AN ABSTRACT OF THE THESIS OF

Blaine H. M. Mooers for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on April 30, 1997. Title: Controlling DNA Packing in Crystals.

Abstract approved: Redacted for Privacy

Pui Shing Ho

Crystals of DNA oligonucleotides are a rich source of information about the conformation of DNA. However, the packing of DNA in crystals limits their utility in several ways. Crystal packing of DNA can distort molecular structure, limit a sequence to one helical conformation, and hinder the study of non-self-complementary sequences. These three problems are addressed in this thesis by developing and testing strategies to control the packing of DNA in new lattices of A-, B-, and Z-DNA.

Different lengths of short oligonucleotides vary in their overall shape and thus in their ability to pack in crystal lattices. Due to distortions introduced by crystal packing, DNA octamers in the A-form crystallize in an extended conformation and have few correlations between base sequence and local helical structure. By reducing the length of alternating and non-alternating dG·dC sequences to six bases, the hexamers crystallize with fewer crystal packing induced distortions and with many correlations between base sequence and local helical structure.

In the past, there have been no crystal structures of the same sequence in two helical conformations. When the octamer sequence d(GCGTACGC), which crystallizes as A-DNA, is extended by one guanine in the 3' direction to form the

nonamer d(GCGTACGCG), the octamer duplex region crystallizes in the B-form. The octamer duplexes stack directly on top of each other while the overhanging base unstacks and slips into the minor groove of the neighboring duplex where it forms an unusual d[G*(G·C)] base triplet. This interduplex interaction was not restricted to one type of lattice since the nonamer sequence d(GCAATTGCG) crystallized in a different lattice with the same d[G*(G·C)] base triplet.

The two backbones of duplexes of non-self-complementary sequences are often so similar that such duplexes will enter the lattice in two orientations rather than one thereby complicating the structure determination problem. The two 5'-overhanging bases of the heptamer duplexes of d(GCGCGCG)·d(TCGCGCG) and d(GCGCGCG)·d(CCGCGCG) crystallize in different lattice environments. As a result, these duplexes crystallize without orientational disorder.

Controlling DNA Packing in Crystals

by

Blaine H. M. Mooers

A THESIS

Submitted to

Oregon State University

in partial fulfillment of
the requirement for the
degree of

Doctor of Philosophy

Completed April 30, 1997

Commencement June 1997

Doctor of Philosophy thesis of Blaine H. M. Mooers presented on April 30, 1997.

APPROVED:

Redacted for Privacy

Major Professor Representing the Department of Biochemistry and Biophysics

Redacted for Privacy

Head of the Department of Biochemistry and Biophysics

Redacted for Privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Blaine H. M. Mooers, Author

ACKNOWLEDGMENTS

I thank Dr. P. Shing Ho for his guidance, support, and example, and for the opportunity to pursue x-ray crystallographic studies of DNA structure. I thank Dr. Gary P. Schroth for his enthusiasm, interest, and encouragement. I am indebted to Dr. Todd F. Kagawa for teaching to me the fine points of DNA crystal growth and crystal structure refinement. I thank Beth Basham for her computer expertise. I thank Brandt Eichman for the many long days and nights of work that were required to finish in a timely manner the research presented in Chapter 4.

I thank my committee members, Dr. W. Curtis Johnson, Dr. Glenn Evans, Dr. Phil McFadden, and Dr. James White for their willingness in learning about the details of DNA structure and crystal packing.

Finally, I thank Dr. John Loeser of the OSU Department of Chemistry for providing the inspiration and encouragment that lead to my decision to pursue a doctorate in molecular biophysics. And last but not least, Prof. Melvin Morris and Dr. Roger Rosentreter, of the University of Montana, who light the flame of passion for research when I was an undergraduate student.

CONTRIBUTION OF AUTHORS

Dr. Shing Ho was involved in the design, analysis, and writing of each manuscript. Dr. Gary Schroth assisted with the crystal growth, x-ray diffraction data collection, structure analysis, and writing of Chapter 2. Dr. Wade Baxter assisted with crystal growth, x-ray diffraction data collection, structure refinement, and structure analysis of the crystal structures presented in Chapter 2. Brandt Eichmann assisted with crystal growth, x-ray diffraction data collection, structure refinement and analysis, and the writing of Chapter 4.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

DEDICATION

I dedicate this thesis to my wife, Gloria, who helped me in many kind ways.

Controlling DNA Packing in Crystals

Chapter 1

Introduction

Crystal structures solved to near atomic resolution provide detailed information concerning the conformation of DNA. Crystal packing, however, limits their usefulness to molecular biology in three ways addressed in this thesis. First, crystal packing distorts the structure of DNA. These distortions are a major concern because small changes in molecular structure can cause profound changes in biological function. Second, crystal structures of identical sequences in two helical conformations are nonexistent. This hinders the understanding of how two helical forms inter-convert. Third, the conformations of the backbones in heteroduplexes are almost identical, so the duplexes can enter the crystal pointed up in one unit cell and pointed down in the next unit cell. The resulting crystal disorder is avoided by limiting crystallographic studies to palindromic sequences in which two strands of one sequence can pair to form a duplex. These three problems are addressed by designing new lattices for the helical forms of DNA found in crystals: A-, B-, and Z-DNA.

Double helical DNA has two antiparallel polynucleotide chains. Each monomer in a chain has three chemical moieties: an aromatic base, a 2' deoxyribose ring, and a negatively charged phosphate group. Adjacent monomers are connected by a phosphodiester link between the 5′ and 3′ hydroxyl groups on the deoxyribose rings (Figure 1.1a). Glycosidic bonds

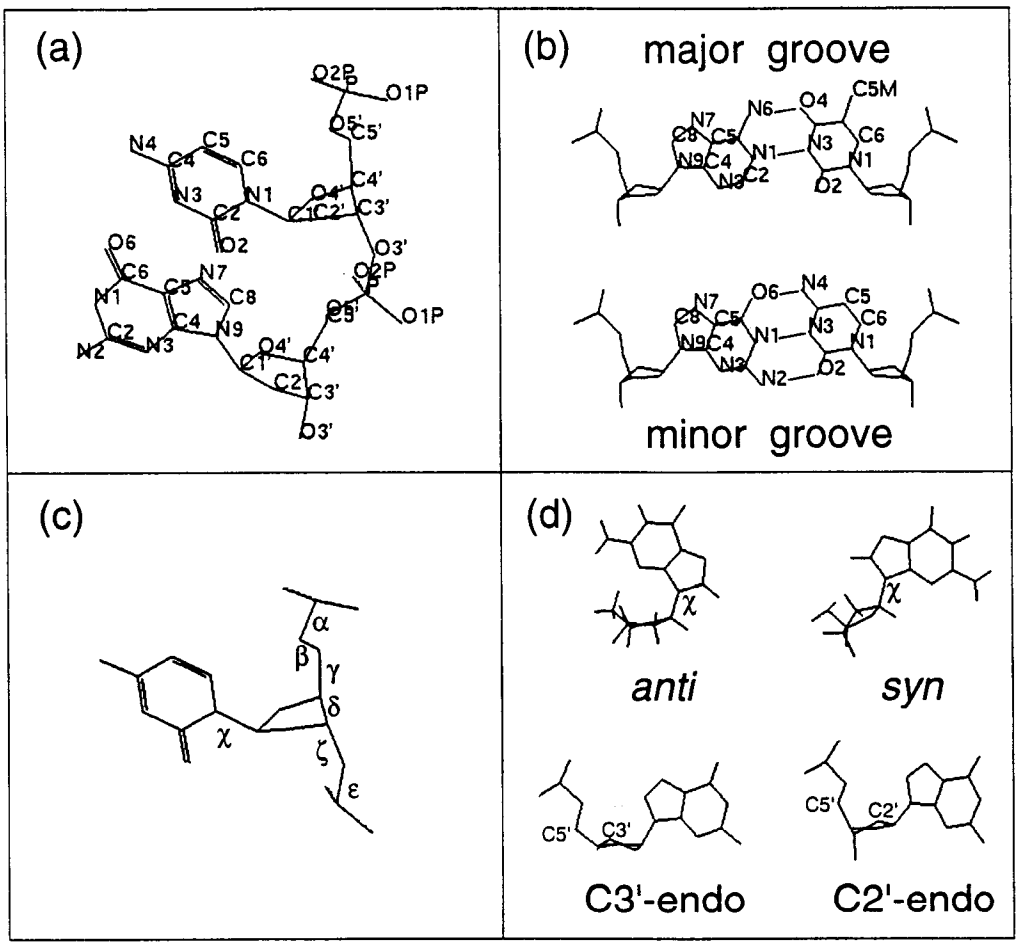


Figure 1.1. (a) The primary structure of a DNA dinucleotide. (b) Watson-Crick AT and GC base pairs. (c) The bonds about which the torsion angles of the backbone are defined--α: 03'-P-O5'-C5', β: P-O5'-C5'-C4', γ: O5'-C5'-C4'-C3', δ: C5'-C4'C3'-O3', ε: C4'-C3'-O3'-P, ζ: C3'-O3'-P-O5', χ (pyrimidines): O4'-C1'-N1-C2, χ (purines): O4'-C1'-N1-C4. (d) Guanine in the *anti* and *syn* conformations about the glycosidic bond (top), and a comparison of the C3'-endo and C2'-endo sugar puckers (bottom).

attach the bases to the backbone by linking the C1′ carbons of the sugar rings and the N9 nitrogens of the purines or the N1 nitrogens of the pyrimidines. Four bases are commonly found in DNA: two purines, guanine and adenine, and two pyrimidines, thymine and cytosine. Two antiparallel DNA strands form a duplex by hydrogen bonding with Watson-Crick base pairs between adenine and thymine and between guanine and cytosine (Figure 1.1b).

The duplex forms a double helix to bury most of the hydrophobic surface areas of the bases. If the duplex could be completely extended, it would have 3.7 Å wide gaps between the 3.3 Å thick base pairs which are spaced 7 Å apart. These gaps would destabilize the duplex, so to avoid this problem, the duplex fills the gaps by folding into a helix. The helix stacks the base pairs on top of each other in a spiral, thereby placing the backbones on the outside of the duplex where the exposure of the hydrophilic phosphates to solvent keeps the duplex soluble in aqueous solutions. The backbones separate two grooves on the surface of the helix. On one side of the base pairs, glycosidic bonds flank the minor groove while the major groove lies on the other side (Figure 1.1b). The depth and width of the grooves vary with helical form.

Watson and Crick (1953) determined the first structure of B-DNA (Figure. 1.2). Their model has 10 base pairs per helical turn (helical repeat). Each base pair is related by a right-handed 36° rotation about the helical axis (helical twist) and a 3.4 Å translation along the helical axis (helical rise). The helical axis passes through the middle of each base pair and is perpendicular to each base pairs (base pair inclination angle 0°). The deoxyribose rings are in the C2'-endo conformation (Figure 1.1d). In this conformation, the C2' and C5' carbons are above the plane formed by three of the four remaining

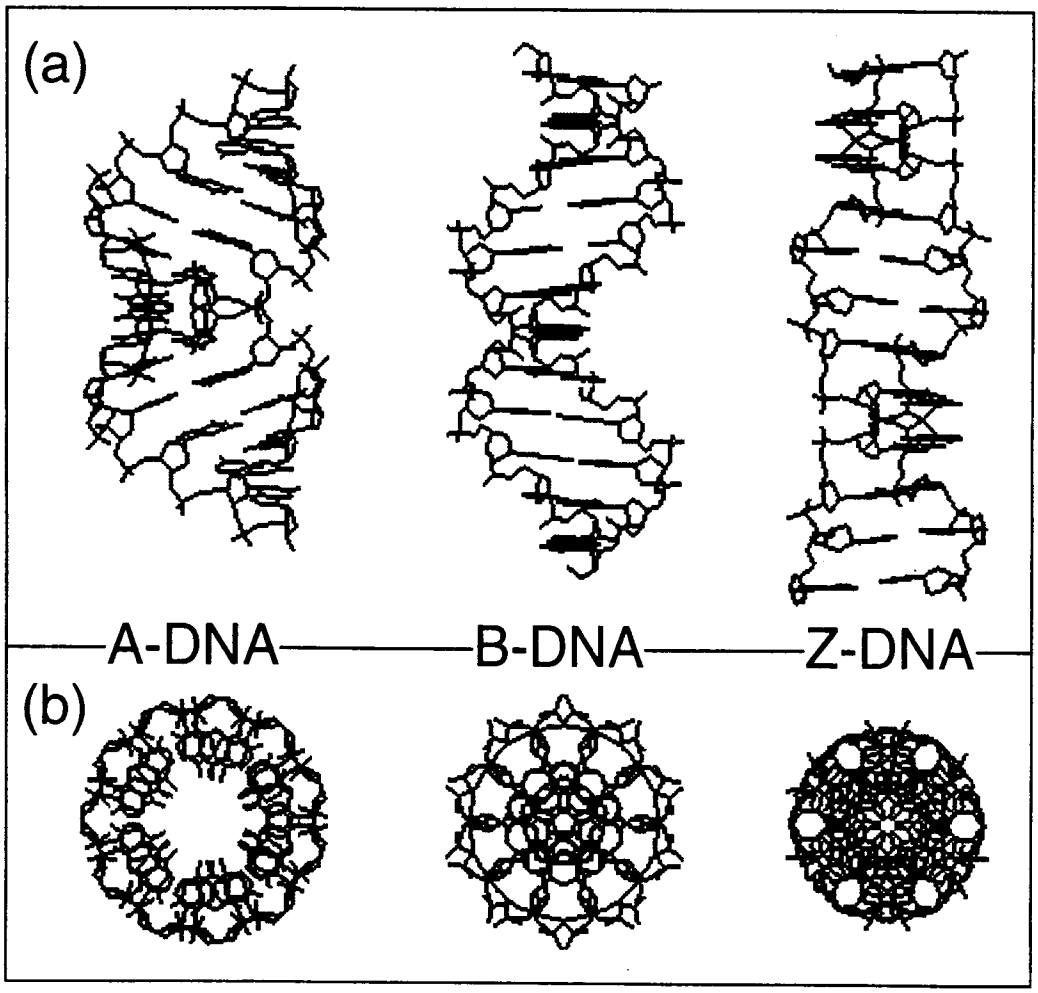


Figure 1.2. (a) Side view of twelve base pairs of A-, B-, and Z-DNA (left to right). Helix length increases from left to right. (b) View down the helix axes of the same models in (a). The base pairs in A-DNA (left) are displaced 4.5 Å towards the minor groove. This leaves a hollow channel in the middle of the helix (b). The base pairs are displaced ~ 2 Å towards the major groove in Z-DNA (right) (b).

ring atoms. The bases are *anti* relative to the sugar rings (Figure 1.1d). Although two grooves are equal in depth, the minor groove is narrow, and the major groove is wide. B-DNA is thought to be the predominant helical form *in vivo* because it is found in aqueous solutions and in DNA fibers in high relative humidity (Peck & Wang, 1982; Franklin & Gosling, 1953c).

Around 1980, fiber diffraction studies of DNA were supplemented with x-ray diffraction studies of DNA single crystals because it was finally possible to obtain large and pure quantities of synthetic DNA oligonucleotides for the growth of crystals. Single crystal x-ray diffraction data can be used to determine the position of every atom in the unique part of the crystal, whereas the fiber data can be used to determine only the average helical parameters.

The first crystal structure of B-DNA, known as the Drew-Dickerson dodecamer, has the sequence d(CGCGAATTCGCG) (Wing *et al.*, 1980). It differs in at least five important ways from fiber model B-DNA (Drew *et al.*, 1988). First, the backbone conformation is variable with the values of the backbone torsion angle δ and the glycosidic angle χ depending on whether the base is a purine or pyrimidine (Figure 1.1c). In addition, correlations between the values of the torsion angles ζ and ε give rise to the BI and BII conformations. The backbones in the fiber model have a conformation that is the average of BI and BII. Third, the bases in a pair twist about their common long axis rather than remain coplanar. This propeller twisting appears to improve base stacking within a strand (Levitt, 1978) and to increase the rigidity of the helix (El Hassan & Calladine, 1996). Third, the minor groove is very narrow in the central AATT region compared to the ends and to the fiber model. Fourth, the helical repeat is closer to 10.6 base pairs per turn, which is the helical repeat of B-DNA in solution (Peck &

Wang, 1981; Rhodes & Klug, 1981). Fifth, the helical twist varies between 28° and 42° in a base sequence dependent fashion rather than remaining 36°. Helical twist varies over a similar range of angles *in vitro* in a base sequence dependent manner (Lomonossoff *et al.*, 1981).

The base sequence dependence of DNA structure was hypothesized by Klug and others (1979) and is now widely accepted. In addition to the Drew-Dickerson dodecamer, most crystal structures of B-DNA have several strong correlations between base sequence and structure at the base step level (Yanagi *et al.*, 1991). In DNA polymers several hundred base pairs in length in solution, several repeating sequences are associated with the intrinsic curvature of DNA (Marini *et al.*, 1982; Hagerman, 1985, 1986). In addition, the base sequence dependence of DNA flexibility is found in the positioning of DNA as it wraps around the histone octamer (Satchwell *et al.*, 1986) and in the shaping of DNA as it is bound by the binding sites of transcription factors (Pabo & Sauer, 1992).

In lower relative humidity (75%), the sodium salt of calf thymus DNA fibers adopt the A-form (Franklin & Gosling 1953a,b). The first molecular model of A-DNA, which was derived from fiber diffraction data (Langridge *et al.*, 1957), has eleven base pairs in one right-handed helical turn. Each base pair rises only 2.6 Å along the helix axis because it is inclined 20° relative to the helix axis and because it is displaced 4.5 Å towards the minor groove. As a result, the minor groove is shallow and broad while the major groove is deep and narrow. The bases are in the *anti* conformation about the glycosidic bonds. The deoxyriboses are in the C3′-endo conformation that places the C3' carbon on the same side as the C5' carbon above the plane of the sugar (Figure 1.1d). This ring conformation separates the 3' and 5' phosphates by about 6 Å compared to about 7 Å in B-DNA.

As a result of significant crystal packing effects, most crystal structures of A-DNA are elongated compared to fiber model A-DNA (Wahl & Sundaralingam, 1997). These structures have an increased helical rise of 2.9 Å, a reduced base pair inclination angle of 12°, and a depressed base pair displacement of 3.8 Å.

A-DNA is of biological interest because it resembles B-DNA. The energy barrier between the two conformations is small, only about 1 kcal/(mol base pair) (Ivanov *et al.*, 1974). Dehydrating conditions induce the transition from B- to A-DNA form. Titrations with the dehydrating agent triflourethanol induce fourteen different sequences in the 5S Ribosomal RNA gene to undergo the B-to-A transition independently (Becker and Wang, 1989). Shortly after the onset of sporulation, which dehydrates cells, small acid-soluble spore proteins bind DNA and induce a transition to the A-form in *Bacillus subtilis* (Mohr *et al.*, 1991). Transcription factors surround their binding sites on DNA with a dehydrated environment which may favor the formation of A-like structures (Ivanov *et al.*, 1995; Guzikevich-Guerstein & Shakked, 1996).

Unlike the structures of A- and B-DNA, which were first determined by fiber diffraction studies, the structure of Z-DNA was first determined by single crystal diffraction studies of the hexamer d(CGCGCG) (Wang *et al.*, 1979). The hexamer forms a left-handed helix with a zig-zagged backbone, hence the name Z-DNA (Figure 1.2). The backbones zigzag because cytosines in *anti* alternate in each strand with guanines in the unusual *syn* orientation about the glycosidic bond (Figure 1.1d). The sugar puckers of the guanines are C3′-endo whereas those of the cytosines are C2′-endo. One helical turn has twelve base pairs, and the helical twist alternates between -9° in CpG steps and -51° in GpC steps. The helical rise averages 3.7 Å. The base pairs

are almost perpendicular to the helical axis, and they are displaced towards the major groove. As a result, the major groove is convex whereas the minor groove is deep and narrow.

The transition to the Z-form can be induced in many ways including high salt concentrations and the negative supercoiling of DNA. Polymers of poly(dG-dC)·poly(dG-dC) can form Z-DNA in solutions with high salt concentrations (Pohl & Jovin, 1972). Alternating pyrimidine/purine sequences that are rich in GC base pairs convert to Z-DNA *in vitro* and *in situ* under the negative superhelical stress generated behind transcribing RNA polymerase (Liu & Wang, 1987; Wittig, *et al.*, 1992). The biological function of Z-DNA, however, is still unknown (Herbert & Rich, 1996), but the ability of specific DNA sequences to adopt different helical forms in changing environments suggests that helical polymorphism may play a role in gene expression (Wells *et al.*, 1978).

DNA duplexes four to twelve base pairs long crystallize from 40 microliter drops of buffered salt solutions by equilibration against reservoirs of precipitating agents such as 2-methyl-2,4-pentanediol (MPD). As the DNA concentrates during the vapor diffusion process, the equilibrium between single-stranded and double-stranded DNA shifts towards double-stranded DNA (Ho *et al.*, 1991). The cation strength (or $CS = Z^2$[cation], where Z is the charge of the cation) of the solution drives the B-to-Z transition in a base sequence dependent manner in sequences that form Z-DNA (Kagawa *et al.*, 1989). A-DNA crystallizes in solutions with lower CS and MPD concentrations than those used to crystallize B-DNA (Timsit & Moras, 1992). When DNA duplexes reach super-saturating concentrations, several duplexes associate to form a crystal nucleus and crystal growth proceeds.

Crystals that reach a size of 0.2 x 0.2 x 0.2 mm are often suitable for x-ray diffraction studies.

