Amer. Chem. Soc.* **94,** 8205-8212
12. M. Levitt and A. Warshal (1978). *J. Amer. Chem. Soc.* **100,** 2607-2613

structures continue to appear, the intermediate sugar puckers are becoming more common. The latest salvo fired by the rigid nucleotide group[14] finishes by saying "...one not only questions the extent to which the restrained refinement procedures [specifically, the Jack-Levitt procedure] determine the reported sugar conformation but also ponders the significance, if any, of the unusually puckered models." The alternative explanation proposed by this group requires the crystallographer to believe that what he sees in the crystal is not the actual conformation, but a mixture of 2'-endo and 3'-endo sugars which is blurred together to look like an intermediate conformation. If this were true, the electron density map should show density for the sugar which is large enough to accommodate sugars of both puckers. This in not the case.

It seems better as a rule to make the theory fit the data, rather than the reverse. Perhaps large and small pieces of DNA have different rules for determining sugar pucker. Such a length-dependent conformation change has a precedent in protein crystallography. Two closely related helical protein structures are the α helix and the 3_{10} helix. In the α helix, the amino hydrogen from one peptide bond hydrogen bonds to the carbonyl oxygen of another peptide bond 3 residues away. In the 3_{10} helix the same bonding occurs, but to a carbonyl group 2 residues away. These ideal structures are seen in small peptide structures. However, chain folding stresses built up in large proteins can distort the helix and cause the amino hydrogen to appear midway between the two carboxylic oxygens, forming a bifurcated hydrogen bond. The helix is neither an α helix nor a 3_{10} helix, but is an intermediate form. This theoretically untidy but realistic state of affairs may have its nucleic acid counterpart in the length-dependent variation of sugar conformation: C3'-endo *vs.* C2'-endo for mono to tetranucleotides, and intermediate conformations in longer hel-

---

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

14. W. K. Olson and J. L. Sussman (1982). *J. Am. Chem. Soc.* **104**, 270-278

ices.

Returning to Wang's octamer, Figure 3 shows that 4 sugar puckers cluster together near C1'-exo, and well away from either C2'-endo or C3'-endo, even though the structure was refined by the Konnert-Hendrickson constrained refinement program rather than the Jack-Levitt. It is interesting that Wang *himself refuses* to admit that he has a sugar pucker which is not C2'-endo or C3'-endo. He refers to his C1'-exo sugars as "a conformation closer to C2'-endo (or C1'-exo) than C3'-endo". [15]

One more piece of information pertinent to oligomer sugar puckers will be available when the refinement of GGTATACC, which is also being refined by the Konnert-Hendrickson program, is reported.

6.1.7 Solution work with CCGG:

Other physical chemical methods have been used to study A-DNA in general and the sequence CCGG in particular. The earliest of these CCGG studies was NMR work by Patel which showed that the sequence melted about 42°C in 0.1M phosphate buffer.[16] This fact was used in selection of this particular sequence. However, of more interest to the crystallographic analysis, is the observation of end-to-end aggregation of the tetramer in solution. This aggregation later manifested itself as the end-to-end stacking about the crystallographic twofold which makes the tetrameric ^{I}CCGG into an octamer.

Patel assigned the B conformation to CCGG by comparing the H-8 proton shift upon melting with that same shift in ribo-CCGG. In the ribo structure,[17] the shift of guanine H-8 upon melting is 0.68 ppm. This was reported as being "consistent with" A-DNA. Patel's finding that guanine H-8 shifted less than 0.1 ppm in deoxy-CCGG led him to conclude that the tetramer was in the B form. This is a reasonable result

15. A. H.-J. Wang *et al.* (1982). ibid
16. D. J. Patel (1977). *Biopolymers* **16**, 1635-1656
17. D. B. Arter, G. C. Walker, O. C. Uhlenbeck and P. G. Schmidt (1974). *Biochem. Biophys. Res. Com.* **61**, 1089-1094

since the conditions of measurement are not dehydrating enough to form A-DNA.

However, a study of the sequences CGCG, GCGC, GGCC and CCGG by circular dichroism[18] indicated that the CCGG tetramer existed in an "unusual" conformation under conditions virtually identical to those used for Patel's NMR study. Three of the tetramers gave what would be considered normal B-DNA CD spectra with maxima near 290 nm and minima near 260 nm, but CCGG, in contrast, had its maximum near 260 nm in a typical A-DNA pattern.[19] This finding may indicate an error in the NMR studies.