In DNA crystals, duplexes occupy about 50 to 75% of the volume, and solvent molecules fill the remaining volume. Some solvent molecules are considered an integral part of the DNA structure (Berman, 1994) because they occupy well defined positions in the first hydration shell of waters surrounding the DNA (Schneider *et al.*, 1993). The disordered solvent molecules outside of the second hydration shell fill the cavities and channels between DNA helices. The contacts between DNA duplexes cover a minor amount of their exposed surface area, so DNA crystals can be thought of as highly ordered gels. DNA crystals dehydrate easily and are too fragile for most crystal physics experiments. Consequently, the physical properties of DNA crystals are inferred from the crystal structures and the crystallization solutions (Ho *et al.*, 1991).

DNA crystal packing is partially driven by the need to keep the hydrophilic phosphates hydrated and the hydrophobic terminal base pairs buried (Quigley, 1991). To remain hydrated, the backbones either face solvent channels or project into the well hydrated major grooves of neighboring duplexes (Timsit & Moras, 1992). The aromatic surfaces of the terminal base pairs are the largest patches of exposed hydrophobic surface on short DNA duplexes in solution. In the crystal structures of B-DNA decamers and Z-DNA hexamers, the stacking of the duplexes end-to-end covers most of the terminal base pair hydrophobic surface area (Figure 1.3a). In the crystal structures of B-DNA dodecamers and A-DNA octamers (Figure 1.3b & c), the terminal base pairs are buried in the minor grooves of neighboring duplexes(Figure 1.3b & c). Favorable van der Waals interactions further stabilize these interduplex contacts.

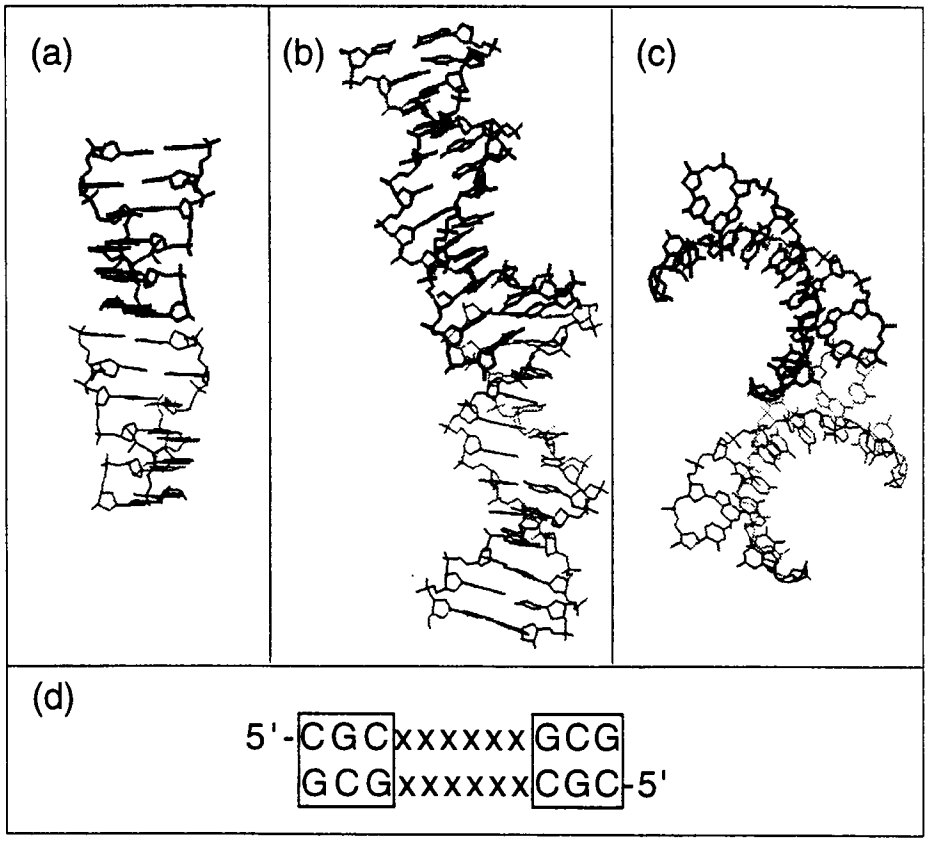


Figure 1.3. Major DNA packing motifs in crystals. (a) The end-to-end stacking of two Z-DNA hexamers. (b) Two base pairs at the end of one duplex overlap with the minor groove of a second duplex in B-DNA Drew-Dickerson type dodecamers. (c) The terminal base pair-to-minor groove packing in the crystal structures of the A-DNA octamers that crystallize in space group $P4_32_12$. The inside of the cresent formed by the duplexes is the major groove surface. (d) The two "packing driving boxes" for the Drew-Dickerson dodecamer, which is also an example of a self-complementary sequence in which two strands of the same sequence can form a duplex by aligning in antiparallel directions.

Hydrogen bonds between neighboring helices also stabilize DNA lattices (Wang & Teng, 1988). Four to eight hydrogen bonds link bases in the terminal base pairs of A-DNA octamers to bases in the minor grooves of neighboring duplexes (Tippin & Sundaralingam, 1996). At each end of the Drew-Dickerson dodecamer, the last two guanines form eight base-to-base hydrogen bonds with guanines in the minor grooves of a neighboring duplex (Wing *et al.*, 1980). Each Z-DNA hexamer is held in place by four hydrogen bonds between the backbones of neighboring duplexes (Ho & Mooers, 1997). If two duplexes are too far apart to form direct hydrogen bonds, water molecules often bridge the two duplexes with hydrogen bonds. Likewise, hydrated cations and cobalt hexammine sometimes bridge duplexes with hydrogen bonds (Gessner *et al.*, 1985).

Considering the negative charge on each phosphate and the density of DNA in crystals while ignoring the anions, DNA crystals have about 2.0-2.5 M negative charge. The electrostatic repulsion between negatively charged backbones is reduced by cation concentrations with CS in excess of the negative charge on the backbones (Wing *et al.*, 1980). Cations are occasionally recognized in crystal structures by their electron density and coordination spheres (Gessner *et al.*, 1989). The close approach of two phosphates from different backbones can be stabilized by a bridging cation as in the crystal structure of the B-DNA decamer d(CGCAATTGCG) (Spink *et al.*, 1995).

The symmetry of the crystal environment around a duplex varies widely. The unit cells of DNA crystals may be found in 65 space groups, and the asymmetric unit (the smallest unit from which the lattice can be generated by symmetry) of the unit cell may contain a dinucleotide, a strand, a duplex, or several duplexes. When a crystallographic symmetry operator

relates one part of a duplex to another, each part is in an asymmetric unit, and each part experiences the same crystal environment. If the whole duplex is in the asymmetric unit, perfect crystallographic symmetry no longer relates one strand of the duplex to the other. Instead, each region of the duplex is in a crystallographically distinct environment. However, the degree to which these environments differ is unpredictable and has to be assessed case by case. When two or more duplexes are in the asymmetric unit, they, too, may or may not be in symmetrically similar environments (Tereshko *et al.*, 1996).

The packing of DNA in crystals is largely controlled by the choice of base sequence, strand length, and duplex terminus, although a change in crystallization conditions can change crystal packing (Shakked *et al.*, 1989). The helical form of DNA oligomers is base sequence dependent and generally predictable (Basham *et al.*, 1995). It can be chosen by designing a sequence using empirically derived rules about how sequence dictates helical form (Peticolas *et al.*, 1988). In addition, certain di- and trinucleotides at key positions in the sequence can be selected to stabilize the packing of duplexes by forming interduplex hydrogen bonds with similar sequence elements in neighboring duplexes (Timsit & Moras, 1994) (Figure 1.4d). These sequence elements have been called "packing driving boxes" (Timsit & Moras, 1992). Sequence length controls the overall shape of the duplex and thus how duplexes pack together. Duplexes near a half or full helical turn in length usually have smaller lattice induced distortions (Verdaguer *et al.*, 1991). The termini of duplexes control the types of interactions that are possible at the interfaces of duplexes. Termini can be blunt or staggered with nucleotides overhanging the 3'- or 5'-ends. Blunt ends can participate in end-to-end stacking or end-to-minor groove packing whereas overhangs generally limit

duplexes to end-on-end stacking (Van Meervelt *et al.*, 1995; Vlieghe *et al.*, 1996b).

A major difficulty with studying DNA in crystals is that lattice contacts often distort the duplex by bending the helix axis, changing the base stacking, and altering the backbone conformation. Base sequence is thought to also introduce structural variations. Consequently, the separation of crystal packing effects from sequence effects is thought to require crystal structures of the same sequence in different lattices and crystal structures of the different sequences in the same lattice (Dickerson *et al.*, 1994). This approach assumes that crystal packing effects and base sequence effects are additive. This assumption may be useful if the interaction between the two of effects is small.

Crystal packing forces sometimes bend DNA helices up to 25°. Although these bends are relatively minor compared to the 90° bend of the TATA box sequence in the binding site of the *Escherichia coli* catabolite activator protein (CAP) (Schultz *et al.*, 1991), they provide information about the sequence dependence of DNA flexibility. In crystal structures of A- and B-DNA, bends generally occur towards the major groove (Young *et al.*, 1996). In B-DNA crystal structures, the CA/TG base step appears unusually flexible because the helix axis at this step has equal propensities to bend towards the major or minor groove. In the $P2_12_12_1$ lattice of 25 Drew-Dickerson type crystal structures, bends vary in magnitude, direction, and position along the helix axis, depending on the base sequence and crystallization conditions (Dickerson *et al.*, 1994, 1996). Dickerson concludes that the same packing forces are not exerted on every dodecamer sequence that crystallizes in this lattice.

Lattice contacts can modify the stacking of base pairs in the helix, thereby changing the helical parameters at the base step level. Of the six degrees of freedom between two base pairs in a base pair step, helical twist is the most sensitive to crystal packing forces because the flat surfaces of the bases pairs provide little resistance to twisting (Levitt, 1978). The resistance of a base step to changes in helical twist decreases as its base stacking grows poorer. For example, in the 5'-TpA-3' base pair step, there is have very little overlap of bases within and between strands. Consequently, this step has a very wide range of helical twist values in A- and B-DNA crystal structures (Hunter & Lu, 1997).

In response to strong lattice contacts, the backbone can change in a concerted fashion because it has six torsional degrees of freedom (α, β, γ, δ, ε, and ζ in the 5' to 3' direction between two phosphates) (Figure 1.1c), in addition to the χ angle about the glycosidic bond. For example, the backbone at the fifth base step in the A-DNA octamer d(CCCCGGGG) is extended in response to crystal packing (Haran *et al.*, 1987). A crankshaft motion (Figure 1.4) about the O5'-C5' bond changes the α and γ torsion angles from their normal *+gauche, -gauche* (30° to 90°), (270° to 330°) conformation to an extended *trans, trans* (150° to 210) conformation. This motion separates the adjacent phosphates by an additional Ångstrom while keeping the deoxyribose in the C3'-endo conformation. Crick & Watson (1954) proposed a similar conformation for an extended backbone many years earlier.

In Chapter 2, the problem of distorted DNA structure is addressed with the specific question of how to reduce the crystal packing induced distortions in A-DNA crystal structures. The majority of the A-DNA crystal structures are eight base pairs in length. Unfortunately, most correlations between local helical variations and base sequence are obscured in the A-

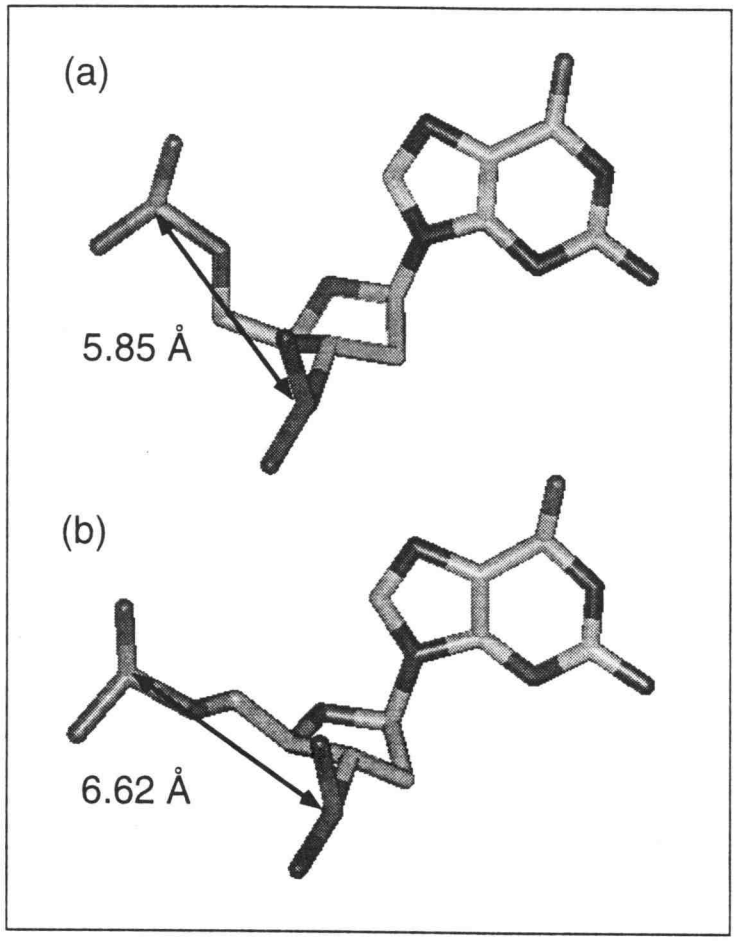


Figure 1.4. (a) Backbone in normal *+gauche, -gauche* conformation of A-DNA. (b) Backbone in the extended *trans, trans* conformation in A-DNA which separates two adjacent phosphates by a distance comparable to that found between two adjacent phophates in B-DNA.

DNA octamer crystal structures by crystal packing induced distortions to the molecular structure (Ramakrishnan & Sundaralingam, 1993a). A-DNA hexamers are hypothesized to crystallize with fewer crystal packing effects because they had a length that is closer to a half helical turn. Four hexamer sequences crystallized in conformations closer to fiber model A-DNA than the A-DNA octamers. The crystal structures of the hexamers demonstrate the utility of changing strand length to control lattice induced distortions.

A second major limitation on the study of DNA in crystals is the absence of crystal structures of identical sequences in more than one helical form. Apparently, the crystallization conditions for alternative helical forms of a sequence continue to elude DNA crystallographers. The lattice, however, may contribute to this problem. After a crystal nucleus has formed, only duplexes in the same conformation as in the nucleus may be incorporated into the lattice during crystal growth.

In Chapter 3, this problem is approached with the specific question of how to crystallize the A-DNA forming sequence d(GCGTACGC) as B-DNA. The addition of a guanosine to the 3' end of the octamer is hypothesized to stabilize the octamer duplex in the B-form by forming interduplex hydrogen bonds in the crystal. The overhang is also expected to block the terminal base pair from packing in the minor grooves of a neighboring duplex in the manner of A-DNA octamers crystal structures. The nonamer sequence d(GCGTACGCG) crystallizes in the B-form with the overhangs forming unusual base triplets in the minor groove of neighboring duplexes. The crystal structure of the nonamer demonstrates a method of crystallizing identical sequences in two helical forms.

Crystallographic symmetry imposes a third restraint on the study of DNA in crystals by preventing the lattice from distinguishing the two

distinct strands of a non-self-complementary sequence. In a such a heteroduplex, the sequence in one strand may complement that of the second strand, but the sequences are different. However, the backbones of both strands may have almost identical conformations. As a result, non-self-complementary duplexes often enter a unique position in the crystal in different orientations from one unit cell to the next. The resulting disorder complicates the structure determination problem because two or more duplexes appear superimposed in the electron density map. This orientational disorder is usually avoided by crystallizing self-complementary sequences which form homoduplexes from two identical strands. This strategy, however, reduces the variety of sequences that can be studied in crystals.

In Chapter 4, the third problem is pursued with the specific question of how to crystallize a heteroduplex without disorder in its orientation in the crystal. To answer this question, the lattice of the Z-DNA heptamer d(GCGCGCG) is selected because it crystallizes with the ends of the duplex in two distinct crystal environments. At one end of the duplex, the guanosine 5'-overhang flips out of the helical stack; at the other end of the duplex, a guanosine 5'-overhang remains stacked. This structure suggests that heteroduplexes with different bases in the 5' terminal position of each strand will crystallize in one orientation if one overhang has a stronger preference to remain stacked. This hypothesis is tested twice by crystallizing the heteroduplexes d(GCGCGCG)·d(TCGCGCG) and d(GCGCGCG)·d(CCGCGCG). In both experiments, the two strands of the heteroduplexes are distinct in the crystal because the guanosine overhang remains stacked and the pyrimidine flips out of the stack. These results suggest that non-self-complementary

sequences with distinct overhangs will crystallize without orientational disorder.

The central question asked in this thesis is how to control the packing of DNA in crystals to remove several limitations on DNA crystallography. Three major challenges caused by crystal packing are identified, and three general strategies to control DNA crystal packing are developed. Changing duplex length reduced lattice induced distortions in A-DNA crystal structures. The addition of 3' overhangs to an octamer sequence that forms A-DNA allows it to crystallizes as B-DNA. The extension of a strand of a Z-DNA hexamer in the 5' direction with different nucleotides allows the crystallization of heteroduplexes without lattice disorder. These results demonstrate crystal engineering methods which extend the opportunities to study DNA structure in crystals by controlling the packing of DNA.

Chapter 2

Alternating and Non-alternating dG-dC Hexanucleotides Crystallize as Canonical A-DNA

Blaine H. M. Mooers, Gary P. Schroth, Wade W. Baxter, and P. Shing Ho

Published in *Journal of Molecular Biology*,
Academic Press Limited, London, England
1995, **249**, 772-784.

2.1. Summary

We have solved the single-crystal x-ray structures of two different hexanucleotides: the alternating sequence d(Gm^{5}CGm^{5}CGC), and the non-alternating sequence d(Gm^{5}CCGGC). Both of these hexamers crystallize readily as A-DNA in the orthorhombic space group $C222_1$. Although hexanucleotides have been previously crystallized as Z-DNA, and in one case as B-DNA, this is the first time hexanucleotides have been crystallized as A-DNA. Both hexamers adopt a typical A-conformation, which is surprisingly more similar to the structure of A-DNA fibers than to other A-DNA single crystals. The structure of d(Gm^{5}CGm^{5}CGC) was solved to a resolution of 2.1 Å (R-factor = 19.6%). This structure has all of the features characteristic of canonical A-DNA, including it's helical repeat (11.2 b.p./turn), helical rise (2.6 Å/b.p.), base pair displacement (-4.7 Å), base inclination angle (16.9°), and sugar puckers that are predominantly 3'-*endo*. The lower resolution, non-alternating structure has similar overall average values for these parameters. We observed several sequence dependent correlations in these parameters, especially in the d(CG) base step. These steps have lower twist and rise values, coupled with high roll angles as compared to d(GC) steps. The molecular interactions involved in crystal packing and the detailed structure of the bound water in the crystals, however, are similar to those of longer 8 and 10 bp A-DNA crystal structures. Although the structural effect of cytosine methylation on A-DNA appears to be minimal, this modification significantly affects the ability of these sequences to crystallize as A-DNA. In conclusion, we present the A-DNA forming class of hexanucleotides, a new crystallographic system for studying DNA structure at near atomic resolution.

2.2. Introduction

Contained within the large number of single crystal oligonucleotide structures determined to date is a wealth of structural information that can, potentially, answer many questions concerning the structure of duplex DNA at the atomic level (Dickerson, 1992; Kennard & Salisbury, 1993; Heinemann *et al.*, 1994). The cloud that veils our understanding of how this detailed crystallographic information relates to the workings of DNA in the cell is the perceived 'tyranny of the lattice' (Dickerson *et al.*,1994). In short, the local conformation of DNA in a single crystal is necessarily constrained by the forces that hold DNA in the regular lattice of the crystal. At one extreme, it has been argued that the global conformation of a DNA duplex, whether it forms right-handed A- or B-DNA, or left-handed Z-DNA, is defined predominantly by the crystal packing forces. This leads to assumptions that certain characteristics of an oligonucleotide, such as length, will force the DNA into a conformation 'allowed' by the crystal lattice, regardless of other intrinsic factors such as sequence. Thus, it has been presumed by some that octanucleotides will only crystallize as A-DNA and hexanucleotides as Z-DNA.

For this study, we had designed a hexanucleotide sequence originally to test the limits on the types of sequences which could be crystallized as Z-DNA. Essentially all previous hexanucleotides have crystallized as Z-DNA, but unlike previous Z-DNA forming hexanucleotides, this sequence starts with a guanine rather than a cytosine base. It has the general sequence motif d(GCGCGC), as opposed to the canonical Z-DNA forming hexamer sequence, d(CGCGCG) (Wang *et al.*, 1979). Although these are both alternating dG-dC sequences, Quadrifoglio *et al.* (1984) had shown that short oligonucleotides of

the form $d(CG)_n$ readily form left-handed Z-DNA in solution, while $d(GC)_n$ do not. In an effort to overcome the inabilities of $d(GC)_n$ to become left-handed, we incorporated methylated cytosines into the sequence, and thus, the sequence we attempted to crystallize as Z-DNA was $d(Gm^5CGm^5CGC)$. It is very well established that methylation of cytosines dramatically stabilizes the Z-DNA conformation (Behe & Felsenfeld, 1981; Fujii *et al.*, 1982) and, therefore, we were hoping to overcome the "end effect" of starting the hexamer sequence with guanine as opposed to a cytosine.

To our surprise this sequence crystallized readily as right-handed A-DNA rather than as Z-DNA. This is the first report of a hexanucleotide crystallizing as A-DNA and is unusual in that it is an alternating dG-dC sequence. Although it is generally accepted that poly-d(G)/poly-d(C) sequences can easily assume the A-conformation, the B-A transition in alternating poly-d(GC) sequences has not been as thoroughly investigated (perhaps because these sequences have such a strong propensity to form Z-DNA). However, it is interesting to note that DNA octamers of alternating purine/pyrimidine sequences, such as d(GTGTACAC) and d(GTGCGCAC), have previously been crystallized as A-DNA (Jain *et al.*, 1987; Bingman *et al.*, 1992). Perhaps certain alternating dG-dC sequences will also crystallize as A-DNA, but that the crystallization properties of these sequences have not yet been fully investigated. We have recently, in fact, crystallized a partially methylated alternating dG-dC octamer with the sequence d(GCGCGCGC) as A-DNA (unpublished results).

We present in this paper the structure of the hexanucleotide $d(Gm^5CGm^5CGC)$ as A-DNA. This structure is compared to previous A-DNA crystal structures and the structure of A-DNA in fibers, and it is found to be more like the canonical fiber conformation. We also present the

structure of a non-alternating hexanucleotide, d(Gm^{5}CCGGC), again as A-DNA, which in comparison, allows us to distinguish sequence specific effects and also distortions to the DNA structure that are induced by crystal packing effects. Finally, a comparison of various demethylated forms of the alternating and non-alternating sequences provides some insight as to the effect of cytosine methylation on the structure and stability of A-DNA. With the inclusion of the phosphorothioate structure reported by Cruse *et al.* (1986), hexanucleotides containing only dG/dC base pairs have now been crystallized in all three of the major DNA duplex forms, as A-, B- or Z-DNA.

2.3. Results

2.3.1 Crystallization and crystal packing of hexanucleotides

The crystallization of various hexanucleotides (Table 2.1) containing methylated cytosine bases in the general sequence d(GCXYGC), where X and Y are either cytosine or guanine nucleotides, was attempted under a variety of solution conditions. In all cases where diffraction quality crystals were obtained, the solutions were characterized by low salt concentrations ([Mg^{+2}] $\leq$ 20 mM). These are typical of the conditions reported to give either A- or B-DNA crystals of longer sequences (Timsit & Moras, 1992), and significantly lower than the cation concentrations for crystallizing Z-DNA hexanucleotides (Ho *et al.*, 1991). The resulting crystals showed the general morphology of flat rectangular plates, reaching dimensions of up to 1.0 mm in length along two edges, but which were usually very thin in the third dimension (typically <0.2 mm in thickness).

The crystals were all isomorphous in the orthorhombic *C222$_1$* space group. The asymmetric unit is defined by one hexanucleotide duplex, with

eight asymmetric units in the unit cell. The unit cell volumes varied between the four crystals from 69,830 $Å^3$ to 72,888 $Å^3$ (Table 1). As also shown in Table 1, the effective resolution limit of the data collected for the crystal of d(Gm^{5}CGm^{5}CGC) was 2.1 Å; however, we included data to 1.9 Å in the refinement process. (In this work the "effective" resolution is defined as the limit at which at least 50% of the expected reflections are observed.) Note that the data on the other three structures are significantly lower in resolution than for the original d(Gm^{5}CGm^{5}CGC) hexamer (Table 2.1).

The structures of all four hexanucleotide sequences were determined to be that of A-DNA (Figure 2.1). The crystal lattice of the hexanucleotides in these crystals is shown in Figure 2.2. The molecules in these crystals are packed in such a way that the DNA duplexes form discrete layers, or sheets, in the *a*-*c* plane of the unit cell (Figure 2.2a). The three-dimensional crystal can be generated by stacking the planar sheets of hexamers along the unique crystallographic *b*-axis. Interestingly, there are no direct DNA-DNA interactions between the sheets of hexamers in these crystals. However, there are several waters that link these sheets together via hydrogen bonds. These planar sheets are unique to A-DNA hexanucleotides since longer A-DNA oligonucleotides pack in spirals rather than planar sheets. The planar organization of the hexamers in these crystals, and the apparent lack of interactions between the layers, may explain the typical thinness of these plates along one dimension of the crystal.

In a perpendicular view, we see that each sheet of the crystal is composed of four hexamer duplexes lying perpendicular to each other to form a rectangular array (Figure 2.2b). These arrays form a small cavity at the junction of the terminal base pairs of four hexamer duplexes and a large

Table 2.1. Crystal and refinement data for four A-DNA hexamers in space group $C222_1$.

	Gm^5CGm^5CGC	Gm^5CGCGC	Gm^5CCGGC	GCCGGC
Unit cell dimensions (Å)				
a	39.67	39.44	38.90	39.10
b	45.98	45.84	45.97	45.97
c	39.96	39.52	39.05	39.11
Volume($Å^3$)	72,888	71,449	69,830	70,297
Reflections				
Collected (Unique)	2937	2102	2816	1677
Completeness (%)	76.9	82.3	73.6	80.4
Refined	1994	1273	1190	929
Resolution Limit				
Effective*	2.1	2.5	2.5	2.7
Refined	1.9	2.2	2.0	2.3
Final R-value (%)	17.6	16.5	17.2	16.5
Solvent Molecules (Waters in Special Positions)	34 (3)	35(2)	31(1)	22(0)

Table 2.1 (Continued)

Mean Temperature Factor (Å2)				
Nucleotide	16.4	15.9	13.2	14.6
Solvent	42.3	44.3	34.6	39.0
R.M.S Deviation from Ideality in:				
Bond Length (Å)	0.013	0.019	0.018	0.018
Bond Angles (°)	3.506	4.122	4.155	3.959

*The effective resolution has 50% data completeness in reciprocal space in a thin shell with a radius that corresponds to the effective resolution.

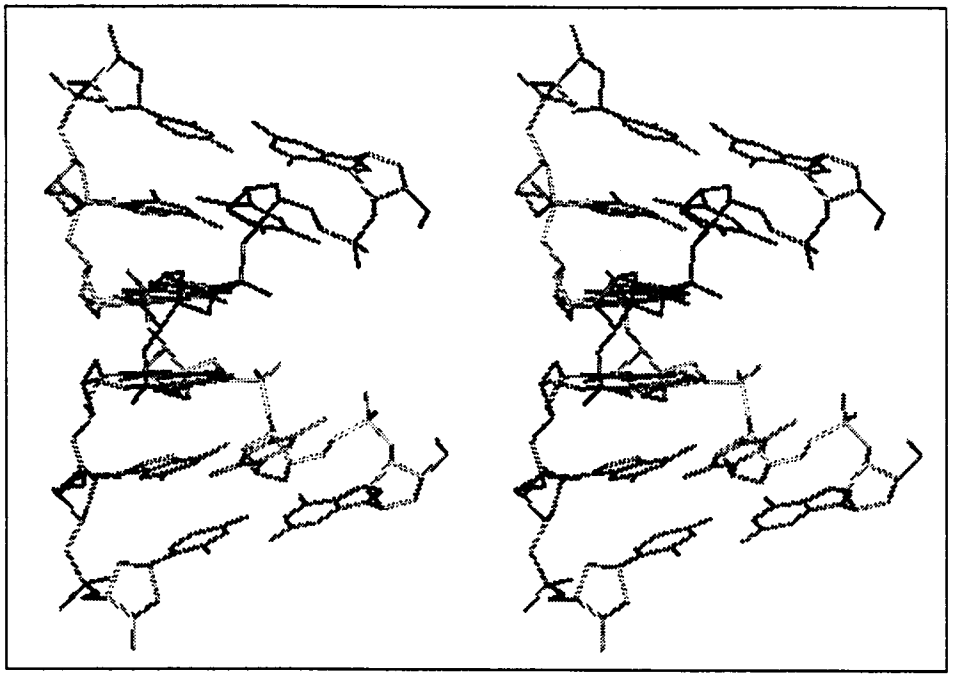


Figure 2.1. A stereo view of the crystal structure of d(Gm^{5}CGm^{5}CGC). A stick model of the structure is shown.

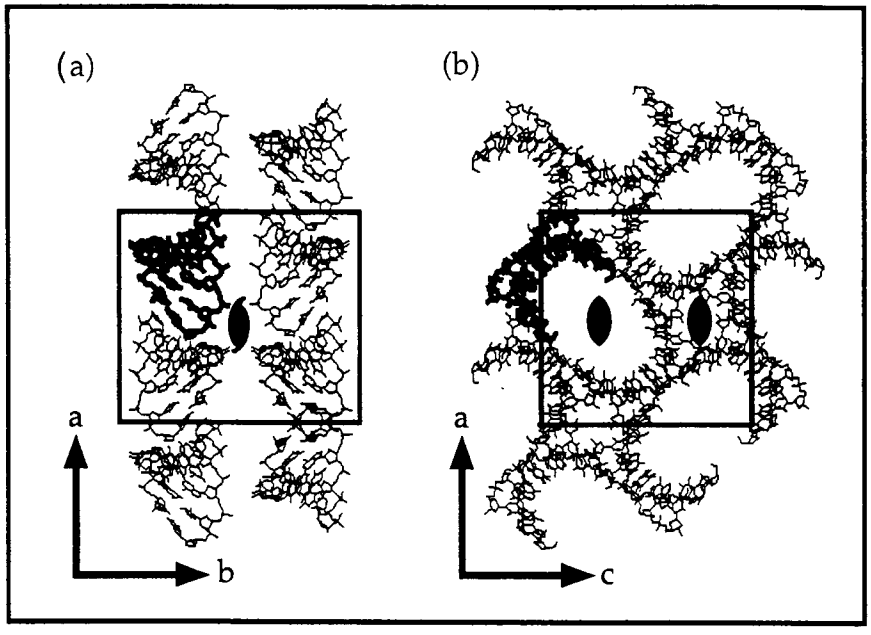


Figure 2.2. Packing of d(Gm^{5}CGm^{5}CGC) duplexes in the crystal lattice. (a) The DNA duplexes form planar sheets that extend out from the *a-b* face of the *C2221* unit cell and stacked along the *b* axis. Duplexes are related from one sheet to the next by the 2-fold screw axis normal to the *a-b*-face. Only one of thirteen 2-fold screw axes normal to the C-face is shown. (b) On the perpendicular face of the unit cell, the duplexes appear as rectangular arrays of hexamers within a sheet. Two-fold rotation axes relate the opposing hexanucleotides that line the large elliptic cavities formed by the arrays and the small cavities at the junction of four loops. Only two of the six two-fold rotation axes that are normal to the B-face are shown.

cavity in the center of each loop. The two-fold symmetry axes of this space group pass through the center of these cavities along the crystallographic *b*-axis. It is important to note that the cavities in this network are too small to contain disordered B-DNA molecules which have been sometimes found in other crystals of A-DNA (Doucet *et al.*, 1989; Bingman *et al.*, 1992).

The specific molecular interactions holding the rectangular arrays together are common to all four hexanucleotide structures. The 5'-terminal guanine base of two neighboring duplexes sit in the minor groove at the 3-10 and 4-9 base pairs of each DNA hexanucleotide. In other words, each hexamer has molecular contacts with the ends of two other hexamers. This is similar to the types of interactions observed in the A-DNA crystal lattice of octa- and decanucleotides (Wang *et al.* 1982; Frederick *et al.*, 1989; Ramakrishnan & Sundaralingam, 1993), although these longer sequences do not form discrete sheets of molecules. The structural consequences of these crystal lattice interactions are therefore length dependent.

2.2.2 Effect of methylation on the crystallization and structure of hexanucleotides

For the alternating d(GCGCGC) type hexamer sequence, we observed a strong correlation between the degree and position of methylated cytosines and the 'crystallizability' of the sequence. For example, the fully unmethylated sequence d(GCGCGC) would not crystallize under any conditions surveyed, while the double- methylated sequence d(Gm^{5}CGm^{5}CGC) crystallized very readily and easily gave the highest quality crystals, in terms of x-ray diffraction. Of the two single-methylated sequences we studied, methylation at the second position, in the sequence d(Gm^{5}CGCGC), gave significantly better crystals than methylation at the central cytosine, e.g. d(GCGm^{5}CGC). In the case of the non-alternating

sequence, of the general type d(GCCGGC), methylation again at the second cytosine appeared to facilitate crystallization. In this case, however, we also obtained diffraction quality crystals of the fully unmethylated sequence d(GCCGGC).

Although the number and the position of methylated cytosines helps in the crystallization of this family of hexanucleotide crystals, the effect of cytosine methylation on the molecular structure of each sequence is not very significant or well correlated with the positions of the methyl groups. There were no noticeable effects of cytosine methylation on the helical parameters, the base-stacking or the crystal packing interactions in these two A-DNA hexanucleotides. Frederick *et al.* (1987) made very similar observations regarding the crystallization and structure of d(GGm^{5}CCGGCC), when comparing the methylated and unmethylated versions of that sequence. Since the effect of cytosine methylation on the crystallization properties of hexamers is substantial, but the structural ramifications of the modification are not obvious, we are going to focus mainly on the sequence specific effects found by comparing the alternating and non-alternating hexamer structures.

2.2.3 The conformation of d(GCXYGC) hexanucleotides is canonical A-DNA

In general the structures of both the alternating and non-alternating hexanucleotide sequences are typical of A-DNA (Figure 1a), having wide minor grooves, and narrow major grooves as compared to B-DNA. The average helical repeat of these hexamers ranges from 11.0 to 11.2 bp/turn, which is identical, within the deviation of the structures, to the 11.0 bp/turn found for both the A-DNA fibers and an average of all other previous A-DNA crystal structures (see Table 2.2).

While the average helical parameters of previous A-DNA crystal structures differ significantly from the fiber structure of A-DNA, those of the hexanucleotides are actually closer to the values of the fiber conformation than to previous crystal structures. In B-DNA, the base pairs of the double helix are essentially perpendicular to the helix axis (with an inclination of approximately -6°). For A-DNA, the base pairs in the fiber are dramatically inclined (19.5°), while those in all previous crystal structures are significantly less inclined (12.2°). The inclination angle for the hexanucleotide structures in this study (17.2° on average) are intermediate between the fiber and the crystal structures of longer oligonucleotides (Table 2.2) but lie closer to the fiber structure. In addition, while the base pairs of B-DNA are intersected by the helix axis, in A-DNA fibers the base pairs are displaced by 4.5 Å away from the axis and toward the minor groove. Similarly, the structures of the hexanucleotide sequences have base pairs that are displaced on average by 4.4 Å, while the average displacement in previous crystal structures of A-DNA is only 3.8 Å. Each base pair in the hexamers rises along the helix axis by an average of 2.6 Å, which is again nearly identical to that of the A-DNA structure in fibers but lower than is usually seen in other crystal forms. Finally, we compared the root mean square (rms) deviations between corresponding atoms of the crystal structures and the fiber models at the nucleotide level as a detailed and length independent measure of how closely one structure matches the canonical A-DNA structure. The average rms deviation for d(Gm^{5}CGm^{5}CGC) to fibers was only 0.98 Å/nucleotide while it was 1.25 Å/nucleotide for the dodecamer d(CCCCCGCGGGGG) (Verdaguer *et al.*, 1991). The conformation of this dodecamer has been reported to be closer to

Table 2.2. Comparison of helical parameters for A-DNA structures.

Structure	Repeat (bp/ turn)	Helical Twist (°)	Helical Rise (Å)	Indination (°)	Roll Angle (°)	Displ.* (Å)
d(Gm^{5}CGm^{5}CGC)	11.2	32.0	2.6	16.2	8.3	-4.8
d(Gm^{5}CGCGC)	11.1	32.5	2.6	16.6	8.6	-4.5
d(Gm^{5}CCGGC)	11.1	32.5	2.6	17.1	9.4	-4.2
d(GCCGGC)	11.0	32.7	2.6	16.8	9.0	-4.3
d(CCCCCGCGGGGG)	10.4	34.7	2.6	13.6	7.1	-4.3
Fiber A-DNA**	11.0	32.7	2.6	19.5	10.9	-4.5
Crystal A-DNA**	11.0	32.7	2.9	12.2	6.8	-3.8

*This is the x-component of the displacement of the base pairs from the helical axis (Babcock *et al.*, 1994). The y-component was less than 0.3 Å in all cases. A negative value refers to the displacement of the helical axis toward the major groove.

**The fiber A-DNA model was generated by the program InsightII. Crystal A-DNA refers to the average parameters of 18 self-complementary A-DNA crystal structures in the Nucleic Acid Database (NDB). The structures in this set contain no mismatches, no unusual base pairs, and only room temperature diffraction data. The fiber model, each of the NDB crystal structures, and the four hexanucleotide crystal structures were analyzed with the program developed by Babcock *et al.* (1994).

that of fiber A-DNA than the crystal structures of A-DNA octamers. Thus the conformation of the hexanucleotide sequences is that of A-DNA and, interestingly, appears to be closer to the A-DNA structure observed in fibers rather than to other A-DNA crystal structures.

2.3.4 The structures of alternating and non-alternating hexanucleotides differ as A-DNA

Within the general class d(GCXYGC), for XY defined as GC, the hexanucleotide has an alternating sequence motif. Similarly, for XY defined as CG, as in d(GCCGGC), the sequence is non-alternating. We find that this difference in sequence is reflected strongly in many of the local helical parameters of the A-DNA conformation of these structures. Some of the major structural indicators of A-DNA include the rise of each base pair along the helical axis, the twist of the base pairs around the helical axis, and the roll angle between base pairs. All of these parameters alternate in the structure of the alternating sequences (Figure 2.3). The helical rise and base pair roll in contrast do not alternate in the non-alternating sequences (Figure 2.4). Other parameters, such as buckle and propeller twist, do not appear to follow any base pair specific behavior. The sites of methylation do not dramatically affect these general patterns for the alternating and non-alternating sequences (Figures 2.3 and 2.4).

The roll angle of the base pairs in the sequence d(Gm^{5}CGm^{5}CGC) alternates between ~4° for the d(GpC) steps and ~14° for the d(CpG) (Figure 3c). These are lower and higher roll angles, respectively compared to that of A-DNA fibers. The roll angles of both the d(CpG) (4.35° ± 3.91°) and d(GpC) (15.25° ± 2.18°) steps are also correlated between the alternating and non-alternating structures, which gives us some confidence that the roll angles are sequence dependent in these structures. The roll angle for the

Figure 2.3. Comparison of (a) helical rise, (b) helical twist, and (c) roll angle for the base steps in the two alternating hexanucleotide A-DNA structures. Open circles represent values for d(Gm^{5}CGm^{5}CGC) and open squares represent d(Gm^{5}CGCGC). The differences between the circles and the squares represents the structural effect of a single methyl group on the alternating hexanucleotides. The heavy line represents the mean of the alternating sequences. Horizontal dashed lines represent the values for fiber A-DNA and B-DNA structures.

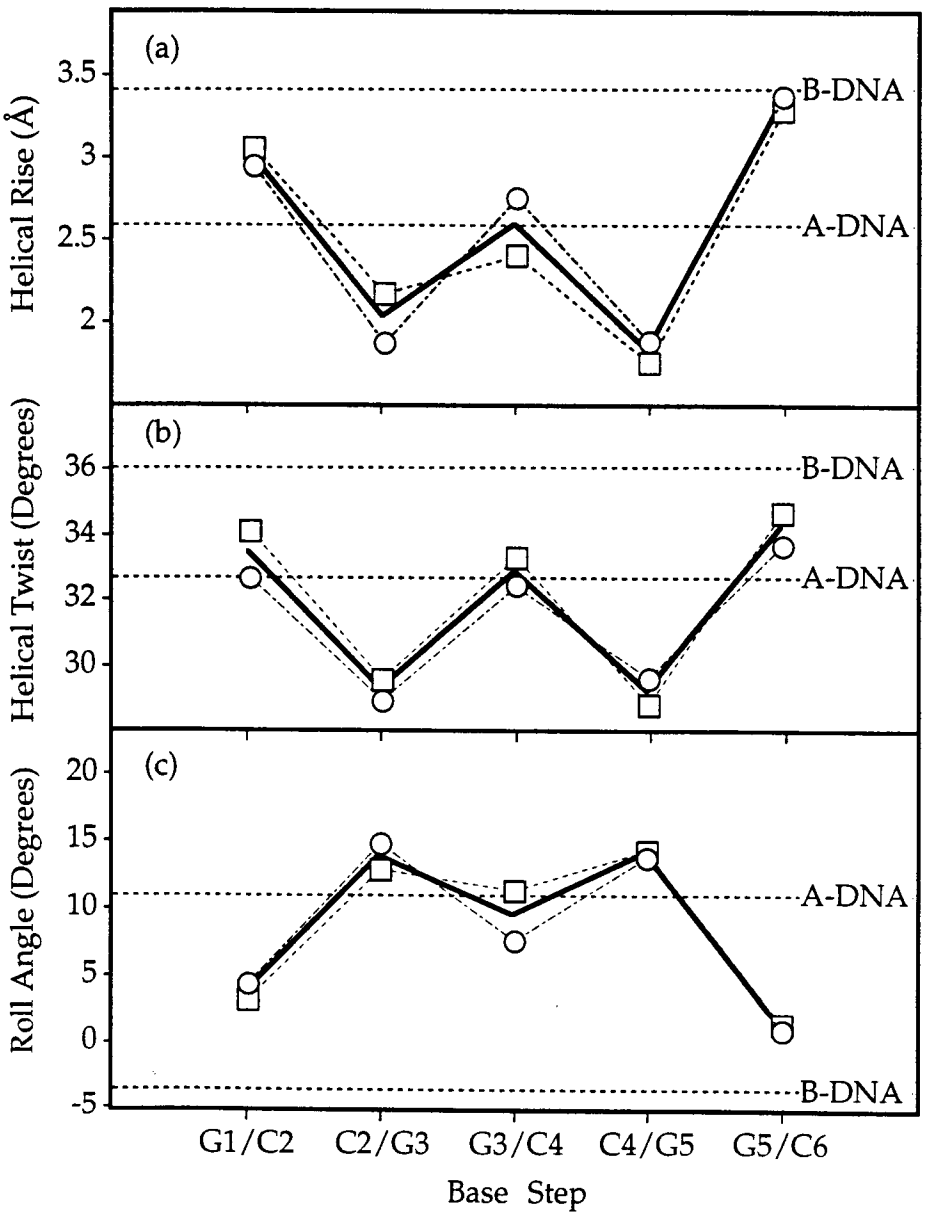

(a)
Helical Rise (Å)
3.5
3
2.5
2
B-DNA
A-DNA
(b)
Helical Twist (Degrees)
36
34
32
30
B-DNA
A-DNA
(c)
Roll Angle (Degrees)
20
15
10
5
0
-5
A-DNA
B-DNA
G1/C2
C2/G3
G3/C4
C4/G5
G5/C6
Base Step

Figure 2.3.