The solution structure of CCGG must still be considered unknown. Circular dichroism, which indicates an A form for CCGG, is perhaps the least reliable technique for structural characterization. On the other hand, the method used by Patel with NMR is not sufficient proof that CCGG is in the B form. The fact that ^{I}CCGG, GGCCGGCC, and CCGGCCGG[20] all crystallize as A-DNA may indicate that CCGG sequences in fact have a preference for the A form even when well hydrated. If this conformational preference exists, it is undoubtedly critical to the recognition of CCGG by restriction endonucleases and methylases such as *Hap* II, *Hpa* II, and *Msp* I.[21]

6.1.8 Other structural studies of A-DNA:

Recently, the Crothers group at Yale has applied the techniques of rotational correlation time and electric dichroism to the study of A-DNA.[22] This technique has been shown to give fairly good estimates of rise per residue and propellor twist of base pairs in B-DNA.[23] For A-DNA, the rotational correlation time is used to estimate

---

18. R. V. Kastrup, M. A. Young and T. R. Krugh (1978). *Biochemistry* **17**, 4855-4865
19. M.-J. B. Tunis-Schneider and M. F. Maestre (1970). *J. Mol. Biol* **52**, 521
20. A. H.-J. Wang *et al.* (1982). ibid
21. R. J. Roberts (1980). *Nuc. Acid. Res. Com.* **8**, r63-r80
22. H. M. Wu, N. Dattagupta and D. M. Crothers (1981). *Proc. Nat. Acad. Sci. USA* **78**, 6808-6811
23. M. Hogan, N. Dattagupta and D. M. Crothers (1978). *Proc. Nat. Acad. Sci. USA* **75**, 195-199

the rise per residue at 2.8Å, which is 0.6Å from the 2.2Å value found in ^{I}CCGG. The propellor twist was estimated by electric dichroism as "between 0° and 7°" from perpendicular to the helix axis. For ^{I}CCGG, the average propellor twist is 15°, agreeing with the outer limit of the dichroism measurement. Crothers interprets his results by saying that the base pairs do not have as large a propellor twist as in B-DNA. In fact the propellor twist is equally large in A and B.[24] These disagreements indicate that electric dichroism and rotational correlation cannot be considered very precise measurement techniques.

6.2 Water structure around ^{I}CCGG

6.2.1 Review of solution studies

The earliest fiber studies[25] showed that the transition from B to A DNA was caused by dehydration. Since that time considerable effort has gone into the study of the precise nature of these water interactions. Ultracentrifuge experiments by Hearst and Vinograd[26] established that the number of waters bound to DNA was a function of water activity. Specifically, they found the following relationship between water activity and DNA bound water:

- At a_w = 0.8, there are 9 waters per nucleotide
- At a_w = 0.9, there are 13 waters per nucleotide
- At a_w = 1.0, there are 50 waters per nucleotide

Falk and his coworkers[27] were able to distinguish waters bound to the DNA by their altered infrared spectrum. This allowed them to determine the numbers of waters bound to the DNA as a function of relative humidity. At 65% relative humidity, there are 5 to 6 waters per nucleotide. At 80% relative humidity, all sites on the DNA are

24. H. R. Drew *et al.* (1981). *Proc. Nat. Acad. Sci. USA* - ibid
25. R. E. Franklin and R. G. Gosling (1953). *Nature* **171**, 740-741
26. J. E. Hearst and J. Vinograd (1961). *Proc. Nat. Acad. Sci. USA* **47**, 825-830
27. M. Falk, K. A. Hartman, Jr. and R. C. Lord (1963). *J. Amer. Chem. Soc.* **85**, 391-394

filled. Further hydration results in swelling of the DNA. At 92% relative humidity, the DNA has 20 waters per nucleotide.

Falk also assigned a specific order to the binding of water molecules to their positions on the DNA.