Figure 2.4. Comparison of (a) helical rise, (b) helical twist, and (c) roll for the two non-alternating hexanucleotide A-DNA structures. Open squares represent values for d(Gm^{5}CCGGC) and open squares represent d(GCCGGC). The differences between the circles and the squares represents the structural effect of a single methyl group on the non-alternating hexanucleotides. The heavy line represents the mean of the non-alternating sequences. Horizontal dashed lines represent the values observed for fiber A-DNA and B-DNA structures.

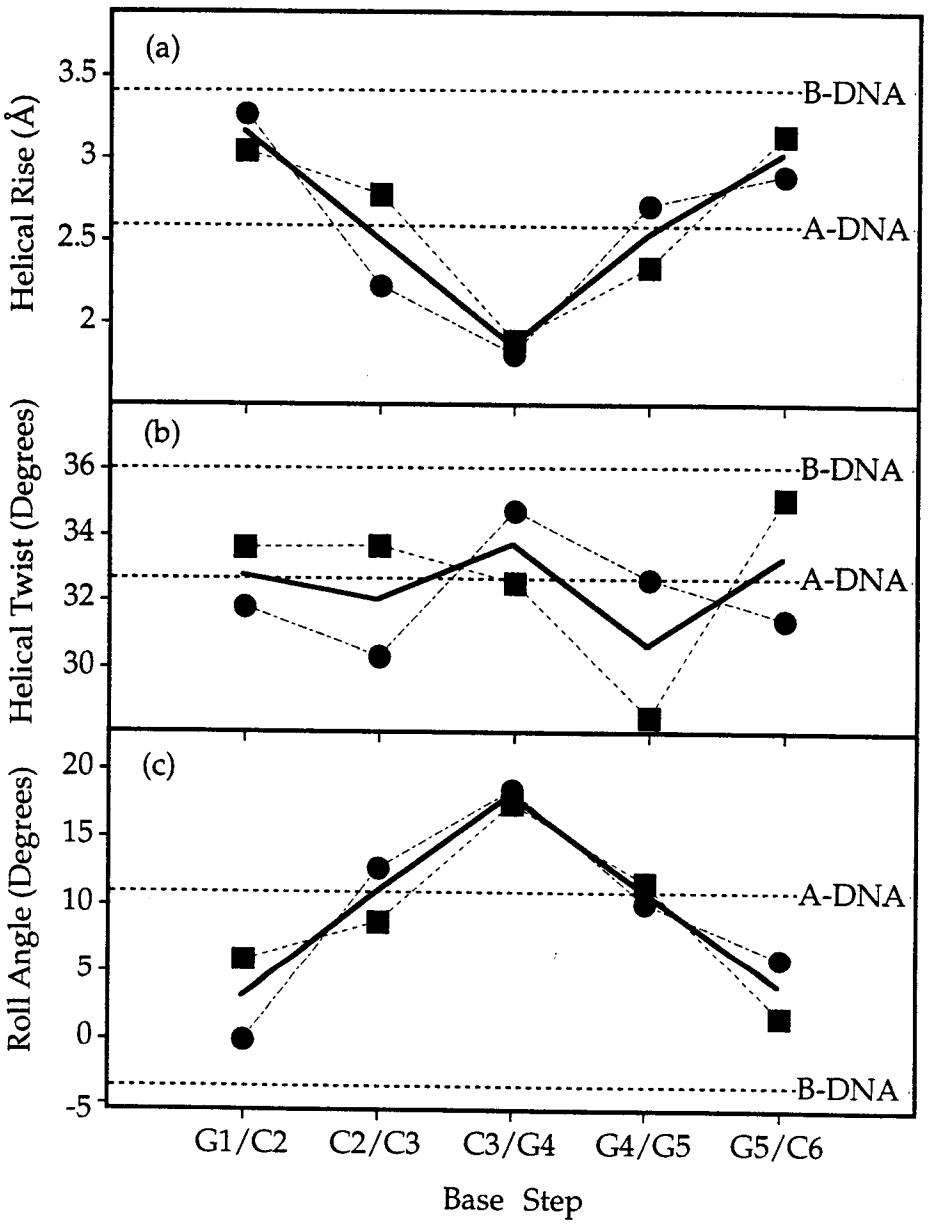

(a)
Helical Rise (Å)
3.5
3
2.5
2
B-DNA
A-DNA
(b)
Helical Twist (Degrees)
36
34
32
30
B-DNA
A-DNA
(c)
Roll Angle (Degrees)
20
15
10
5
0
-5
A-DNA
B-DNA
G1/C2
C2/C3
C3/G4
G4/G5
G5/C6
Base Step

Figure 2.4.

d(CpC/GpG) steps in the non-alternating hexamers is approximately 10.70° ± 1.70°.

Besides having a very high roll angle, the d(CpG) step has other unique features, especially within the alternating dG-dC sequence. The average rise for the d(CpG) step in all of these structures is only about 2.0 Å (Figure 2.3a), which is very low compared to an average of 2.6 Å for A-DNA fibers, and 2.9Å for other A-DNA crystal forms (Table 2.2). The helical twist values of the d(CpG) steps in the alternating sequence d(Gm^{5}CGm^{5}CGC) are also unusual. The average twist values for the d(CpG) steps in the alternating hexamers is about 29°, but the single d(CpG) step in the non-alternating structure has a typical A-DNA twist value of about 33°. When taken together these data suggest that the d(CpG) step is easily distorted into conformational extremes, and that in strictly alternating dG-dC sequences, the d(CpG) step in A-DNA will have the following properties: much lower than average rise, a lower than average twist angle, and a larger than average roll angle values.

In the final analysis, the low and high roll angles at each dinucleotide step compensate to give an average helical roll that is very similar to that of A-DNA in the fiber (Figure 2.3c). This trend is repeated for the helical rise and twist. The alternating sequence shows an alternating pattern in these structural indicators, while the non-alternating sequence generally does not. In the end the variations average to give values that are very close to those of fiber A-DNA. This degree of conformational variability at the base-pair level is probably also inherent in the structures of longer DNA, including the fibers. These local variations, however, simply cannot be resolved by fiber diffraction of randomly aligned sequences.

2.3.5 Structure of the phosphoribose backbone

The conformation of the furanose ring of the nucleotide ribose sugars (sugar pucker) are generally thought to be indicative of the DNA conformation. For A-DNA, the sugar is usually in the C3'-*endo* conformation, where the third carbon in the sugar ring lies slightly below the plane of the furanose ring. The sugar puckers observed in the current hexanucleotides were predominately C3'-*endo*. The exceptions were the G1 and G7 residues of d(Gm^{5}CCGGC) and G9 of d(Gm^{5}CGm^{5}CGC), which all adopt a C2'-*exo* conformation. The pseudorotation angles of these nucleotides, however, are very near the separation between C2'-*exo* and C3'-*endo*. Since all three of these guanines are involved in the intermolecular contacts, this suggests that their sugar conformations are distorted by crystal packing forces in the lattice.

There is a distortion to the DNA backbone that appears to be dependent on whether the sequence is alternating or non-alternating. For the alternating d(GCGCGC) sequence, a guanine is placed at the third position. In this situation, the ribose O4' of the terminal base forms a hydrogen bond with the N2 amino group of the central purine (Figures 2.5a and b). In the non-alternating sequence d(GCCGGC), a cytosine sits at the third position and does not allow formation of this hydrogen bond (Figures 2.5c and d). The structural consequence of this single hydrogen bond is to distort the phosphoribose backbone of the nucleotide at the point of interaction.

This hydrogen bond pushes the G7' nucleotide into the backbone of G3, forcing the latter to become extended. As a consequence, the 5'-end of the phosphoribose backbone must become extended and, therefore, adopts a

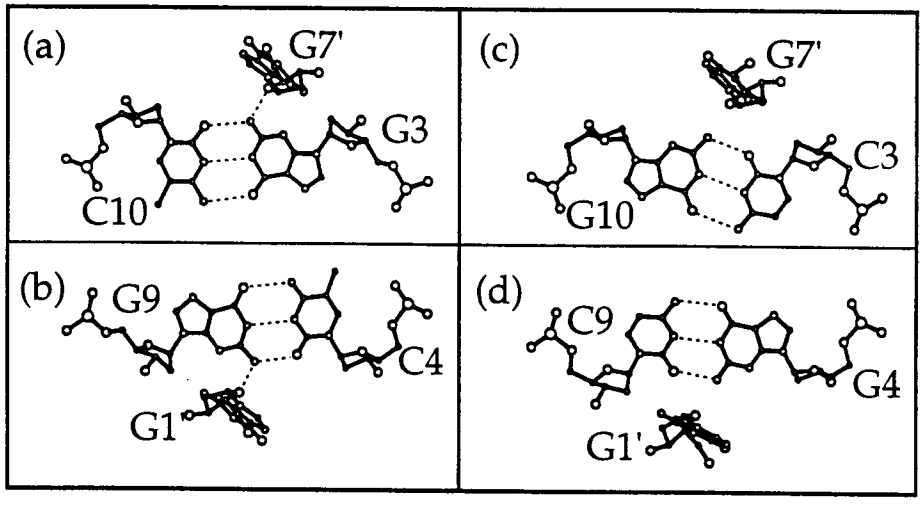


Figure 2.5. Intermolecular interactions in the crystals of the alternating (a and b) and the non-alternating (c and d) hexanucleotide sequences as A-DNA. (a) A single hydrogen bond links the G3 base of d(Gm^{5}CGm^{5}CGC) and the terminal guanine of a symmetry related duplex (G7'). (b) The analogous hydrogen bond between G9 and G1' of the symmetry related hexamer is shown for the d(Gm^{5}CGm^{5}CGC) structure. Both G3 and G9 have backbones in the extended *trans,trans* conformation. (c) The intermolecular contact is shown for the C3-G10 base pair of d(Gm^{5}CCGGC) and the terminal guanine of a symmetry related duplex (G7'). The cytosine C3 lacks a hydrogen bond donor in the vicinity of the O4' of G7. (d) A similar intermolecular contact is shown on the other side of the molecular dyad between the G4-C9 base pairs of d(Gm^{5}CCGGC) and the G1' base of a symmetry related duplex.

trans,trans conformation. This extended conformation is characterized by a low α torsion angle (139°± 4°compared to the average 294° ± 9° for the other residues) and a high γ torsion angle (190° ± 4° compared to the average 57° ± 4° for other residues). The α torsion angle is the O3'-P-O5'-C5' angle, and the γ torsion angle is the O5'-C5'-C4'-C3' angle. A similar interaction is observed at the G9 position where the ribose of the G1' from an adjacent duplex sits in the minor groove. An analogous hydrogen bond between the N2 amino group and the ribose of the intervening guanine residue and subsequent extension of the phosphoribose backbone is observed for these residues positioned across the molecular dyad (Figures 2.5a and b).

In the case of the non-alternating d(Gm^{5}CCGGC) structure, the third position is occupied by a cytosine. Thus the inter-duplex hydrogen bond cannot form because of the absence of the hydrogen bond donor (Figures 2.5c and d) as described above. Consequently, the backbone in the third position of these hexamers adopts the normal *-gauche,+gauche* conformation.

Similar interactions and distortions to the phosphoribose backbone are observed in the structures of octanucleotides that crystallize in the $P4_32_12$ space group (Jain *et al.*, 1989). In the octamers, the ribose of the terminal G9 nucleotide interacts with the base at the fifth position of the neighboring duplex. The *trans,trans* conformation is always observed when a purine is at this fifth position. Thus, the extension of the backbone at the central base pair results from crystal lattice distortions to the DNA structure and is not intrinsic to the structure of A-DNA. Any guanine at the central position of a hexanucleotide (G3) or octanucleotide (G5) structure that show this end-to-minor groove interaction will be forced into a *trans,trans* conformation.

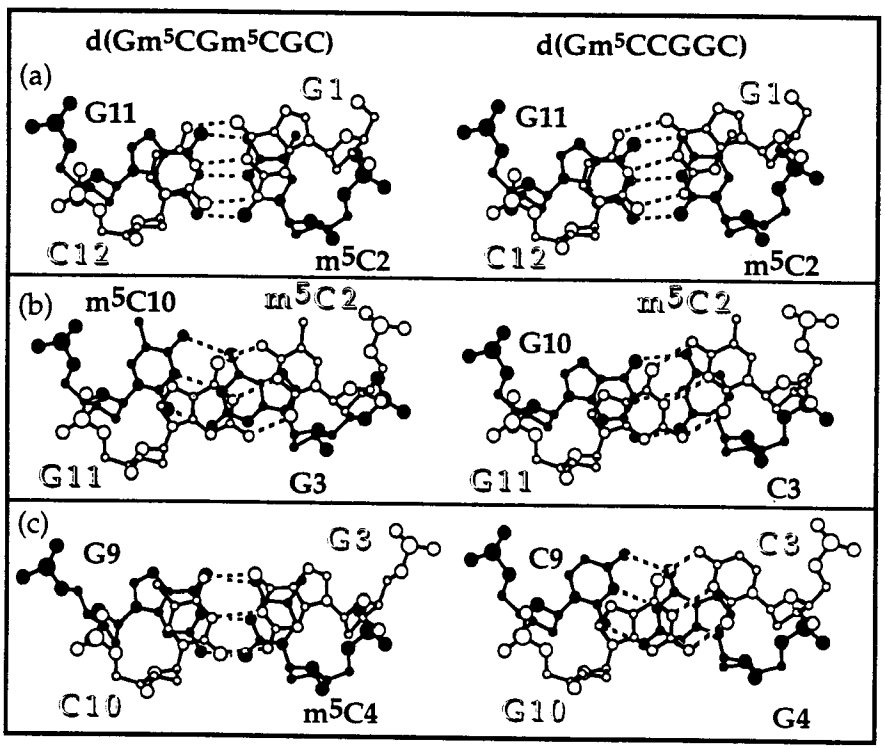


Figure 2.6. Comparison of base stacking in d(Gm^{5}CGm^{5}CGC) (on the left) with those of d(Gm^{5}CCGGC) (on the right). The base steps 1-2 (a), 2-3 (b), and 3-4 (c) represent the type of base stacking observed in the crystal structures. The 4-5 and 5-6 base steps (not shown) are very similar to the stacking of the 2-3 and 1-2 base steps, respectively. The atoms of the top base pair in each base step are represented by open circles, and the atoms of the bottom base pair in each step are represented by closed circles. The hydrogen bonds between bases are represented by dashed lines.

2.3.6 Effects of sequence and methylation on base stacking

The bases of the hexamer sequences stack in a manner that is typical of other A-DNA crystal structures (Figure 2.6). In the purine-pyrimidine steps of A-DNA and B-DNA, bases within a strand generally stack over each other (as in the d(GpC) steps shown in Figures 2.6a and 2.6c). In pyrimidine-purine steps (Figures 2.6b and 2.6c), the purine bases stack between opposite strands. This interstrand overlap is exaggerated in A-DNA where the pyrimidine-purine steps have large slide values (Calladine & Drew, 1984). This pattern of base stacking is a characteristic of previously determined crystal structures of alternating purine-pyrimidine A-DNA sequences (Jain *et al.*, 1989; Takusagawa, 1990) and of fiber A-DNA structures (Chandrasekaran *et al.*, 1989).

The bases of the homopyrimidine d(CpC) steps in d(Gm^{5}CCGGC) are intermediate between the d(GpC) and the d(CpG) extremes. Although the bases stack primarily along the same strand, some degree of interstrand overlap is observed. This pattern of base stacking is characteristic of previously determined crystal structures of sequences with homopurine and homopyrimidine stretches such as d(GGGGCCCC) (McCall *et al.*, 1985).

There is no apparent effect of methylation on the base stacking within these A-DNA structures. The degree of interstrand guanine overlap in the d(m^{5}CpG) steps of d(Gm^{5}CGm^{5}CGC) and the central d(CpG step) of d(Gm^{5}CCGGC) are nearly identical (Figures 2.6b and 2.6c). The cytosines in these steps have no interstrand overlap and little intrastrand stacking. Similarly, the distorted *trans,trans* conformation of d(Gm^{5}CGm^{5}CGC) in the second step does not affect base stacking (Figure 6b). This is consistent with the observations for A-DNA octanucleotides (Jain *et al.*, 1989).

2.3.7 Solvent structure of A-DNA hexanucleotides

The arrangement of solvent molecules around a DNA structure plays an important role in defining the stability of the structure (Saenger *et al.*, 1986; Westof, 1988; Ho *et al.*, 1991). The hexamers described here have the typical A-DNA solvation pattern of lightly solvated minor grooves and heavily solvated backbones and major grooves (Schneider *et al.*, 1992). No pentagonal arrangements of solvent molecules, however, were observed in the major groove of these structures, as has been reported previously for several A-DNA octamers (Kennard *et al.*, 1986).

Solvent structure plays a specific role in holding the DNA together within the DNA sheets and between the layers in this *C222*$_1$ lattice. The most important solvent molecules for this space group are the waters that link the sheets DNA hexamers together along the crystallographic *b*-axis in the lattice. One DNA-solvent-DNA interaction appears in both the alternating and non-alternating structures. This water bridges the O2P phosphate oxygen of the nucleotide at the 10 position (C10 in the alternating structure and G10 of the non-alternating structure) with the identical phosphate oxygen of a symmetry related duplex (water A in Figure 2.7). Two waters bridge the O2P phosphate oxygen of the symmetry related G3 nucleotides in only the alternating sequence (waters B and B' in Figure 2.7).

2.4. Discussion

We have crystallized a set of alternating and non-alternating d(G/C) containing hexanucleotide sequences as A-DNA. This set includes sequences having methylated cytosines. Both the number and position of methylation

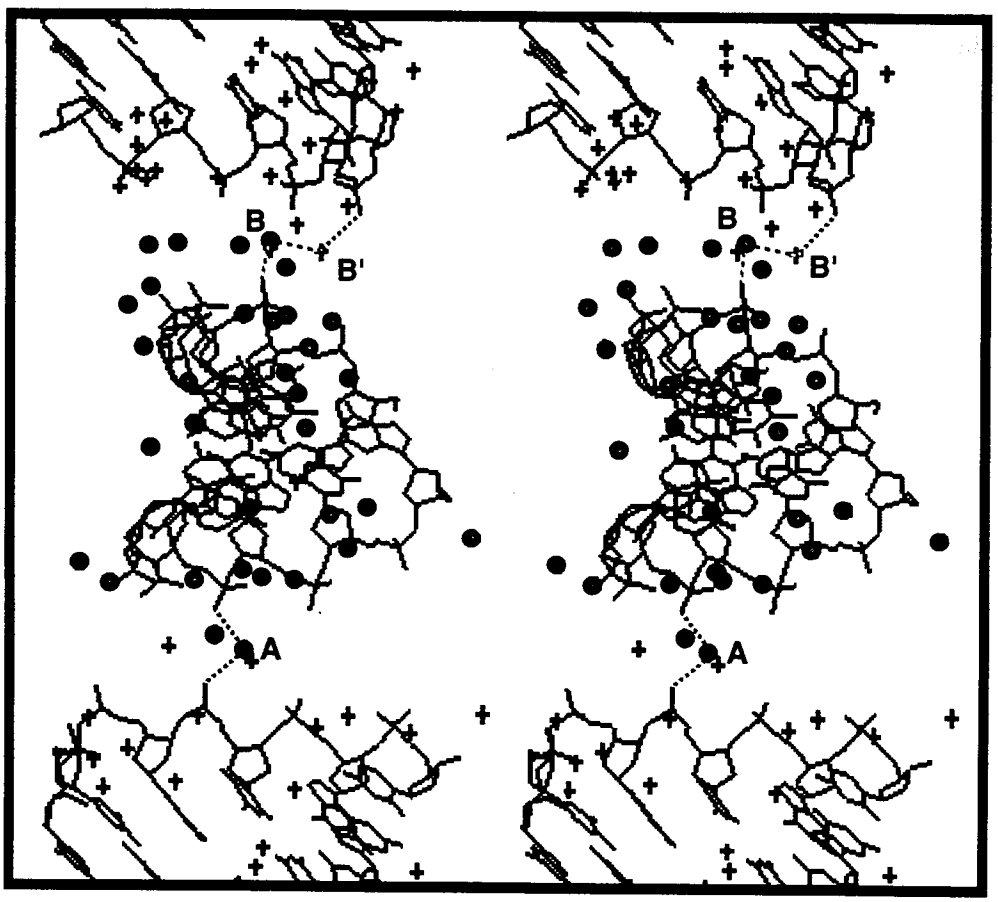


Figure 2.7. The hydration pattern of an A-DNA hexamer. A stereo view normal to the crystallographic *b*-axis of the hydrated hexamer d(Gm^{5}CGm^{5}CGC) is shown. One unique hexanucleotide is shown sandwiched between two symmetry related duplexes. The 34 waters assigned to the hexamer in the asymmetric unit are represented by spheres, while the waters assigned to the symmetry related duplexes are shown as crosses. The water molecule that links the C10 phosphate of the unique hexamer to the C10 phosphate of a symmetry related duplex is labeled A. The two waters joined by hydrogen bonds that link the two symmetry related G3 phosphate oxygens are labeled B and B'.

sites within the sequences strongly affects their ability to crystallize and the quality of crystals in terms of x-ray diffraction, but do not generally affect the overall structure. This concurs well with the observations of Frederick *et al.* (1987), that diffraction quality crystals of the sequence d(GGm^{5}CCGGCC) can be readily obtained, but that d(GGm^{5}Cm^{5}CGGCC) formed showers of microcrystals (i.e., this sequence crystallized too easily). The structure of d(GGm^{5}CCGGCC) however did not differ from the unmethylated version of this octamer. Since the methyl group at the C5 position of a cytosine base lies in the major groove of A-DNA, this cannot directly affect the crystal packing forces that hold the lattice together. The interactions within the crystal lattices of this octanucleotide and the current hexanucleotide sequences places the ends of neighboring duplexes into the minor groove of A-DNA. It is possible that methylation does not directly stabilize the A-DNA conformation, but instead simply lowers the intrinsic solubility of the oligonucleotide, thereby helping the DNA to come out of solution as a crystal. It is more likely that methylation affects the ability of the various sequences to adopt the A-conformation. We find that the alternating sequence of the type d(GCGCGC) appears to require methylation to crystallize as A-DNA, while the analogous non-alternating d(GCCGGC) sequence does not. The presence of two GpG steps makes this latter hexamer a better A-DNA sequence (Wang *et al.*, 1982), suggesting that the stabilizing effect of methylation is not required for crystallization.