1. Below 60% relative humidity:

    - First two waters go to charged phosphate oxygens.
    - The next waters go to O3' and O5'
    - Then water goes to O1'

2. Above 65% relative humidity

    - Base carbonyls and ring nitrogens are hydrated

Wolf and Hanlon have used circular dichroism to monitor the A to B transition as a function of relative humidity.[28] Linking their work with that of Falk, they have found that B-DNA has about 18 waters/nucleotide, while A-DNA may have as few as 4 waters/nucleotide. The A to B transition midpoint occurs about 13 waters/nucleotide. They have divided bound waters into two groups, confusingly called A and B. Group A consists of those 11 waters which attach directly to the DNA and disturb its infrared spectrum. For a single residue, there are 5 on the base, 4 on the sugar and 2 on the phosphate. Group B waters are those waters that do not attach directly to the DNA itself but also do not show the spectrum of bulk water. Presumably they attach to the group A waters.

6.2.2 Correlations with ^{I}CCGG:

By consulting the first layer hydration stereo drawings shown in Figures 5 and 6, certain correlations can be made between the preceeding solution studies and the ^{I}CCGG structure. Falk allocates the first two waters to the charged phosphate

28. B. Wolf and S. Hanlon (1975). *Biochemistry* **14**, 1661-1670

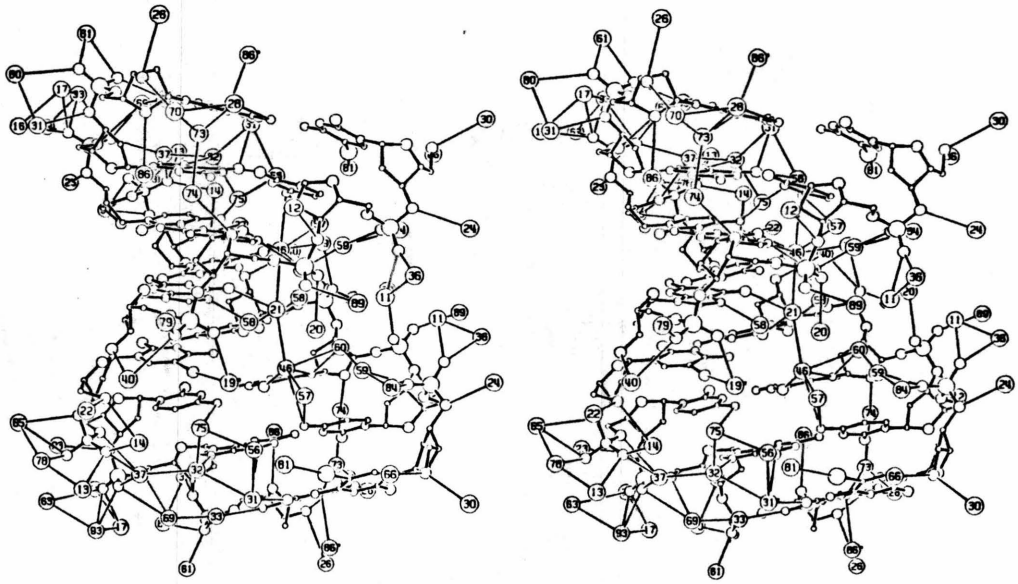

Figure 5. Major groove view of first layer hydration of ^{I}CCGG

oxygens. This is certainly in agreement with the crystal structure. Phosphate oxygens account for 42% of the bonds between first layer waters and DNA. These bonds have an average bond length of 3.0Å, indicating a relatively strong hydrogen bond. The phosphates are heavily hydrated in general, but in the case of the phosphate between cytosine 6 and guanine 7, there is only one water on each oxygen.

Falk allocates the next waters to O5' and O3', and then to the sugar oxygen, O1'. This disagrees with ^{I}CCGG. In ^{I}CCGG, the next most numerous bonds belong to the bases. Those in the major groove account for 23% and those in the minor groove account for 7% of the first layer bonds between DNA and water. In the major groove, these bonds are strong, with an average bond length of 2.9Å. In the minor groove they are much weaker, having an average length of 3.4 Å.

After the sugars, Falk proposes that the bases hydrate next. ^{I}CCGG indicates that the order should be reversed. In ^{I}CCGG the sugars have the longest, and hence the

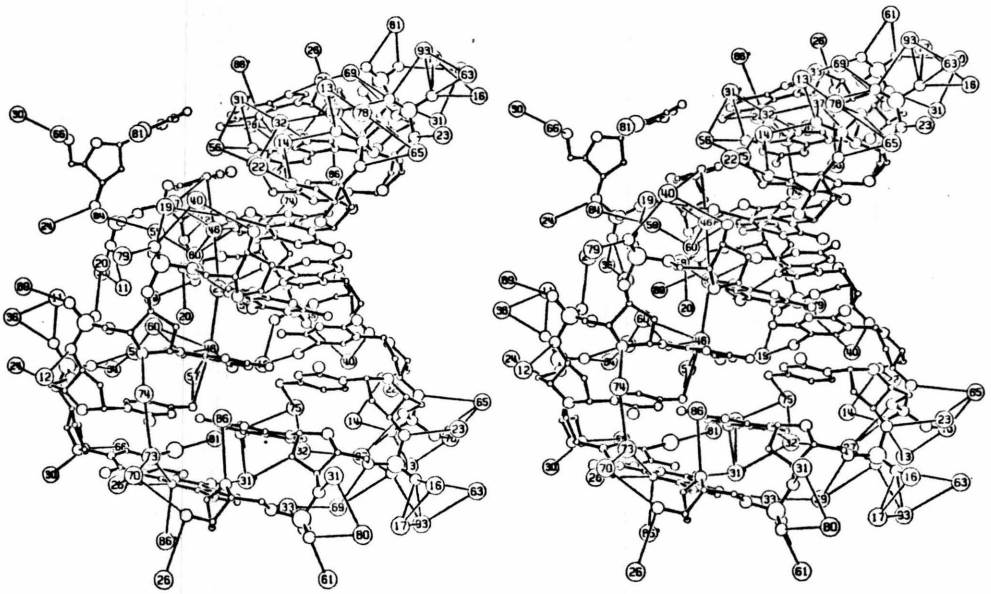

Figure 6. Minor groove view of first layer hydration of ^{I}CCGG

weakest, bonds of any of the groups. They account for 25% of the first layer water to DNA bonds. Falk seems to be correct when he ranks O5' and O3' before O1'. Sugar O1' atoms have an average bond length of 3.4Å while the O3' atoms are stronger as indicated by a average hydration bond length of 3.3Å.

The sugar oxygen O5' in a special case in A-DNA. Examination of Figure 5 shows that O5' is located along the major groove side of the sugar phosphate backbone. If a water is bridging between two phosphate oxygens, the O5' oxygen often will be close enough because of A-DNA stereochemistry to form a hydrogen bond. In ^{I}CCGG there are five O5' bonds with an average bond length of 3.2 angstroms. However, such a binding scheme would change with a change in DNA conformation.

6.2.3 New ordering of hydration sites on DNA

With these considerations, ^{I}CCGG gives us a new ordering of hydration on DNA.

1. Two waters on the phosphate oxygens. (in agreement with Falk)

2. Next, three waters on the major groove side of the base pairs. At this point we have accounted for the 4 to 5 waters/nucleotide that Falk sees below 60% relative humidity. We have also accounted for the four waters that Wolf and Hanlon see as the minimum for A-DNA.

3. It is not clear which area would hydrate next.

    - The O5' oxygens are positioned in the major groove in such a fashion that they are sterically hindered and cannot be reached by a water molecule unless that molecule also is bound to a phosphate oxygen. It is possible that one of the waters already on a phosphate would shift its position to a O5' and allow room for another water to bind to the phosphate oxygen. A fact consistent with this idea is that those phosphates with the most hydration are those phosphates with their O5' atoms hydrated. See residues two and three of Figure 5 for example.

    - The O3' atoms are accessible from the minor groove side of the helix. These are regularly hydrated in ^{I}CCGG with an average bond length of 3.3Å. These probably hydrate about the same time as ...

    - The minor groove guanine N2 and N3 atoms. These are hydrated in ^{I}CCGG when they are not sterically hindered by the 4_3 related helix.

4. We can add a water to O3', one water each to guanine N2 and N3, and one more water at a phosphate oxygen, for a total of 9 waters. According to Hearst and Vinograd this corresponds to $a_w = 0.8$. Thus we have reached the 80% relative humidity that Falk observed was sufficient to fully cover the DNA without causing swelling. We now have 3 waters in the major groove, 2 in the minor, 3 on the phosphate, and 1 on the O3' of the sugar.