How this stabilization comes about is not entirely clear, since there is very little distortion to the DNA or the overall solvent structure caused by this modification. Perhaps the dependence of the methylation effect on position will provide the answer, but we will need to study a larger set of

methylated and unmethylated structures, in conjunction with other theoretical and biochemical approaches to draw any general conclusions on the effects of methylation on the stabilization of A-DNA. Toward this end we have recently crystallized the octamer sequence d(Gm^{5}CGCGm^{5}CGC) in order to gain additional insight into the effects of methylation on A-DNA structure.

The overall structure of these hexanucleotides is that of canonical A-DNA, as defined from fiber diffraction studies. This is even more so than the crystal structures of longer oligonucleotides. The average base pair displacement and rise across the A-DNA hexanucleotide family are identical to those of the fiber structures (Arnott & Hukins *et al.*, 1989). In contrast, the crystal structures of longer oligonucleotides are characterized by base pair displacements that are closer to the helical axis by 0.5 Å, and by helical rises that are 0.3 Å longer along the axis compared to fiber A-DNA. The A-DNA structures of eight, ten and twelve base pair oligonucleotides are thus significantly longer and narrower helices as compared to both the hexamer and fiber structures.

These differences do not appear to correlate directly to the crystal packing density. The volume per base pair in the hexanucleotides (~1500Å^3/bp) falls well within the range of those for all previous A-DNA crystals. The molecular packing interactions are similar to those of octamer crystals in the $P4_32_12$ space group, with the same structural distortions induced by crystal packing at the points of DNA-DNA interaction. The hexamer structure in the crystals, however, does appear to be significantly more variable at the base pair level than the crystal structures of longer sequences. The average standard deviation in the inclination angles, for example, of the hexanucleotides is 12°, while that of all previous 8, 10, and

12 base pair A-DNA structures is 7°. It appears that the base pairs are less constrained by the crystal packing in the hexanucleotide crystals. This is also evident from the high degree of symmetry across the molecular dyad axis and between the complementary strands (Figures 3 and 4). Perhaps the lack of interactions between the sheets of the crystal lattice reduces the number of intermolecular contacts that constrain the structure in the hexanucleotides, thus allowing greater variability in the DNA conformations in a sequence dependent manner. Intermolecular interactions in crystals of longer A-DNA sequences are often stabilized by two hydrogen bonds at each intermolecular contact (Ramakrishnan & Sundaralingam, 1993).

In these crystallization studies, we had originally designed the sequence d(Gm^{5}CGm^{5}CGC) to potentially form Z-DNA. This sequence incorporates motifs that are characteristic of Z-DNA, including the alternating dG-dC base pairs and methylation of cytosine at the C5 positions. Why did this sequence not adopt the left-handed conformation? First, methylation also apparently tends to stabilize the A-DNA conformation. Thus A- and Z-DNA are competing conformations once B-DNA has been destabilized. Perhaps a more significant determinant which causes d(Gm^{5}CGm^{5}CGC) to be A-DNA, while d(m^{5}CGm^{5}CGm^{5}CG) very easily forms Z-DNA (Fujii *et al.*, 1982) is that the A-forming sequence starts with a guanine, while the Z-forming sequence starts with a cytosine. Quadrifoglio *et al.* (1984), showed that, in solution, short alternating sequences readily adopt the left-handed conformation only when the base at the 5'-end is a cytosine. For sequences that start with a guanine, Z-DNA cannot be completely induced until the length of the sequence exceeds 14 base pairs. This indicates that the inability of short d(GC)$_n$ sequences to form Z-DNA is the result of this unusual end effect.

What is the root of this strong end-effect? If we examine the classic Z-DNA forming sequence d(CGCGCG), we see that the bases along each strand alternate between the *anti* and *syn* conformations for the cytosines and guanines, respectively (Wang *et al.*, 1979). Treating this sequence as packets of dinucleotides, this sequence can be thought of as three separate *anti-syn* d(CG) dinucleotide steps which are stabilized in the Z-conformation by the high degree of stacking between the base pairs. This additionally keeps the aromatic bases inaccessible to solvent. A *syn-anti* step in Z-DNA, however, has the bases unstacked and more exposed to solvent. For the sequence d(CGCGCG) there are two destabilizing *syn-anti* steps separating the three stabilizing *anti-syn* steps. Thus, d(CGCGCG) overall would be expected to be stable as Z-DNA.

In this context, the sequence d(GCGCGC) would be described by three Z-destabilizing *syn-anti* steps and only two stabilizing *anti-syn* steps, making the sequence overall unstable as Z-DNA. A similar explanation for why d(GTGTACAC) formed A-DNA rather than Z-DNA was given by Jain *et al.* (1987). As this sequence is extended, the number of *syn-anti* steps approaches that of the *anti-syn* steps and the Z-DNA stabilizing and destabilizing effects are equalized. Long sequences of $d(GC)_n$ therefore would behave similarly to long sequences of $d(CG)_n$.

The structures presented here of the alternating and non-alternating d(C/G) containing hexanucleotides now expands the number of different DNA lengths that crystallize as A-DNA. By comparing the structure of a specific conformation in different crystal lattices, the intrinsic properties and the flexibility of a DNA form can eventually emerge (Heinemann *et al.*, 1994). Alternatively, we can also state that hexanucleotide sequences, which crystallize predominantly as Z-DNA and in one case as B-DNA, can now be

crystallized in any of the three standard DNA duplex conformations. In short alternating dG-dC sequences, effects such as cytosine methylation that destabilize the standard B-conformation of the DNA duplex will force the sequence to adopt a new conformation. If the sequence starts with a cytosine that new conformation is likely to be Z-DNA, especially if it is an alternating pyrimidine/purine sequence. If it starts with a guanine, the results presented in this paper suggest that the new conformation will be A-DNA. Therefore, hexanucleotides as a family, represent a common length in which the properties of A-, B-, and Z-DNA can be compared.

2.5 Materials and Methods

2.5.1 Oligonucleotide synthesis and crystallization

The hexanucleotides used for this study (Table 2.1) were synthesized using phosphoramidite chemistry on an Applied Biosystems synthesizer in the Center for Gene Research and Biotechnology at Oregon State University. Size exclusion chromatography on a Sephadex G-10 column was used to remove blocking groups, prematurely terminated oligonucleotides, and to desalt the oligonucleotides. The oligonucleotides were then concentrated by lyophilization, resuspended in dilute sodium cacodylate buffer, and used for crystallization without further purification.

All crystals were grown at room temperature by the vapor diffusion method from 40-60 μl droplets in a sitting drop setup. The initial crystallization conditions varied slightly between hexamers, but generally crystals were grown in solutions containing 1 mM DNA (single strand concentration), 15 mM sodium cacodylate buffer (pH 7), 0.5 to 5 mM

spermine, 1 to 20 mM $MgCl_2$, and 5% volume MPD (2-methyl-2,4-pentanediol). These solutions were equilibrated against a reservoir of 35% MPD. Large platelike crystals were obtained for most sequences after 7-14 days, although some, especially the less methylated sequences, took several weeks. The crystal of d(Gm^{5}CGm^{5}CGC) used for data collection was approximately 1.0 X 1.5 X 0.15 mm in size while the others were much smaller, especially in the first two dimensions.

2.5.2 Data collection

Single crystals were mounted in a glass capillary tube for x-ray diffraction data collection. Data from the d(Gm^{5}CGm^{5}CGC) crystal were collected on a Rigaku rotating anode diffractometer at 15°C using CuKα radiation. Diffraction data from the d(Gm^{5}CGCGC), d(Gm^{5}CCGGC), and d(GCCGGC) crystals were collected on a Siemens P4 diffractometer with a sealed tube copper source. The intensities for all data sets were corrected for polarization, Lorentz factor, and absorption. The crystals are all isomorphous and in the orthorhombic space group *C222$_1$* (Table 2.1).

2.5.3 Molecular replacement solution

Since d(Gm^{5}CGm^{5}CGC) is the first hexamer crystallized in this orthorhombic space group, it had a unique phase solution which required the determination of the molecule's orientation and position within the unit cell. Trial A-, B-, and Z-DNA models were generated using the program InsightII (Biosym Technologies, Inc.) with optimized helical parameters based on canonical fiber conformations (Arnott & Hukins, 1972). These models were then subjected to exhaustive rotation and translation searches in the *C222$_1$* space group with the program *AMoRe* (Navaza, 1994) using

8-4.0 Å data. These searches yielded an A-DNA solution that was significantly better than either the B-DNA or Z-DNA solutions. The best solutions for each model, as judged by the R-factors, were then subjected to ten rounds of rigid body refinement with 8-3.5 Å data in *AMoRe*. After rigid body refinement, the R-factor of the best A-DNA solution (41.2%) was 9.9% better than that of the best B-DNA solution (51.1%) and 9.2% better than the best Z-DNA solution (50.4%).

The refined molecular replacement solution for the A-DNA model was then subjected to successive rounds of positional refinement in X-PLOR starting with 15- 8 Å data and finishing with 8-3.2 Å data (Brünger, 1992). The R-value at this stage was 37.3%, but the initial electron density maps were not very well defined. An isomorphous crystal of a brominated derivative of the original hexamer, d(GBr^{5}CGm^{5}CGC), was used to help determine the position of the DNA duplex. The bromine positions were found by applying direct methods (SHELXTL PC^{TM}, Release 4.1, Siemens Analytical X-ray Instruments, Inc.) on the difference data set: d(GBr^{5}CGm^{5}CGC) minus d(Gm^{5}CGm^{5}CGC). These bromine positions served to verify the orientation and position of the best A-DNA solution provided previously by *AMoRe*. The best refined A-DNA model from *AMoRe* was then subjected to simulated annealing in X-PLOR using 8-2.7 Å data with a slow cool protocol starting at 3000°K followed by 80 cycles of positional refinement. This procedure substantially improved the fit of the model to the data, reducing the R-factor from 37.3% to 23%. The subsequent $2F_O$-F_C electron density map revealed that the refined A-DNA model now fit the data very well.

2.5.4 Structure refinement and analysis

All four crystal structures were refined using similar protocols. After simulated annealing of the model against data at 8-2.1 Å data for d(Gm^{5}CGm^{5}CGC), and against the 8-2.5 Å data for the other three structures, the overall and individual temperature factors were refined. Then solvent positions were identified by searching an (Fo-Fc) electron density map using PEAKMAX from the CCP4 suite of crystallographic programs (Collaborative Computational Project, Number 4, 1994). Difference electron density peaks above 2σ were treated as potential solvent molecules. The CCP4 program, WATPEAK, was used to select the crystallographically unique peaks and to measure their distance to potential hydrogen bonding sites on the DNA model. Potential positions that were within 2.4 to 3.6 Å of DNA hydrogen bonding partners were examined for their fit to both (F_O-F_C) and ($2F_O$-F_C) electron density maps. Several water molecules were located in special positions in the unit cell and were, therefore, fixed in position during the refinement. Each round of water addition was followed by a round of conventional position refinement and isotropic temperature factor refinement. Final R-factors for these A-DNA hexanucleotide structures ranged from 16.5% to 19.6% and are given in Table 1. The ($2F_O$-F_C) electron density maps for two base pairs of alternating d(Gm^{5}CGm^{5}CGC) structure are shown in Figure 2.8. The atomic coordinates for all structures will be deposited in both the Protein Data Bank of the Brookhaven National Laboratory (Bernstein *et al.*, 1977) and the Nucleic Acid Database (Berman *et al.*, 1992).

Helical parameters were determined with a program developed and kindly provided to our group by Marla Babcock at Rutgers University

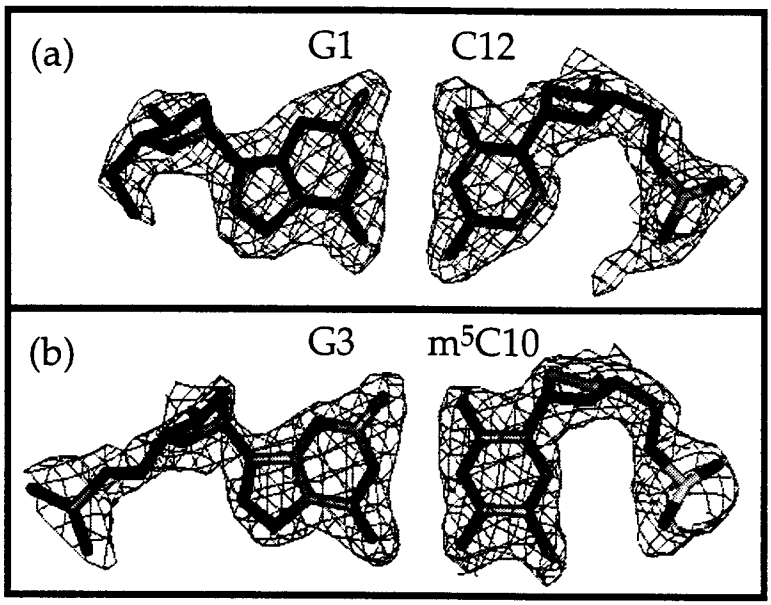


Figure 2.8. $2F_O$-F_C map at 2.1 Å resolution of (a) the G1-C7 terminal base pair and (b) the G3-m^{5}C10 base pair of the d(Gm^{5}CGm^{5}CGC) structure. The map was contoured at 1.0 σ. The water molecules have been deleted from the model.

(Babcock *et al*, 1994). The program NASTE (Nucleic Acid STructure Evaluation), a helical analysis program under development in our lab, was used to measure the backbone torsion angles and the ribose ring pseudorotation angles.

2.6 Acknowledgements

This work has been supported by grants from the American Cancer Society (NP-740C), and the National Science Foundation (MCB-9304467). BHMM was supported by an AASERT grant (N00014-9311355) from the Office of Naval Research and GPS by a postdoctoral fellowship from the American Cancer Society (PF-3749). We thank Dr. Loren D. Williams for his help in collecting data on the sequence d(Gm^{5}CGm^{5}CGC). We would also like to thank Dr. Todd F. Kagawa for assistance with the early work on the molecular replacement solution, Monika Ivancic for working on the structure refinement of one hexamer, the rest of the Ho lab for comments on the manuscript.

2.7. References

Arnott, S. & Hukins, D. W. L. (1972). Optimized parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* **47**, 1504-1509.

Babcock, M. S., Pednault, E. P. D. & Olson, W. K. (1994). Nucleic acid structure analysis: mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* **237**, 125-156.

Behe, M. & Felsenfeld, G. (1981). Effects of methylation on a synthetic polynucleotide: The B-Z transition in poly (dG-m^{5}C)•poly(dG-m^{5}C). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 1619-1623.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, D., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). The Nucleic Acid Database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751-759.

Berstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. F., Brice, M. D., Rodgers, J. D., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Bingman, C., Li, X., Zon, G. & Sundaralingam, M. (1992). Crystal and molecular structure of d(GTGCGCAC): investigation of the effects of base sequence on the conformation of octamer duplexes. *Biochemistry*, **31**, 12803-12812.

Brünger, A. (1992). *X-PLOR Manual, Version 3.1.* Yale University, New Haven, USA.

Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

Calladine, C. R. & Drew, H. R. (1984). A base-centered explanation of the B-to-A transition in DNA. *J. Mol. Biol.* **178**, 773-782.

Chandrasekaran, R., Wang, M., He, R.-G., Puigianer, L. C., Byler, M. A., Millane, R. P. & Arnott, S. (1989). A re-examination of the crystal structure of A-DNA using fiber diffraction data. *J. Biomol. Struct. Dynam.* **6**, 1189-1202.

Cruse, W. B. T. , Salisbury, S. A., Brown, T., Cosstick, R., Eckstein, F. & Kennard, O. (1986). Chiral phosphorothioate analogues of B-DNA, the crystal structure of Rp-d | Gp(S)CpGp(S)CpGp(S)C | . *J. Mol. Biol.* **192**, 891-905.

Dickerson, R. E. (1992). DNA structure from A to Z. In *Methods in Enzymology, Vol. 211* (Lilley, D. M. J. & Dahlberg, J. E. eds.), pp. 67-111.

Dickerson, R. E., Goodsell, D. S. & S. Neidle. (1994). "...the tyranny of the lattice...". *Proc. Nat. Acad. Sci., U.S.A.* **91**, 3579-3583.

Duocet, J., Beoit, J. P. Cruse, W. B. T., Prange, T. & Kennard, O. (1989). Coexistence of A- and B-form DNA in a single crystal lattice. *Nature (London)*, **337**, 190-192.

Frederick, C. A., Saal, D., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. & Rich, A. (1987). The crystal structure of d(GGm^{5}CCGGCC): the effect of methylation on A-DNA structure and stability. *Biopolymers*, **26**, S145-S160.

Frederick, C. A., Quigley, G. J., Teng, M.-K., Coll, M., van der Marel, G. A., van Boom, J. H., Rich, A. & Wang, A. H.-J. (1989). Molecular structure of an A-DNA decamer d(ACCGGCCGGT). *Eur. J. Biochem.* **181**, 295-307.

Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J.H. & Rich, A. (1982). Molecular structure of (m^{5}CG)$_3$: the role of the methyl group on 5-methyl-cytosine in stabilizing Z-DNA. *Nucleic Acids Res.* **10**, 7879-7892.

Heinemann, U., Alings, C. & Hahn, M. (1994). Crystallographic studies of DNA helix structure. *Biophysical Chem.* **50**, 157-167.

Ho, P. S., Kagawa, T. F., Tseng, K., Schroth, G. P. & Zhou, G. (1991). Prediction of a crystallization pathway for Z-DNA hexanucleotides. *Science*, **254**, 1003-1006.

Jain, S., Zon, G. & Sundaralingam, M. (1987). The potentially Z-DNA forming sequence d(GTGTACAC) crystallizes as A-DNA. *J. Mol. Biol.* **197**, 141-145.

Jain, S. & Sundaralingam, M. (1989). Effect of the crystal environment on conformation of the DNA duplex. *J. Biol. Chem.* **264**, 12780-12784.

Jain, S., Zon, G. & Sundaralingam, M. (1989). Base only binding of spermine in the deep groove of the A-DNA octamer d(GTGTACAC). *Biochemsitry*, **28**, 2360-2364.

Kennard, O., Cruse, W. B. T., Nachmann, J., Prange, T., Shakked, Z. & Rabinovich, D. (1986). Ordered water in an A-DNA octamer at 1.7Å resolution. *J. Biomol. Struct. Dynam.* **3**, 623-647.

Kennard, O. & Salisbury, S.A. (1993). Oligonucleotide X-ray structures in the study of conformations and interactions of nucleic acids. *J. Biol. Chem.* **268**, 10701-10704.

McCall, M., Brown, T. & Kennard, O. (1985). The crystal structure of d(GGGGCCCC): a model for poly (dG).poly(dC). *J. Mol. Biol.* **183**, 385-396.

Navaza, J. (1994). *AMoRe*: an automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.

Quadrifoglio, F., Manzini, G. & Yathindra, N. (1984). Short oligonucleotides with d(G-C)n sequences do not assume left-handed conformation in high salt conditions. *J. Mol. Biol.* **175**, 419-423.

Ramakrishnan, B. & Sundaralingam, M. (1993). Crystal packing effects on A-DNA helix parameters: a comparative study of the isofroms of the tetragonal and hexagonal family of octamers with differing base sequences. *J. Biomol. Struct. Dynam.* **11**, 11-26.

Saenger, W., Hunter, W. N. & Kennard, O. (1986). DNA conformation is determined by the economics in the hydration of phosphate groups. *Nature (London)*, **324**, 385-388.