5. According to Wolf and Hanlon, there are 11 group A waters. (*ie.* waters that bind directly to the DNA and disturb its spectrum.) If ^{I}CCGG can serve as a pattern, these two extra waters are probably added around the phosphate oxygens.

6. According to Wolf and Hanlon, there are 8 or 9 more waters (group B waters) to add before bulk solvent is reached. These probably correspond to the second and higher layer waters of ^{I}CCGG. Since the midpoint of A to B transition occurs at 13 waters and we already have 11 by now, by the time these group B waters are all added, we will have B-DNA. Some of these undoubtedly go into higher layers of the spine structure[29] in the AT-rich minor groove of B-DNA. No direct comparison can be made with ^{I}CCGG since the tetramer has no AT base pairs.

6.3 Spermine

Spermine was necessary for the crystallization of ^{I}CCGG. In fact this molecule is commonly used to crystallize nucleic acids. The three A-DNA structures currently under refinement, ^{I}CCGG, GGCCGGCC, and GGTATACC were all crystallized with spermine. Even the B-DNA dodecamer[30] and t-RNA were crystallized using this molecule. Both the B-DNA dodecamer and ^{I}CCGG were crystallized with low spermine to DNA ratios. In the case of ^{I}CCGG, the ratio moles spermine/moles ^{I}CCGG =0.09 because it was found that this low ratio promoted larger crystal growth.

In ^{I}CCGG the spermine cannot be located with confidence. In the initial anomalously phased electron density maps, there appeared a band of density stretching across the major groove from one phosphate chain to the other which was believed to be the spermine. The spermine in the dodecamer has been assigned to a similar position across the major groove.[31] However, most of this band later refined away and the rest was replaced by discrete water molecules. In GGCCGGCC

29. H. R. Drew and R. E. Dickerson (1981). *J. Mol. Biol.* **151**, 535-556
30. R. Wing *et al.* (1980). - ibid
31. H. R. Drew and R. E. Dickerson (1981). - ibid

no mention is made of the spermine location.[32] In GGTATACC the spermine has been tenatively positioned spanning the major groove and passing through the DNA two-fold.[33] These structures will have to await further refinement before anything is known with certainty.

6.4 Summary of Advances from ^{I}CCGG

Advances in DNA structure due to the solution of ^{I}CCGG can be classed in three groups:

1. Single crystal information about A-DNA: This includes adjustment of most of the helix paramenters and a major change (-12° to +17 °) in propellor twist.
2. Study of water structure around A-DNA including:
    - The first crystallographic study of A-DNA water structure.
    - Location and reordering of the water binding sites in DNA
    - Identification of the minor groove of CG-rich DNA as a relatively hydrophobic area on the DNA.
3. Identification of a possible recognition mechanism for proteins which recognize the sequence CCGG.

---

32. A. H.-J. Wang *et al.* (1982). ibid
33. Z. Shakked - personal communication

Appendix Refined ICCGG Coordinates

APPENDIX REFINED ^{I}CCGG COORDINATES

6.1 Coordinate Transformation

^{I}CCGG was refined in space group $P2_1$. To convert back to the original fractional $P4_32_12$ system use the following set of equations:

- X $P4_32_12$ = Z $P2_1$/41.1 Å
- Y $P4_32_12$ = X $P2_1$/41.1 Å
- Z $P4_32_12$ = Y $P2_1$/26.7 Å

Coordinates of ^{I}CCGG			
Atom Name	X $P2_1$	Y $P2_1$	Z $P2_1$
1 O5'	28.69225	21.57280	26.84247
1 C5'	28.00716	21.44655	25.57956
1 C4'	27.28270	20.12717	25.61525
1 C6	29.23810	18.36084	28.15337
1 I	31.35796	19.14377	30.13559
1 C5	29.92827	17.94553	29.28505
1 N4	30.23647	16.30988	30.98613
1 C4	29.60312	16.70859	29.84845
1 N3	28.67426	15.92104	29.30568
1 O2	27.15561	15.58567	27.69916
1 C2	28.02016	16.30556	28.20795
1 N1	28.28528	17.52457	27.64145
1 C1'	27.54564	17.96915	26.41919
1 C2'	26.10956	18.44748	26.69147
1 O1'	28.24947	19.13747	25.95752
1 C3'	26.25705	19.96494	26.71915
2 O3'	25.02866	20.68398	26.45670
2 P	24.03911	21.12987	27.64156
2 OL	22.91927	21.93358	27.09589
2 OR	24.81969	21.78647	28.71227
2 O5'	23.42960	19.76349	28.19466
2 C5'	22.49431	18.99651	27.39885
2 C4'	22.27689	17.72490	28.17177
2 C6	25.29520	17.72942	30.21312
2 C5	26.39516	17.99886	31.02411
2 N4	27.83986	17.29674	32.76624