Schneider, B., Cohen, D. M., Schleifer, L., Srinivasan, A. R., Olson, W. K. & Berman, H. M. (1993). A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophysical J.* **65**, 2291-2303.

Takusagawa, F. (1990). The crystal structure of d(GTACGTAC) at 2.25Å resolution: are the A-DNA's always unwound approximately 10° at the C-G steps? *J. Biomol. Struct. Dynam.* **7**, 795-809.

Timsit, Y. & Moras, D. (1992). Crystallization of DNA. In *Methods in Enzymology, Vol. 211* (Lilley, D.M.J. & Dahlberg, J.E. eds.), pp. 409-429.

Verdaguer, N., Aymamí, J., Fernández-Forner, D., Fita, I., Coll, M., Huynh-Dinh, T., Igden, J. & Subirana, J. A. (1991). Molecular structure of a complete turn of A-DNA. *J. Mol. Biol.* **221**, 623-635.

Wang, A. H.-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van der Marel, G.A., van Boom, J.H. & Rich, A. (1979). Molecular Structure of a left handed double helical DNA fragment at atomic resolution. *Nature (London)*, **282**, 680-686.

Wang, A. J.-H., Fujii, S., van Boom, J. H. & Rich, A. (1982). Molecular structure of the octamer d(G-G-C-C-G-G-C-C): modified A-DNA. *Proc. Nat. Acad. Sci., U.S.A.* **89**, 534-538.

Westof, E. (1988). Water: an integral part of nucleic acid structure. *Ann. Rev. Biophys. Biophys. Chem.* **17**, 125-144.

Chapter 3

Extra-helical 3'-overhangs Form d[G*(G·C)] Base Triplets in the Minor Grooves of B-DNA Nonamers

Blaine H. M. Mooers and P. Shing Ho*

Formatted for submission

3.1 Summary

Bases overhanging the ends of oligonucleotide duplexes form interduplex hydrogen bonds which facilitate the crystallization of protein-DNA complexes and interesting DNA structures. We have solved the crystal structures of d(GCAATTGCG) and d(GCGTACGCG), which are characterized by eight base pairs of standard B-DNA and single guanines overhanging the 3'-ends. The octamer duplex section is able to stack end-to-end with a neighboring duplex because the overhanging guanines twist out of the helical stacks to avoid collisions with the neighboring duplex. Each extra-helical guanine projects into the minor groove of a neighboring duplex and forms an unusual d[G*(G·C)] base triplet. The sequence d(GCAATTGCG) crystallizes in two different unit cells, and the structures are solved to resolutions of 2.25 Å and 2.3 Å. The structure of the sequence d(GCGTACGCG) is solved to a resolution of 2.5 Å. In all three structures, the base triplets are similar . We discuss the geometry of stacking B-DNA octamer duplexes and demonstrate how a -72° twist angle at the junction between duplexes places 3'-overhangs in the minor groove and 5'-overhangs in the major groove of a neighboring duplex.

3.2 Introduction

Non-Watson-Crick base pairs frequently stabilize the tertiary interactions between nucleic acid duplexes. For example, many of the 26 different non-Watson-Crick base pairs that share two cyclic hydrogen bonds stabilize folded RNAs (Rich and RafBhandary, 1976). In particular, a

symmetric r(G·G) base pair stabilizes the "L" shaped fold of *E. coli* $tRNA^{Cys}$ by forming hydrogen bonds between the N2 and N3 atoms on minor groove faces of guanine 15 and guanine 48 (Hou *et al.*, 1993). Non-Watson-Crick base pairs are also found in DNA mismatches and in unusual DNA structures, either as base pairs or as components of base triples or quadruples (Brown & Hunter, 1997). Here, we report the structure of a symmetric d(G·G) base pair in an unusual d[G*(G·C)] base triplet. The guanine 3'-overhangs project into the minor grooves of B-DNA duplexes where they form triplets in the crystal structures of two nonamer sequences.

Base triplets can also form in the wide and shallow minor groove of RNA double helical regions when a single stranded region approaches a double helical region. For example, residue A153, the third residue in the GAAA tetraloop in the group I self-splicing intron, forms a base triplet in the minor groove of a Watson-Crick r(C·G) base pair by hydrogen bonding with the N2 amino group of G250 (Cate *et al.*, 1996a). Likewise, the second and third adenines in the GAAA tetraloop in the crystal structure of the hammerhead ribozyme form base triplets in the minor groove of r(C·G) base pairs in the double helical stem of the tetraloop of a neighboring ribozyme (Pley *et al.*, 1994). In contrast, base triplet formation in the narrow minor grooves of the B-DNA is surprising, although not without precedent. The two base pairs at each end of the duplex in crystal structures of the B-DNA dodecamers starting with a 5'-CGY... sequence (Dickerson *et al.*, 1987) and the B-DNA octamer d(CGCTAGCG) (Tereshko *et al.*, 1996) form two interduplex symmetric d(G·G) base pairs in the minor grooves of neighboring duplexes. In the dodecamer and the octamer, the guanines in these d(G·G) base pairs are in the *anti* conformation, and their deoxyribose are in the C2'-endo conformation which is typical form B-DNA. On the other hand, the

overhanging guanines are extra-helical because they have rotated into the *syn* conformation and their deoxyribose are in the C3'-endo conformation which is not typical of B-DNA. The unusual conformation of the overhanging guanines permits the nonamers to stack end-to-end whereas the octamer and dodecamer duplexes do not stack end-to-end.

The end-to-end stacking of two B-DNA duplexes that are not a full or half helical turn in length requires changing the twist angle at the junction between duplexes from the 36° twist expected between two adjacent base pairs in the B-form. Several "underwound" stacks have been stabilized with overhanging nucleotides. For example, 5'- and 3'-overhangs stabilize the end-to-end stacking of the central octamer duplex region of the decamer d(CGCAATTGCG) (Spink *et al.*, 1995). The 3'-overhanging guanine projects into the minor groove whereas the 5'-overhanging cytosine projects into the major groove. The terminal base pair at the junction between duplexes is related by a rotation angle of about -72°. If the 5'-overhanging base is a purine, it will clash with the terminal Watson-Crick base pair of the neighboring duplex. This clash can be avoided by the staggered, non-coaxial stacking of duplexes, as in the crystal structure of the B-DNA nonamer d(GCGAATTCG) (Van Meervelt *et al.*, 1995; Vlieghe *et al.*, 1996a) and the B-DNA decamer d(GGCCAATTGG) (Vlieghe *et al.*, 1996b).

We used x-ray crystallography to determine how guanine 3'-overhangs stabilized the packing and structure of the octamer sequences d(GCAATTGC) and d(GCGTACGC). The first octamer sequence has a central AT rich region found in the sequences of other B-DNA crystal structures, so the crystallization of the nonamer sequence d(GCAATTGCG) as B-DNA is not surprising. On the other hand, we have crystallized the second octamer as A-DNA so the crystallization of the nonamer d(GCGTACGCG) as B-DNA

was not assured. The crystal structures of these two nonamer sequences demonstrate that only guanine 3'-overhangs are necessary to stabilize the coaxial stacking of B-DNA octamer duplex regions. This finding expands the repertoire of structures that can be used to design stable multibranched nucleic acids and protein*DNA complexes.

3.3 Results

We have solved the crystal structures of the DNA sequences d(GCAATTGCG) and d(GCGTACGCG). The underlined sequences form octamer duplexes which crystallized in the B-form. One of the most interesting features of these crystal structures is the d[G*(G·C)] base triplets formed by the 3'-guanines overhanging each end of the central octamer duplexes. Each overhanging guanine projects into the minor groove of a neighboring duplex where it forms a base triplet with the terminal d(G·C) base pair (Figure 3.1c). In contrast, 5'-overhangs of the nonamer d(GCGAATTCG) and the decamer d(GGCCAATTGG) form two types of d[(G·C)*G] base triplets (Figure 3.1 a & b) in the major groove (Van Meervelt *et al.*, 1995, Vlieghe *et al.*, 1996b). We report the structural features of the d[G*(G·C)] base triplet and the influence of the triplet on duplex stacking, crystal packing, solvent interactions, and the helical features of the B-DNA duplex regions.

3.3.1 Structure of the d[G*(G·C)] base triplet

In the d[G*(G·C)] base-triplets reported here, the third strand guanine is hydrogen bonded to the guanine of the Watson-Crick base pair via the N2 amino group and the N3 nitrogen in the purine rings to form a d(G·G) base

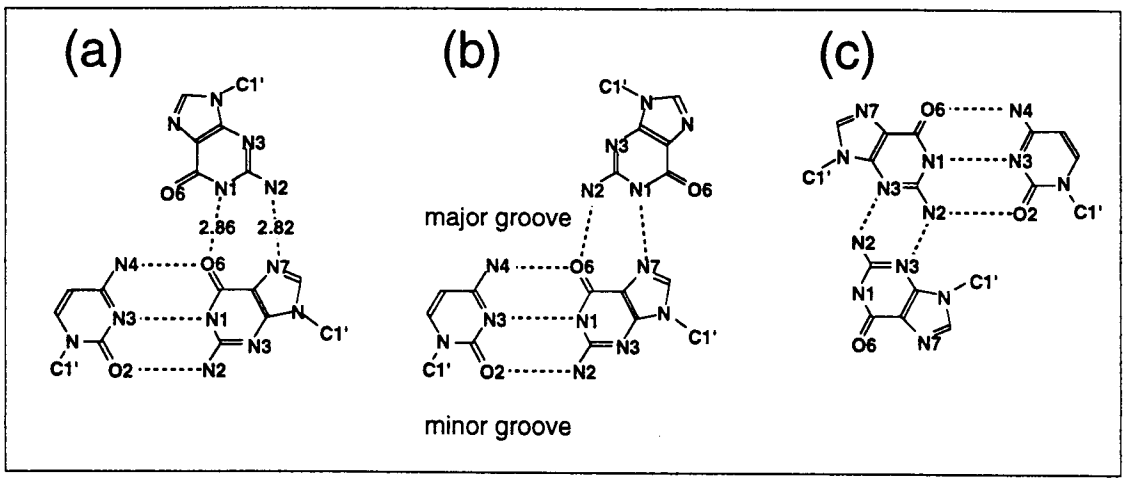


Figure 3.1. Hydrogen bonding patterns in DNA base triplets. Standard d[(C·G)*G] base-triplets are formed between Watson-Crick base-pairs and a guanine in Hoogsteen (a) and reverse Hoogsteen (b) arrangements. The triplet in (a) was observed in the crystal structure of d(GCGAATTCG) (Van Meervelt *et al.*, 1995) (average hydrogen bond distances (Å) from this crystal structure are indicated). The triplets in both (a) and (b) were observed in the crystal structures of d(GGCCAATTGG) (Vlieghe *et al.*, 1996b). (c) The d[G*(G·C)] triplet observed in the current structures has the guanine sitting in the minor groove.

Table 3.1. Overhanging guanine base-to-base hydrogen bond distances in the crystal structures of d(GCGTACGCG) and of d(GCAATTGCG). The dihedral angle between the overhanging guanine and the guanine of the Watson-Crick base pair is given in degrees. This angle is the average of the complements of two diheral angles: C8-N1-N1*-C8* and C8-C2-C2*-C8*. The translational symmetry mates are marked by an asterisk.

	Hydrogen Bond Distances (Å)		Base Plane Angle
d(GG) Base Pair	N2--N3	N3--N2	(°)
d(GCGTACGCG)			
G9-G1*	2.70	2.84	36.93
d(GCAATTGCG)-A			
G9-G28	2.99	2.88	40.08
G18-G19*	3.20	3.01	34.75
G27-G10	2.72	2.73	40.93
G36-G1*	2.74	2.77	34.10
d(GCAATTGCG)-B			
G9-28	2.92	3.00	37.04
G18-G19*	2.84	2.83	38.17
G27-G10	2.78	2.86	33.33
G36-G1*	2.83	3.17	38.17

pair (Figure 3.1C). The resulting d(G·G) base pair has also been found at the lattice contacts between two B-DNA duplexes in crystal lattices (Wing *et al.*, 1980, Spink *et al.*, 1995, Tereshko *et al.*, 1996), but in these cases the d(G·G) base pair was in a four base hydrogen bond network rather than a true base triplet. Thus, the nonamer structures reported here show that the base triplet formed in the minor groove can stand on its own, without the need for hydrogen bonding interactions with a fourth base.

The N2--N3 potential hydrogen bond lengths have a mean value of 2.89 Å, with a range of 2.72 Å to 3.20 Å, for the eight crystallographically unique triplets in the two crystal structures of the nonamer d(GCAATTGCG) (Table 3.2). The two unique N2--N3 distances in the structure of d(GCGTACGCG) are 2.70 Å and 2.84 Å. These distances overlap the N2-N3 distances of d(G·G) base pairs found in high resolution crystal structures of (a) parallel stranded duplexes of d(CpG) (2.87 Å and 2.97 Å) (Cruse *et al.*, 1983; Coll *et al.*, 1987), (b) cyclic diguanylic acid (3.01 Å) (Egli *et al.*, 1990), (c) a 1:1 complex of 9-ethylguanine with 1-methylcytosine (3.01 Å) (O'Brien, 1967), and (d) guanine hydrochloride (3.08 Å) (Broomhead, 1951).

The minor groove of B-DNA is too narrow (~6 Å wide, nucleus to nucleus distance) to accommodate a guanine that is coplanar with a Watson-Crick base pair (Figure 3.3a & b). Instead, the guanine of the third strand tilts at an average angle of 37.1° in order to approach close enough to the guanine of the Watson-Crick base pair to form two hydrogen bonds (Fig. 3.2a). This angle compares well with the angles of 38.3°and 33.7° formed by the d(G·G) base pairs in the two crystal structures of parallel stranded DNAs (Fig.3.2b), where neither guanine was constrained by packing in a minor groove (Cruse *et al.*, 1983; Coll *et al.*, 1987). Recent *ab initio* quantum chemical methods

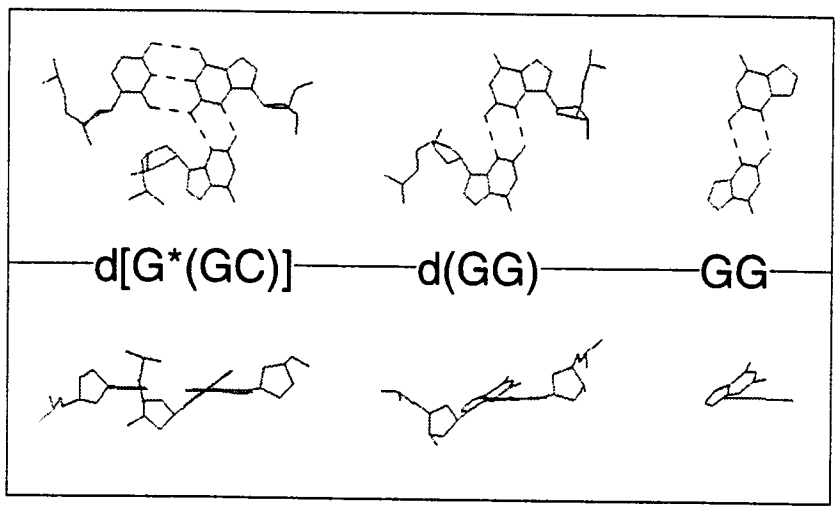


Figure 3.2. Top views (a) and side views (b) of a d[G*(G·C)] base triplet on the left, a d(G·G) base pair in the middle, and a G·G base pair on the right. The side views emphasizes the large angle between two guanine bases in a triplet, in a base pair with deoxyribose backbones, and in a base pair without backbones. The d[G*(G·C)] base triplet is from the crystal structure of d(GCGTACGCG). The d(G·G) base pair is from the crystal structure of d(CpG) which crystallized in parallel-stranded duplex (Coll *et al.*, 1987). The G·G base pair was calculated by *ab initio* quantum mechanical methods (Sponer *et al.*, 1996).

Figure 3.3. (a) A view down the helix axis of a 2Fo-Fc electron density map of the d[G*(G·C)] triplet from the second monoclinic form of d(GCAATTGCG). The electron density map was contoured at the one sigma contour level with diffraction data to a resolution of 2.1 Å. (b) Side view of one overhanging guanine projecting into one minor groove. The three potential hydrogen bonds that may stabilize the extra-helical guanine are represented by dashed lines. (c) The close approach of two O1P phosphates of two antiparallel overhanging guanines is stabilized by a hydrated magnesium complex. The four water oxygens that complete the octahedral coordination sphere of the magnesium are represented by the letter W.

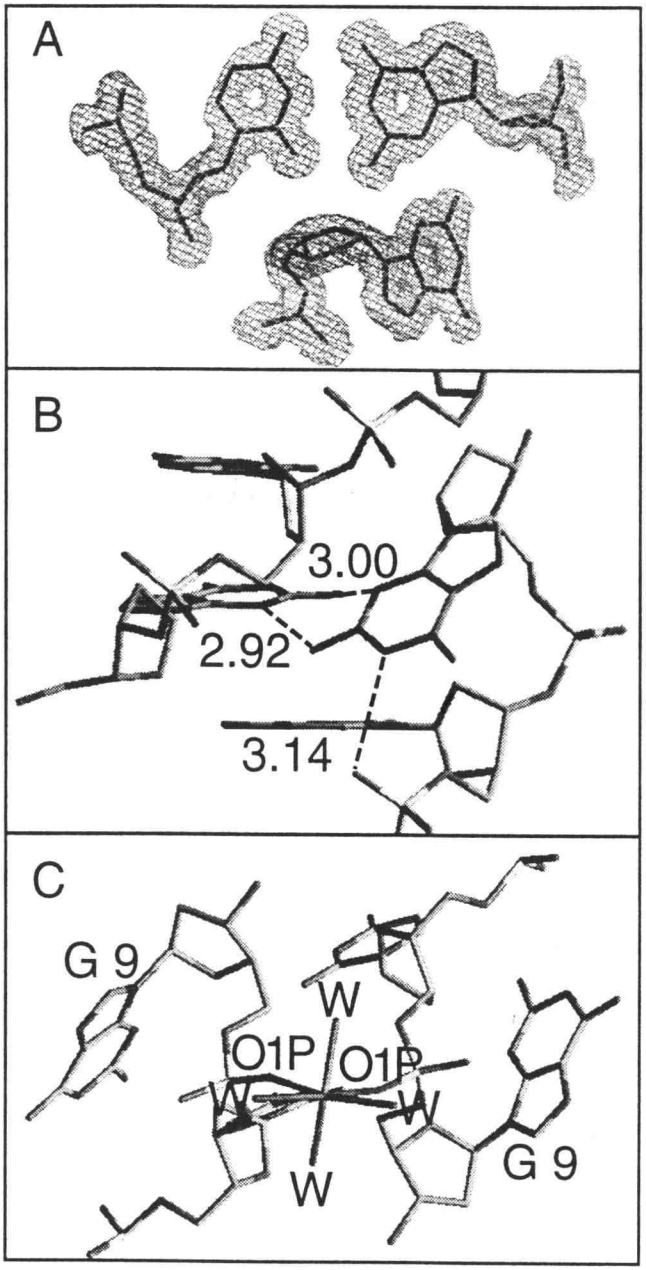


Figure 3.3.

produced a theoretical structure of a G·G base pair (Fig. 3.2c) with an angle of 42.9° between guanine base planes (Sponer *et al.* 1996). The physical basis of this large tilt angle has not been thoroughly explored, but the slight sp^3 character of the guanine amino group is thought to be one of several factors favoring hydrogen bond formation between tilted guanines (Jiri Sponer, personal communication). In the crystal structures of the nonamers, the backbone of the overhanging guanosine accommodates the large tilt of the base by a rotation about the glycosidic bond from the *anti* conformation (χ = -108°) to the high *anti* conformation (χ = -84° to -69°) , which is actually part of the *syn* range, and by a change in the sugar pucker of the deoxyribose ring from C2'-endo to C3'-endo.

Intramolecular interactions between the overhanging guanine (GUA 9) and the its 3'-neighboring cytosine (CYT 8) stabilize its extra helical packing in the minor groove. The overhanging guanine had van der Waals contacts with the C5 and C6 atoms of CYT 8. In addition, a hydrogen bond links the guanine's N1 and the CYT 8's O2P phosphate oxygen (Figure 3.3b). Thus, the overhanging guanines form hydrogen bonds with bases in the terminal base pairs of both duplexes at a junction between helices. The deoxyribose 03′ atoms of the overhanging guanines are too far (4-5 Å) from the amino N2 of the guanine in the second base pair of the neighboring duplex to form a hydrogen bond, as in the structure of d(CGCAATTGCG) (Spink *et al.*, 1995).

3.3.2 Triplet formation and duplex stacking

In all three crystal structures, the eight base pair duplexes form B-DNA and are stacked end-to-end to form columns of duplexes. Although the duplexes are stacked with their helical axes aligned, they do not quite form

Figure 3.4. Comparison of the stacking of B-DNA octamer duplexes. (a) Two d(GCAATTGCG) duplexes (top) crystallized in a staggered fashion to avoid a clash between the 5'-overhanging guanine in black and the terminal Watson-Crick base pair of the duplex below it (Van Meervelt *et al.*, 1995). The two Watson-Crick base pairs (bottom) at the junction between two duplexes are related by a -72° rotation angle with staggered helical axes. (b) A model of two B-DNA duplexes stacked in a pseudocontinuous fashion (top) with a rotation angle of 36° between the two terminal base pairs at the junction between duplexes (bottom). (c) Two duplexes of d(GCAATTGCG) from the second monoclinic crystal structure stacked with their helical axes aligned (top) and with the terminal base pairs related by a rotation angle of -72° (bottom).

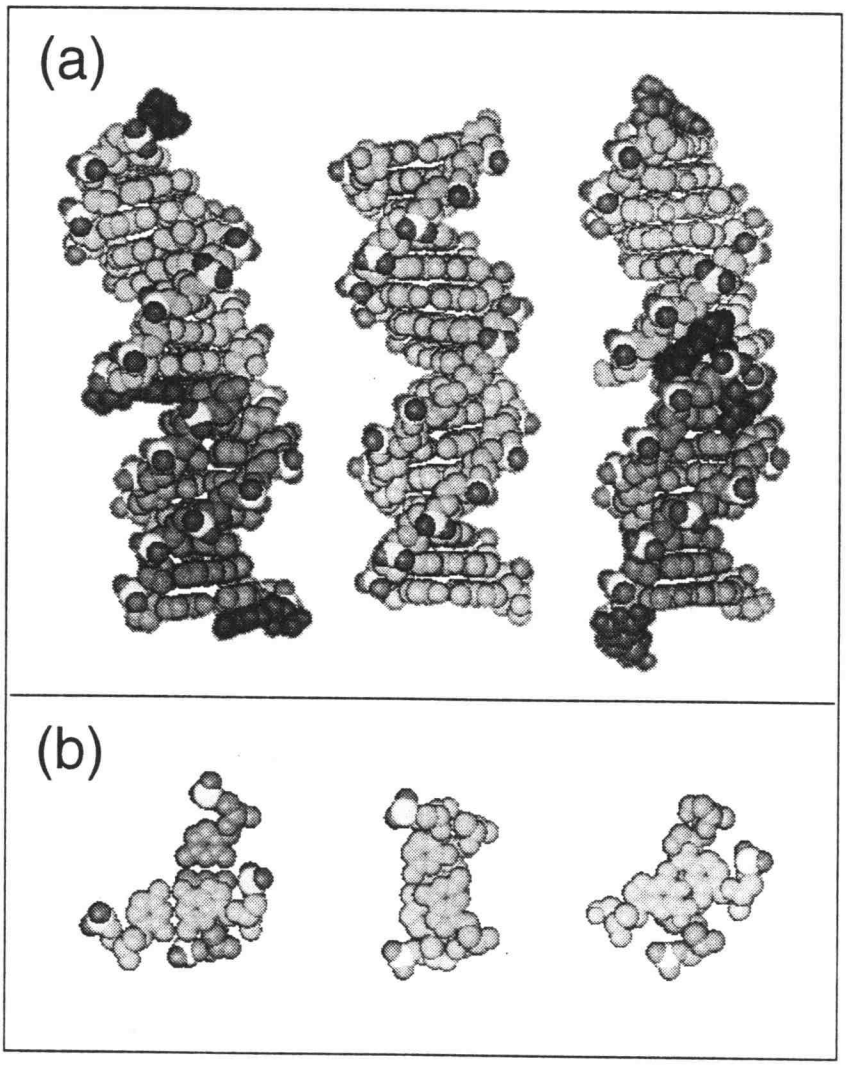


Figure 3.4.

pseudo-continuous helices (Figure 3.4c). The angle between the stacked terminal base-pairs of two duplexes is about -72°, as opposed to about +36° seen in the pseudo-continuous helices in the crystal structures of B-DNA decamers (Figure 3.4b). In this orientation, the extension of each strand by one guanine in the 3′ direction causes a collision between the overhanging guanine and the backbone of the neighboring duplex. This collision is avoided by unstacking of the dangling guanine and its projection into the minor groove (Figure 3.4c).

3.3.3 Hydrated magnesiums and backbone contacts stabilize triplets

The two 3′-overhanging nucleotides at the junction between duplexes are on the same side of the stack, and their phosphates pass each other in opposing directions at a shorter a distance (~3 Å) than one would expect due to the electrostatic repulsion between the negatively charged phosphates. In the crystal structure of d(GCGTACGCG), the two phosphate O1P atoms are bridged by a $Mg^{+2}(H_2O)_4$ ion which also sits on the crystallographic 4_2 symmetry axis (Fig. 3.5a). In the second structure of d(GCAATTGCG), the O1P atoms of the overhanging guanines in the middle of the asymmetric unit (GUA 9 and GUA 27) are bridged by a $Mg^{+2}(H_2O)_3$ ion while in the first structure, the O1P atoms of the overhanging guanines (GUA 18 and GUA 36) at the ends are bridged by a $Mg^{+2}(H_2O)_3$ ion. The remaining overhangs in both structures are bridged by electron densities which are too poorly defined to be modeled as metal complexes, although metals probably occupy these sites. The coordination of two close approaching phosphate oxygens by one magnesium has been observed in other nucleic acid crystal structures (Holbrook *et al.*, 1977; Cate *et al.*, 1996a,b).

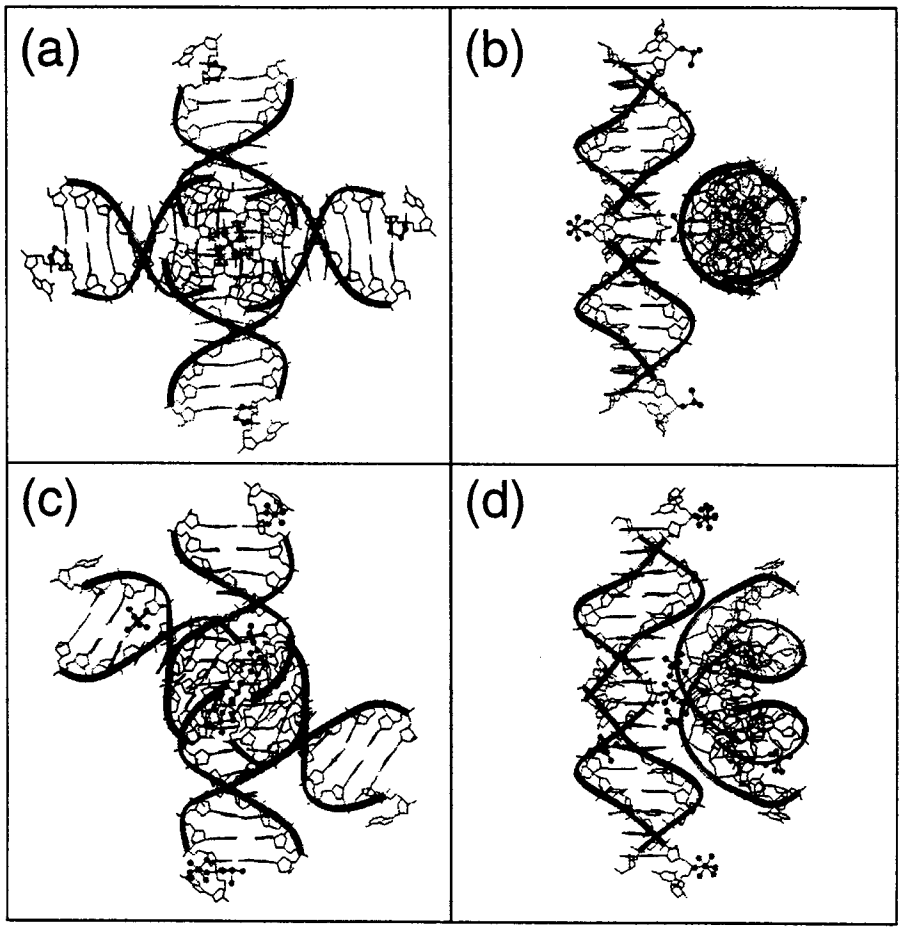


Figure 3.5. The packing of four DNA duplexes in the crystal structures d(GCGTACGCG) and d(GCAATTGCG). The backbones are outlined with ribbons. (a) Stacks of duplexes of the sequence d(GCGTACGCG) pack together in the tetragonal lattice to form a cross that has 90° between the arms. (b) Rotation by 90° about the vertical reveals the contact between the stacks which is mediated by one magnesium water complex. (c) Stacks of duplexes of the sequence d(GCAATTGCG) pack in a monoclinic lattice to form a cross that has 63° between the arms. (d) Rotation by 90° about the vertical reveals the contact between the stacks which is mediated by three magnesium water complexes and which induces bends in the axes of the duplexes.

The nonamer structures, however, are of higher resolution and thus provide views of all or most of the coordinating waters of the bridging magnesium's octahedral coordination sphere.

In addition to the phosphates of the overhangs, the deoxyribose rings of the overhangs pass very close to the deoxyribose rings of cytosine 8 in the opposing strands (Figure 3.3b). The four deoxyribose rings have shape complementarity, and their close approach is stabilized by favorable van der Waals interactions, the burial of hydrophobic surface area, and the magnesium-water complex bridging the phosphates. A similar structural motif involving four ribose rings in RNA has been called a "ribose zipper" (Cate *et al.*, 1996a; Pley et al., 1994), which is further stabilized by reciprocal O2' to O2' hydrogen bonds and O2′ to base O2 or N3 hydrogen bonds.

3.3.4 Triplets do not limit crystal geometry

The d[G*(G·C)] triplets linking together the nonamers do not restrict the crystal packing of the columns of duplexes to one particular geometry. In high concentrations of magnesium, the sequence d(GCAATTGCG) crystallized in the columns of duplexes that are crossed at an angle of 62° in the first monoclinic lattice (Figure 3.5c) and at an angle of 63° in the second lattice. In low concentrations of magnesium, the columns of duplexes of d(GCGTACGCG) are crossed at an angle of 90° (Figure 3.6a).

In contrast, B-DNA duplexes with 5-overhangs crystallize with their helical axes aligned parallel to adjacent duplexes in the lattice. This is a consequence of the staggered stacking found in the structures with guanine 5'-overhangs: d(GCGAATTCG) (Van Meervelt *et al.*, 1995) (Figure 3.4a) and d(GGCAATTCGG) (Vlieghe *et al.*, 1996b). However, duplexes of the decamer d(CGCAATTGCG) (Spink *et al.*, 1995) do not have this limitation because

they have a smaller pyrimdine 5-overhang. As a result, they are almost coaxially stacked. The crystal packing of this decamer suggests that nonamers with 3'-overhangs have the potential to crystallize with all duplexes being parallel.

3.3.5 Overhang phosphates coordinate magnesiums which stabilize lattice contacts

The two closely approaching phosphates of the overhangs provide coordination sites for hydrated magnesiums which in turn stabilize the contacts between neighboring stacks of duplexes. The magnesium-water complexes link the minor groove face of a helical junction to the major groove face or backbones of helical junctions in neighboring layers of stacked duplexes. In the lattice of d(GCGTACGCG), the crossover point of two stacks is mediated by the $Mg_2^+(H_20)_4$ complex mentioned earlier (Figure 3.3b) which links the minor groove face of one junction to the major groove face of the junction in the adjacent layer of duplexes via four hydrogen bonds between the water ligands and the guanines (Figure 3.5b). This is the major interaction between columns of duplexes because these columns duplexes have no side-to-side contacts. The columns in a layer are spaced 27.51 Å apart, which is greater than the ~20 Å diameter of standard B-DNA duplexes.

In both crystal structures of d(GCAATTGCG), hydrated magnesium complexes span two types of crossover points (A and B), neither of which is identical to the crossover in the lattice of d(GCGTACGCG). At a type A crossover, the helical junction is pinned between two layers of stacks while at a Type B crossover, the helical junction is pulled ~2 Å to one side by intermolecular hydrogen bonds mediated by hydrated magnesium complexes (Figure 3.5d). At crossover type A, hydrated magnesium complexes link the major groove surface of the helical junction to the minor

groove surface of a neighboring column of duplexes while the minor groove surface has solvent mediated contacts with the backbones of a second column of duplexes. At crossover type B, the major groove surface of the junction faces a solvent channel while one to three hydrated magnesium complexes link the minor groove surface to the major groove surface of a junction in a neighboring column of duplexes.

3.3.6 Global helical features of the duplexes

The overall features of the backbones and helical structure of the octamer duplex sections were characteristic of classical B-DNA in all three structures. The sugar puckers of the deoxyribose rings in the octamer duplex sections are in the C2'-endo family, which is typical for B-DNA. The base pairs in the ocatmer duplex in formed by d(GCGTACGCG) are related by an average rotation angle of 35.9° about the helical axis (helical twist) and an average translation aonf the helical axis of 3.46 Å/bp. Likewise, the duplexes of d(GCAATTGCG)-A have average helical twists of 36.0° and 35.4° along with average helical rises of 3.40 Å and 3.37 Å, and duplexes of d(GCAATTGCG)-B have average helical twists of 35.2° and 36.1° along with average helical rises of 3.39 Å and 3.37 Å. These values are typical of classical B-DNA (Saenger, 1983).

All five duplexes have wide major grooves and narrow and deep minor grooves which is also characteristic of B-DNA. The minor grooves are widest at the ends of the duplexes and are narrowest in the middle of the duplex as seen in most crystal structures of B-DNA. The presence of a guanine base in the minor groove may be expected to widen the minor groove near the ends of the duplex, but this is not the case as shown by comparing the minor groove widths near the ends of the duplex at the same

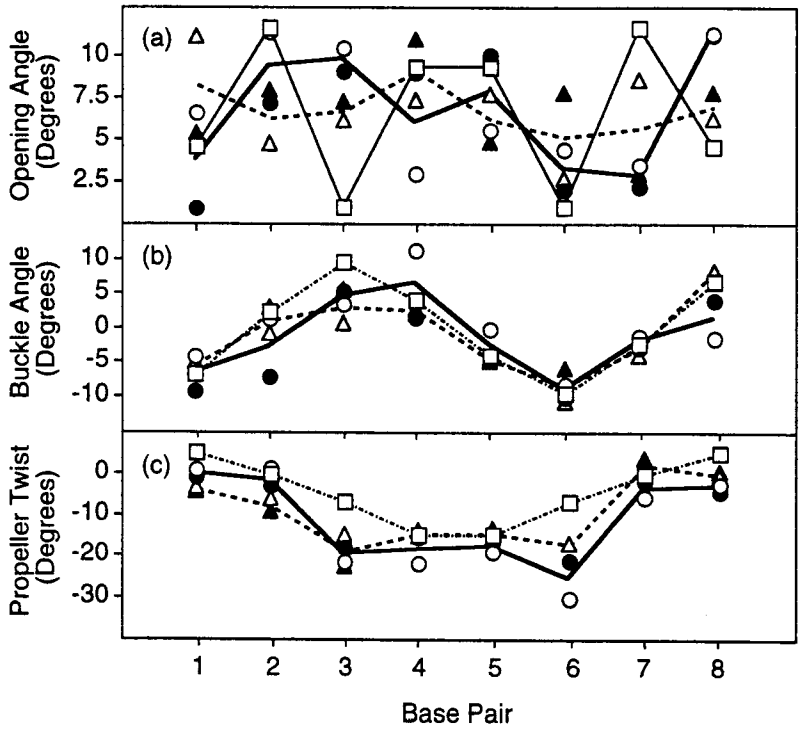


Figure 3.6. The base-pair parameters (a) opening angle, (b) buckle angle, and (c) propeller twist of the crystal structures of d(GCGTACGCG) and the d(GCAATTGCG) are plotted by base pair. Opening angle is the rotation angle between two bases in a pair about their common vertical z-axis. Buckle angle is the rotation between two bases about their common short x-axis. Propeller twist is the angle between two base s about their common long axis. These structures were analyzed with program by Babcock *et al.* (1994). The values for the structure of d(GCGTACGCG) are plotted with open squares. The values for the unique duplexes in the crystals of d(GCAATTGCG) are plotted in opposing directions. The values of the first crystal of d(GCAATTGCG) are plotted with circles--open circles for residues 1-18 and the closed circles for residues 19-36. Their mean values are plotted with a solid line. The values of the second crystal of d(GCAATTGCG) are plotted with triangles--open triangles for residues 1-18 and the closed triangles with residues 19-36. Their mean values are plotted with a dashed line.

Figure 3.7. Plots by base step of the base step parameters (a) helical twist, (b) helical rise, (c) roll angle, (d) slide, and (e) tilt angle of the crystal structures of d(GCGTACGCG) and the d(GCAATTGCG). The Cartesian coordinate frame of each base pair step was used to calculate the parameters with the program by Babcock *et al.* (1994). The x-axis is oriented between the short axes of the base pairs, and the y-axis is oriented between the long axes of the base pairs. Helical twist is the rotation angle between two base pairs about their common vertical z-axis. Helical rise is the distance between two base pairs along the vertical z-axis. Roll angle is the rotation about their common long axis. Positive values correspond to opening of the base pairs towards the minor groove. Slide is the translation of one base pair relative to another along their common y-axis. Tilt is the rotation about the common x-axis and results in the opening of the base pairs towards one of the two backbones. The values for the structure of d(GCGTACGCG) are plotted with open squares. The values for the unique duplexes in the crystals of d(GCAATTGCG) are plotted in opposing directions because the duplexes are more similar when aligned in this way on account of lattice induced bending of the helix axis. The values of the first crystal of d(GCAATTGCG) are plotted with circles--open circles for residues 1-18 and the closed circles for residues 19-36. Their mean values are plotted with a solid line. The values of the second crystal of d(GCAATTGCG) are plotted with triangles--open triangles for residues 1-18 and the closed triangles with residues 19-36. Their mean values are plotted with a dashed line.

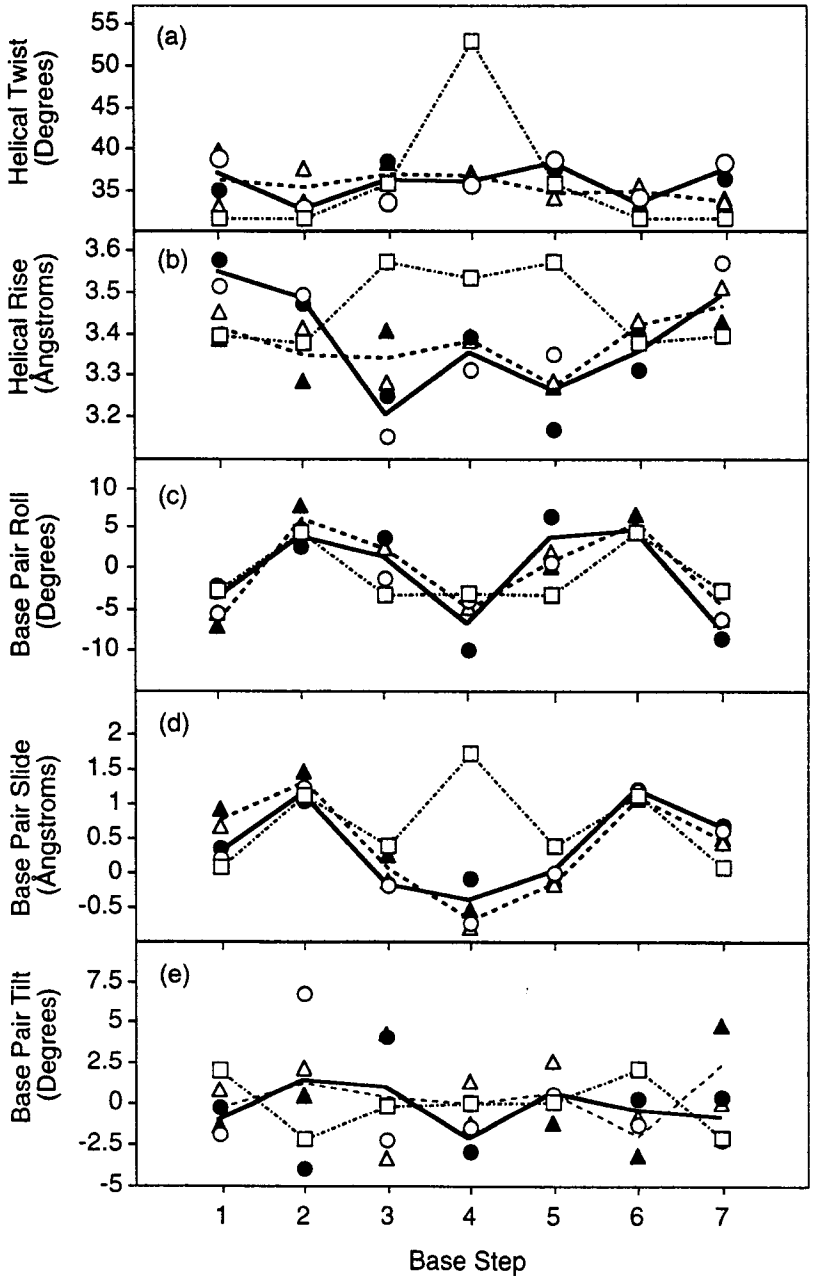

(a)
Helical Twist (Degrees)
55
50
45
40
35
(b)
Helical Rise (Angstroms)
3.6
3.5
3.4
3.3
3.2
(c)
Base Pair Roll (Degrees)
10
5
0
-5
-10
(d)
Base Pair Slide (Angstroms)
2
1.5
1
0.5
0
-0.5
(e)
Base Pair Tilt (Degrees)
7.5
5
2.5
0
-2.5
-5
1
2
3
4
5
6
7
Base Step

Figure 3.7

distance from the molecular dyad of the crystal structure of the B-DNA decamer d(CCAGGCCTGG) and the nonamers, (the minor groove width is the shortest phosphorous to phosphorous distance across the minor groove minus 5.8 Å to account for the van der Waals radii of the phosphate groups). At this position, the decamer structure has a width of 8.3 Å, the structure of d(GCGTACGCG) has a width of 7.7 Å, and the four structures of d(GCAATTGCG) have widths that range from 5.9 Å to 7.6 Å. The narrower widths in the nonamers show that the overhanging guanines are not widening the minor grooves. Triplet formation would also be expected to cause opening of the terminal base pairs towards the minor groove (negative opening), where the opening angle is defined as the rotation angle between two bases in a pair about their common vertical axis. The opening angles range from -5.4° to 6.3°which are not significantly different from the opening angles of base pairs in other positions in the duplexes (Figure 3.7a).