Coordinates of ^{I}CCGG (Continued)			
Atom Name	X $P2_1$	Y $P2_1$	Z $P2_1$
2 C4	26.75803	17.05347	31.97635
2 N3	26.06058	15.92442	32.13538
2 O2	24.35825	14.62628	31.47336
2 C2	25.00031	15.66821	31.35999
2 N1	24.62090	16.55954	30.39708
2 C1'	23.46608	16.23465	29.50011
2 C2'	22.08907	16.51270	30.10254
2 O1'	23.55472	17.10706	28.35756
2 C3'	21.78563	17.91083	29.58649
3 O3'	20.37787	18.23874	29.63866
3 P	19.76572	19.12598	30.82596
3 OL	18.33730	19.41718	30.55515
3 OR	20.59134	20.35608	30.91347
3 O5'	19.94411	18.23288	32.16106
3 C5'	19.12143	17.05138	32.42966
3 C4'	19.75101	16.25188	33.55945
3 C4	24.00597	16.30518	35.57001
3 N3	24.09447	15.32175	36.48775
3 N2	25.44263	14.46164	38.19722
3 C2	25.19838	15.39832	37.23824
3 N1	26.12482	16.38202	37.05864
3 O6	26.99878	18.21805	36.08577
3 C6	26.07622	17.40488	36.14095
3 C5	24.89333	17.32845	35.34311
3 N7	24.48891	18.12122	34.32776
3 C8	23.32352	17.57869	33.93729
3 N9	22.97954	16.46872	34.66301
3 C1'	21.78003	15.60839	34.48987
3 C2'	20.74260	15.79318	35.59572
3 O1'	21.14400	16.07207	33.28452
3 C3'	19.75330	16.77562	34.98082
4 O3'	18.45969	16.72884	35.61133
4 P	18.21407	17.45447	37.02138
4 OL	16.78127	17.41945	37.38628
4 OR	18.83313	18.78960	36.92830
4 O5'	18.98996	16.57605	38.11295
4 C5'	18.50179	15.28004	38.53394
4 C4'	19.37990	14.79678	39.68837
4 C4	23.59505	17.07462	40.23163
4 N3	24.29764	16.42038	41.20238
4 N2	26.37357	16.33452	42.31299
4 C2	25.55707	16.86191	41.35280
4 N1	26.07082	17.84772	40.56342
4 O6	26.01482	19.39754	38.94162
4 C6	25.41417	18.52025	39.56081
4 C5	24.05269	18.08461	39.42586
4 N7	23.08513	18.53397	38.59720

Coordinates of ^{I}CCGG (Continued)			
Atom Name	X $P2_1$	Y $P2_1$	Z $P2_1$
4 C8	22.01505	17.78011	38.90099
4 N9	22.26778	16.86240	39.88634
4 C1'	21.32088	15.86535	40.47156
4 C2'	20.09926	16.54770	41.06621
4 O1'	20.75835	15.10870	39.38306
4 C3'	19.16602	15.34300	41.10596
4 O3T	19.52498	14.39168	42.12839
5 O5'	34.78635	20.12749	42.98569
5 C5'	35.27895	18.81458	42.61200
5 C4'	34.45020	17.85727	43.42087
5 C6	31.73105	19.26041	41.34337
5 I	31.67685	20.97504	38.84328
5 C5	30.90445	19.85645	40.38106
5 N4	28.71814	20.21132	39.52414
5 C4	29.52802	19.67038	40.47629
5 N3	29.00020	18.94954	41.47049
5 O2	29.29889	17.75771	43.33614
5 C2	29.79163	18.38547	42.40004
5 N1	31.16138	18.51201	42.33446
5 C1'	32.05501	17.87569	43.40401
5 C2'	32.56702	16.43178	43.23494
5 O1'	33.23610	18.64339	43.57718
5 C3'	34.03992	16.53682	42.75937
6 O3'	34.95102	15.28222	42.88747
6 P	34.50200	14.10829	41.89252
6 OL	35.33524	12.88520	41.91089
6 OR	34.44452	14.88707	40.65947
6 O5'	32.97849	13.75417	42.25365
6 C5'	32.55234	13.12596	43.48189
6 C4'	31.10315	12.80313	43.15482
6 C6	30.72118	15.33912	40.36136
6 C5	30.78351	16.15477	39.22829
6 N4	29.68010	17.14008	37.36162
6 C4	29.62383	16.34882	38.47510
6 N3	28.46944	15.79190	38.84682
6 O2	27.35480	14.51688	40.31023
6 C2	28.40799	15.02847	39.94553
6 N1	29.52588	14.77855	40.68929
6 C1'	29.39911	13.87601	41.88620
6 C2'	29.43439	12.39521	41.54175
6 O1'	30.48848	14.07591	42.82198
6 C3'	30.88866	12.05563	41.83347
7 O3'	31.13889	10.63095	41.87053
7 P	31.37346	9.80809	40.50525
7 OL	31.72272	8.40919	40.81284
7 OR	32.38309	10.51128	39.69695
7 O5'	29.96111	9.81350	39.74106

Coordinates of ^{I}CCGG (Continued)			
Atom Name	X $P2_1$	Y $P2_1$	Z $P2_1$
7 C5'	28.93590	8.87078	40.14951
7 C4'	27.72371	9.02498	39.26656
7 C4	27.38431	12.57121	36.17882
7 N3	26.20630	12.42556	35.52907
7 N2	24.88411	13.32930	33.82657
7 C2	26.02950	13.32393	34.55948
7 N1	26.95070	14.29975	34.30614
7 O6	28.87912	15.43367	34.64487
7 C6	28.16124	14.47759	34.94005
7 C5	28.36948	13.48883	35.94809
7 N7	29.41437	13.31603	36.78545
7 C8	29.03844	12.28167	37.56090
7 N9	27.79099	11.80051	37.24321
7 C1'	26.95395	10.76017	37.90857
7 C2'	26.76616	9.44390	37.16843
7 O1'	27.53400	10.43892	39.17290
7 C3'	27.78574	8.52793	37.82599
8 O3'	27.43430	7.12893	37.67673
8 P	27.89388	6.31362	36.37465
8 OL	27.47580	4.89591	36.44881
8 OR	29.31975	6.57677	36.15659
8 O5'	27.09305	6.97316	35.15828
8 C5'	25.68741	6.69514	34.99144
8 C4'	25.22739	7.35363	33.72791
8 C4	27.21689	11.21548	31.91980
8 N3	26.42334	11.73470	30.96709
8 N2	26.22287	13.51322	29.46544
8 C2	26.91524	12.85171	30.43387
8 N1	28.09398	13.39447	30.85117
8 O6	29.95340	13.51386	32.11237
8 C6	28.92255	12.90638	31.83022
8 C5	28.40456	11.69939	32.39354
8 N7	28.90691	10.92131	33.37175
8 C8	27.98183	9.95850	33.51059
8 N9	26.92023	10.10149	32.65511
8 C1'	25.65117	9.33398	32.60010
8 C2'	25.60973	8.19201	31.59740
8 O1'	25.51009	8.74218	33.89490
8 C3'	25.98424	6.99538	32.45929
8 O3T	25.57268	5.74169	31.88991
WAT 9	22.87546	22.29451	34.44115
WAT 10	39.03223	15.35583	41.09807
WAT 11	35.91507	10.77356	40.56984
WAT 12	28.86957	9.58239	44.13828
WAT 13	24.71236	24.72469	28.39394
WAT 14	27.38318	22.07094	29.72832
WAT 15	38.12708	18.84697	43.27396

Coordinates of ^{I}CCGG (Continued)			
Atom Name	X $P2_1$	Y $P2_1$	Z $P2_1$
WAT 16	15.71326	17.84462	30.60759
WAT 17	16.70154	20.03473	32.73630
WAT 18	18.44806	24.69693	31.50084
WAT 19	31.54509	5.90616	34.04625
WAT 20	33.80396	8.74271	37.94617
WAT 21	32.73351	13.35038	32.72876
WAT 22	30.30489	24.16092	25.95244
WAT 23	20.04610	19.56577	26.80925
WAT 24	38.18828	25.18158	44.26433
WAT 25	28.65294	24.65044	32.84888
WAT 26	18.56360	16.41988	44.67209
WAT 27	39.36414	23.34517	39.65976
WAT 28	37.59586	19.04645	23.30791
WAT 29	25.43233	25.77808	33.54160
WAT 30	24.65195	15.19827	25.25446
WAT 31	16.39568	14.27169	35.18245
WAT 32	13.00669	13.91278	35.80127
WAT 33	16.02460	12.95030	39.98839
WAT 34	37.13176	11.13017	38.13202
WAT 35	28.80469	5.34241	41.39368
WAT 36	22.68784	3.88231	25.47751
WAT 37	23.65076	21.13669	31.19217
WAT 38	15.81119	23.26941	30.06293
WAT 39	32.16869	23.36649	29.05861
WAT 40	33.70923	23.66675	26.71416
WAT 41	31.86336	21.78467	33.00409
WAT 42	25.96515	12.04826	42.72205
WAT 43	40.19809	15.03417	38.61783
WAT 44	20.60822	20.37468	24.13387
WAT 45	35.78171	15.88939	33.46552
WAT 46	31.17519	15.42055	34.84662
WAT 47	35.20650	14.97240	36.19202
WAT 48	33.35721	17.91434	39.25293
WAT 49	30.17421	6.03266	43.39156
WAT 50	33.75751	25.97881	36.44420
WAT 51	35.28368	24.92014	29.57631
WAT 52	20.67754	23.09845	36.18973
WAT 53	35.64240	23.38826	38.70226
WAT 54	38.23360	22.37488	34.96750
WAT 55	34.88393	21.52393	35.40392
WAT 56	29.22678	20.02292	35.31171
WAT 57	32.36420	17.53281	36.04160
WAT 58	34.00751	16.31920	31.29778
WAT 59	33.42847	14.76941	38.26389
WAT 60	37.01839	13.77567	32.17072
WAT 61	16.18442	20.01645	37.46584
WAT 62	35.95784	26.69983	35.95711

Coordinates of ^{I}CCGG (Continued)			
Atom Name	X $P2_1$	Y $P2_1$	Z $P2_1$
WAT 63	37.78879	25.31223	33.65392
WAT 64	26.97086	24.55333	24.21744
WAT 65	22.92049	21.61957	23.88181
WAT 66	35.61533	22.10085	40.86673
WAT 67	37.81128	21.10889	40.61458
WAT 68	40.20052	21.34871	41.85068
WAT 69	20.90021	20.79294	34.24814
WAT 70	40.03119	21.17653	22.75374
WAT 71	35.32204	5.72781	40.51955
WAT 72	32.96938	19.28310	33.25403
WAT 73	23.40503	13.24294	41.77881
WAT 74	21.09926	17.08371	45.12212
WAT 75	28.35091	19.85747	32.31110
WAT 76	27.05965	26.26587	28.73302
WAT 77	22.98029	25.37993	23.30289
WAT 78	24.33783	24.15001	25.73256
WAT 79	26.44994	3.68093	38.37366
WAT 80	14.01204	16.37453	35.58904
WAT 81	32.65770	22.63315	36.23270
WAT 82	38.94791	19.18996	35.84273
WAT 83	33.58118	8.29796	35.32224
WAT 84	35.76973	16.90804	39.39597
WAT 85	31.51845	23.95116	32.21445
WAT 86	18.77907	25.86765	44.35048
WAT 87	40.53183	17.53714	39.12218
WAT 88	24.72504	26.69980	24.72260
WAT 89	39.90665	5.59533	48.03493
WAT 90	30.00676	25.13036	30.13039
WAT 91	20.97115	23.88373	29.67679
WAT 92	40.43021	26.69954	40.42981
WAT 93	17.74277	21.83675	31.25867
WAT 94	35.98708	18.49216	36.56966