The global helical axes of the octamer duplexes sections (drawn with the program CURVES) have several minor bends. The minor bends at the second and sixth base steps in the all of the nonamer structures are associated with the same positive roll angle which bends the helical axis towards the major groove (Figure 3.6c). However, the bends at every other base step of this type in the neighboring duplexes in the stack face opposing directions and thus cancel each other. This cancellation allows the duplexes to stack in straight columns. There is a minor weave in the columns of d(GCAATTGCG) duplexes due to every other helical junction being pulled ~2Å to one side by contacts with one neighboring column of duplexes. These bends, which are apparent in Figure 3.5d, are due to crystal packing geometry of the monoclinic lattice rather than base sequence since the sequence at the crossovers is identical in the tetragonal and monoclinic 3'-overhang

nonamers. The larger bends on the monoclinic structures have large positive or negative tilt components at the second and third base steps (Figure 3.7e)

3.3.7 Base triplet formation affected the terminal but not the central base steps

The ocatmer duplex sections have sequences of the type d(GCxxxxGC), where x is any nucleotide. The terminal GC base steps are impacted by the intermolecular contacts in the minor groove with overhanging guanines while the xxxx base steps show strong correlations with base sequence. The Cx base steps at the junction between these two regions experience the roll bend that has already been discussed.

The GC regions are affected by both (a) hydrogen bonding in the minor groove of the guanine in the terminal base pair to the overhanging guanine and (b) hydrogen bonding between a hydrated magnesium complex at the helical crossover and the N7 and O6 atoms of the 5'-terminal guanine. The intermolecular contacts increase the slide and reduce the roll values from what has been observed in GC steps in B-DNA crystal structures in general to what has been observed for GC steps near the ends of dodecamers of the sequence type 5'-CGCxxxxxxGCG-3' (where x is any nucleotide) (Hunter and Lu, 1997) and the octamer d(CGCTAGCG) (Tereshko *et al.*, 1996).

The AATT tetramer shares with the Drew-Dickerson dodecamer d(CGCGAATTCGCG) the base sequence dependent features of high propeller twist of the A·T base pairs and the associated narrowing of the minor groove in the center of the duplex. The nonamer d(GCAATTGCG) had more severe and extensive narrowing of the minor groove (3.6 Å minimum width) compared to the nonamer d(GCGTACGCG) (4.4 Å minimum width) and

compared in the GC rich decamer d(CCAGGCCTGG) (5.2 Å minimum width) (Heinemann & Alings, 1989). The central AATT tetramer of d(GCAATTGCG) has a rms deviation of 0.8 Å with the central AATT tetramer of the Drew-Dickerson dodecamer. In contract, the central GTAC tetramer of d(GCGTACGCG) has a rms deviation of 1.8 Å with the central tetramer of the Drew-Dickerson dodecamer.

The central GTAC tetramer of d(GCGTACGCG) is unique since a central GTAC tetramer sequence has not been previously crystallized as B-DNA. In models of B-DNA generated with canonical B-DNA helical parameters, TA base steps have very little overlap of the bases within a strand or between strands. The large twist and postive slide values in the crystal structure further of d(GCGTACGCG) further reduce this overlap. The central TA base step has a very large twist and large positive slide values which abolishes overlap between the purine and the pyrimidine ring. Instead, each six-membered ring lies on top of an exocyclic base atom of the neighboring nucleotide. This overlap may be energetically favorable due to the interaction of a permanent dipole with the delocalized electrons of the base. In fact, adenines and thymine bases stack in this fashion (Bugg *et al.*, 1970) in the crystalline state which suggests that the TA step is exhibiting geometry for bases that are not constrained by being in a helix. The large twist of this step violates Calladine's prediction that pyrimidine/purine steps have low twist to avoid clashes of the purine bases in the minor groove (Calladine, 1982). In this case, the clashes are alleviated by a large rise value. The large twist of this step compares well with of the two duplexes of d(CGCTCTAGAGAGCG) which have helical twists of 45.2° and 43.2° and with the central base step of d(CGCTAGCG) which has a helical twist of 48.9 Å (Tereshko *et al.*, 1996). The two adjacent A·T base pairs have intermediate

propeller twists of ~16° compared to ~8° in isolated A·T base pairs of the decamer d(CCAGGCCTGG) (Hienemann & Alings, 1989) and ~20°+ in the four A·T base pairs of the nonamer d(GCAATTGCG). This is reflected in intermediate values of minor groove width of the nonamer d(GCGTACGCG). The narrow minor grooves and high propeller twist of A/T rich regions are sequence dependent features that have been observed in most B-DNA crystal structures (Berman, 1997).

3.3.8 Triplets disrupt hydration patterns

The presence of the triplet in the minor groove interrupted the line of waters known as the spine of hydration. The spines of hydration were also interrupted in both crystal structures of d(GCAATTGCG) by a hydrated magnesium complex which sits in the minor groove of one duplex and is hydrogen bonded to adenine 21, adenine 22, and thymine 33 (Figure 3.5c). Further details of the hydration pattern are not presented here because not all of the first shell waters could not be modeled accurately with the medium resolution diffraction data.

3.4 Discussion

We have determined the crystal structures of two nonanucleotide sequences, both characterized by eight base pairs of standard B-DNA, with a single 3′ terminal guanine overhanging each end of the duplex. Although the two sequences crystallize in different space groups, the stacking of the octamer duplex sections is very similar. The duplexes are stacked end-to-end such that the overhanging base sits in the minor groove of an adjacent duplex to form a d[G*(G·C)] base triplet. Thus, the stacking of the B-DNA

duplexes and the base triplet observed here are not anomalies, but are characteristic of the sequence type d(GCxxxxGCG), where x is any base.

When we compare these structures to the structure of the nonamer d(GCGAATTCG) (Van Meervelt *et al.*, 1995), we observe that the type of base triplet formed by the overhanging guanine and the adjacent DNA duplex depends on which end of the DNA strand is extended (Figure 3.4). Why should triplet type be correlated with overhang type? When we compare the stacking of the B-DNA duplexes with 5'-overhangs to those with 3'-overhangs, we find that the angles between the terminal base pairs are essentially the same (Figure 4a & c). If we start with this duplex-duplex interaction as the primary crystal lattice interaction, it becomes clear why extending the 5'-end of the duplex places the overhanging base into the major groove (Figure 4a), while extension of the 3'-end places the extra base in the minor groove of the adjacent duplex (Figure 4c). Once this is established, the geometrical constraints of the grooves dictate the details of the end-on-end stacking between duplexes. In the case of the 5'-overhang, the major groove is wide enough to allow the guanine to sit in the same plane as the Watson-Crick base pair if one duplex slides along the long axis of the terminal base pair of the adjacent duplex. The sliding motion relieves a clash between the 5'-overhanging base and the terminal base pair of the neighboring duplex, enabling the overhanging base to remain stacked in its strand. This sliding motion, however, breaks the alignment of the helical axes of the neighboring duplexes. In the case of 3'-overhanging guanines, the guanine must reside in a minor groove which is too narrow to accommodate a third base that is coplanar with the Watson-Crick base pair. Instead, the 3'-overhanging guanine slips into the minor groove with its long axis tilted with respect to the guanine of the terminal Watson-Crick

base pair to which it hydrogen bonds. By unstacking, the 3'-overhang avoids a clash with the backbone of the neighboring duplex. This permits the stacking of the octamer duplexes with their helical axes aligned.

This does not, however, imply that the stacking between the ends of the eight base pair duplexes is an overriding lattice interaction. Indeed, when we remove the 3'-terminal guanine from d(GCGTACGCG), leaving only the duplex d(GCGTACGC), the octamer sequence crystallizes as A-DNA in the space group $P6_1$ with the terminal d(G·C) base pair sitting on the minor groove surface of an adjacent duplex, as is standard for A-DNA octamers that crystallize in this space group (Shakked *et al.*, 1981). Thus, although the end-to-end stacking of B-DNA duplexes is an available lattice interaction for crystallizing B-DNA octanucleotides, it is not sufficiently strong in the case of d(GCGTACGC) to overcome the stronger interaction of the d(G·C) base pair sitting in the minor groove of the A-DNA duplex.

In fact, the end-on-end stacking of B-DNA octamers may also be a weaker interaction than interduplex d(G·G) base pair formation in the minor groove. The only crystal structure of a B-DNA octamer duplex, d(CGCTAGCG), crystallizes with overlapping ends stabilized by two interduplex d(G·G) base pairs in the minor groove of a neighboring duplex (Tereshko *et al.*,1996). These d(G·G) base pairs are similar to the d(G·G) base pairs in the base triplets of the nonamers reported here.

Hydrogen bond counting can be used to explain why the unstacking of the 3'-terminal guanines to form a triplet is more favorable in B-DNA nonamers than the formation of sometime type of unusual d(G·G) base pair between the stacked 3'-overhangs of two adjacent duplexes. The formation of stacked duplexes joined by d(G·G) base pairs sharing two hydrogen bonds would add four interduplex hydrogen bonds per duplex whereas the

formation of two d[G*(G·C)] base triplets at the junction between duplexes adds eight interduplex hydrogen bonds per duplex (or twelve hydrogen bonds if the N3 to O1P hydrogen bonds are counted).

The structures of the nonamer sequences d(GCGTACGCG) and d(GCAATTGCG) suggest that the crystallization of stacked B-DNA duplexes stabilized by d[G*(G·C)] base triplets will require sequences of the type d(GCxxxxGCG). Comparisons of the end-on-end stacking of these nonamers to those duplexes with 5'-overhangs demonstrate that the stacking angle between duplexes is an important consideration in the design of oligonucleotides with overhangs to stabilize the crystallization of the DNA*protein complexes and unusual nucleic acid structures.

3.5 Materials and Methods

3.5.1 Synthesis, purification & crystallization

The nine base oligonucleotides d(GCGTACGCG) and d(GCAATTGCG) were synthesized using phosphoramidite chemistry on an Applied Biosystems DNA synthesizer. Size exclusion chromatography on a Sephadex G-25 column was used to remove salts, blocking groups, and prematurely terminated oligonucleotides. The oligonucleotides were lyophilized, resuspended in 30 mM sodium cacodylate buffer (pH 6.0), and used for crystallization without further purification.

Crystals of the two sequences were grown at room temperature by vapor diffusion in sitting drop setups. Needle-like crystals with dimensions of 0.1 mm x 0.15 mm x 2.0 mm of the sequence d(GCGTACGCG) appeared in two weeks from a solution initially containing 0.5 mM DNA (single-strands), 30 mM sodium cacodylate (pH 6.0), 10 mM $MgCl_2$, 1.0 mM spermine, and

10% (v/v) 2-methyl-2,4-pentanediol (MPD), with equilibration against a reservoir of 50% MPD. Meanwhile, 0.1 mm x 0.1 mm x 0.1 mm trapezoidal crystals of the sequence d(GCAATTGCG) appeared in a month from solutions containing 0.5 mM DNA (single-strands), 30 mM sodium cacodylate (pH 6.0), 150 mM $MgCl_2$, 0.5 mM spermine, and 7.5% MPD, with equilibration against a reservoir of 25% MPD. Several crystals reached dimensions of 0.3 mm x 0.3 mm x.0.2 mm after macroseeding in solutions containing 0.25 mM DNA (single-strands), 30 mM sodium cacodylate (pH 6.0), 150 mM $MgCl_2$, 0.25 mM spermine, and 20% MPD, with equilibration against a reservior of 22.5% MPD.

3.5.2 X-ray diffraction data collection

X-ray diffraction data for the crystals were collected at room temperature using a Siemens P4 diffractometer (Cu-K_α radiation from a sealed tube source). Diffraction data for the crystal of d(GCGTACGCG) were collected with a Siemens Point Counter while diffraction data for the crystal of d(GCAATTGCG) were collected with a Siemens HI-STAR area detector. The diffraction pattern from the crystal of d(GCGTACGCG) had $4/mmm$ symmetry while those of d(GCAATTGCG) had $2/m$ symmetry. Crystallographic data are presented in Table 3.2.

3.5.6 Structure solution and refinement

The structure of the sequence d(GCGTACGCG) was solved by using the information from the Patterson map and the symmetry of the crystal lattice. The Patterson map generated from the observed structure factors had planar densities spaced 3.4 Å along the *a-b* diagonal, suggesting that the duplex was in the B-form with its helix axis was aligned parallel to a *a-b*

diagonal. The unit cell volume suggested that only one DNA strand was in the asymmetric unit, so the second strand had to be generated by placing the molecular dyad of the duplex on the crystallographic two-fold rotation axis along the *a-b* diagonal. The resulting duplex was positioned with its helical axis superimposed on a crystallographic two-fold screw axis in the *a-b* face of the unit cell. The two possible orientations of the duplex around the helical axis were tested by refinement in X-PLOR (Brünger, 1992) using a new sets of parameters for nucleic acids (Parkinson *et al.*, 1996). The model with the correct orientation had a R-factor of 35.7% after initial positional refinement to 2.7 Å data. The overhanging nucleotide was found by subsequent model building with XtalView (MacRee, 1992) and further refinement with X-PLOR.A final R-factor of 20.4% (R_{free}=28.9%) was obtained after including diffraction data to 2.21 Å and adding 17 waters and a 0.5 Mg^{2+} ion which sits on a 4_2-axis.

The structure of d(GCAATTGCG) was thought to be similar to that of d(GCGTACGCG), because Patterson maps generated from the d(GCAATTGCG) diffraction data also had planar densities separated by 3.4 Å. The unit cell volumes suggested that two duplexes were in each asymmetric unit, so two stacked duplexes from the crystal structure of d(GCGTACGCG) were used as the search model in the molecular replacement program AMoRe (Navaza, 1994). The correct solution for the first monoclinic structure, d(GCAATTGCG)-A, had the two duplexes aligned along one of the short diagonals that passes through the center of the unit cell, in agreement with the Patterson map. This solution had a R-factor of 23.7% after rigid body refinement, simulated annealing (starting at 3000°K), conventional position refinement, and individual B-factor refinement with data from 8-

Table 3.2. Diffraction data and refinement statistics for the crystal structures of d(GCGTACGCG) and d(GCAATTGCG).

Parameter	GCGTACGCG	GCAATTGCG-A	GCAATTGCG-B
	Unit Cell Data		
a (Å)	38.92	38.23	39.16
b (Å)	38.92	27.86	29.07
c (Å)	33.03	46.07	46.02
β(°)		114.19	114.11
Space Group	$P4_22_12$	$P2_1$	$P2_1$
Asymmetric unit	1 strand,	4 strands,	4 strands,
	17 waters,	72 waters,	20 waters,
	0.5 Mg2+	4 Mg2+	3 Mg2+
	Diffraction Data		
Total Reflections	1,920	14,526	9,837
Unique Reflections	1,126	5,936	5,791
Resolution Range (Å)	28.2-2.2	19.1-1.8	13.7-2.0
$R_{sym}(I)^*$ for I > 0 (%)	4.2	6.5	8.0
	Refinement Statistics		
R-factor (%)	20.41	19.63	18.45
R-free (%)	28.95	27.17	27.06
Effective Resolution	2.5 Å	2.25 Å	2.3 Å
Resolution Range (Å)	8-2.21	8-2.25	8-2.30
Completeness (%)	64.4	79.4	77.9
Reflections	920, F> 1σ(F)	3372, F> 2σ(F)	3540, F> 2σ(F)

Table 3.2 (Continued)

rms bond lengths (Å)	0.015	0.010	0.012
rms bond angles (deg.)	1.660	1.674	1.497

*$R_{sym}(I)$= 100 x (Σhkl (| I - <I> | / | <I> |)) / n where I is the integrated intensity of a reflection, <I> is the average of all observations of the reflection and its symmetry equivalents, and n is the number of unique reflections. All positive, non-zero reflections (2σ intensity cutoff) were included in scaling and merging.

2.35 Å in X-PLOR. The final model has a R-factor of 19.6% (R_{free} = 27.2%), 72 waters, and 4 hydrated magnesium complexes with data from 8-2.22 Å (Table 3.2). The correct solution of the second monoclinic structure d(GCAATTGCG)-B had the two duplexes aligned along the *b-c* face diagonal, which was also consistent with its Patterson map. The final model had a R-factor of 18.5% (R_{free} = 27.1%), 38 waters, and 3 hydrated magnesium complexes with data from 8-2.3 Å (Table 3.2).

The final coordinates and structure factors for the crystal structures of d(GCGTACGCG), d(GCAATTGCG)-A, and d(GCAATTGCG)-B have been deposited in the Nucleic Acid Database (Berman *et al.*, 1992). Their reference codes are XXXX, XXXX, and XXXX, respectively.

3.6. Acknowledgements

This work has been supported by a grant from the National Science Foundation (MCB-9304467). B. H. M. M. was supported by an AASERT grant from the Office of Naval Research (N00014-9311355). We thank Dr. Jiri Sponer of the Academy of Sciences of the Cech Republic for providing atomic coordinates of the theoretical guanine--guanine base pair.

3.7. References

Babcock, M. S., Pednault, E. P. D. & Olson, W. K. (1994) Nucleic acid structure analysis: mathematics for local cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* **237**, 125-156.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Geblin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992) The

Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751-759.

Broomhead, J. M. (1951) The structures of pyrimidines and purines. IV. The crystal structure of guanine hydrochloride and its relation to that of adenine hydrochloride. *Acta Crystallogr.*, **4**, 92-100.

Brown, T. & Hunter, W. N. (1997) Non-Watson-Crick base associations in DNA and RNA revealed by single crystal x-ray diffraction methods: mismatches, modified bases, and nonduplex DNA. *Biopolymers (Nucleic Acid Sciences)*, **44**, 91-103.

Brünger, A. (1992) *X-PLOR Manual, Version 3.1*. Yale University, New Haven, USA.

Bugg, C. E., Thomas, J. M., Sundaralingam, M., & Roa, S. T. (1971) Stereochemistry of nucleic acids and thier constituents. X. Solid-state base-stacking patterns in nucleic acid constituents and polynicleotides. *Biopolymers*, **10**, 175-219.

Calladine, C. R. (1982) Mechanism of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.* **161**, 343-352.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.

Cate, J. H. & Doudna, J. A. (1996) Metal-binding sites int he major groove of a large ribozyme domain. *Structure*, **4**, 1221-1229.

Coll, M., Solans, X., Font-Altaba, M., & Subriana, J. A. (1987) Crystal and molecular structure of the sodium salt of the dinucleotide duplex d(CpG). *J. Biomol. Struct. Dyn.*, 4,797-811.

Cruse, W. B. T., Egert, E., Kennard, O., Sala, G. B., Salisbury, S. A., & Viswamitra, M. A. (1983) Structure of a mispaired RNA double helix at 1.6 Å resolution and implications for the prediction of RNA secondary structure. *Biochemistry*, **22**, 1833-1839.

Dickerson, R. E., Goodsell, D. S., Kopka, M. L., & Pjura, P. E. (1987) The effect of crystal packing on oligonucleotide double helix structure. *J. Biomol. Struct. Dyn.* **5**, 557-579.

DiGrabriele, A. D., & Steitz, T. A. (1993) A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024-1039.

Egli, M., Gessner, R. V., Williams, L. D., Quigley, G. J., van der Marel, G., van Boom, J. H., Rich, A., & Frederick, C. A. (1990) Atomic-resolution structure of the cellulose synthetase regulator cyclic diguanylic acid. *Proc. Natl. Acad. Sci. USA* **87**, 3235-3239.

Goodsell, D. S., Kaczor-Grzeskowiak, M., & Dickerson, R. E. (1994) The crystal structure of C-G-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps. *J. Mol. Biol.*, **239**, 79-96.

Heinemann, U., & Alings, C. (1989) Crystallographic study of one turn of G/C rich B-DNA. *J. Mol. Biol.* **210**, 369-381.

Holbrook, S. R., Sussman, J. L., Warrant, R. W., Church, G. M., & Kim, S.-H. (1977) RNA-ligand interactions: (I) magnesium binding sites in yeast $tRNA^{Phe}$". *Nuc. Acid. Res.* **4**, 2811-2820.

Hou, Y.-M., Westhof, E., & Giege, R. (1993) An unusual RNA tertiary interaction has a role for the specific aminoacylation of a transfer RNA. *Proc. Natl. Acad. Sci. USA*, **90**, 6776-6780.

Hunter, C. A. & Lu, X.-J. (1997) DNA base-stacking interactions: A comparison of theoretical calculations with oligonucleotide X-ray crystal structure. *J. Mol. Biol.* **265**, 603-613.

MacRee, D. (1992). A visual protein crystallographic software system for X11/Xview". *J. Mol. Graphics* **10**, 44-46.

Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. sect. A*, **50**, 157-163.

O'Brien, E. J. (1967) Crystal structures of two complexes containing guanine and cytosine derivatives. *Acta Crystallogr.* **23**, 92-106.

Parkinson, G. N., Vojtechovsky, J., Clowney, L., & Berman, H. M. (1996) New parameters for the refinement of nucleic acid-containing structures. *Acta Cryst., Sect. D.*, **52**, 57-64. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) *Nature*, **372**, 111-113.

Ramakrishnan, B. and Sundaralingam, M. (1993) High resolution crystal structure of the A-DNA decamer d(CCCGGCCGGG). Novel intermolecular base-paired G*(G·C) triplets. *J. Mol. Biol.*, **231**, 431-444.

Rich, A. and RajBhandary, U. L. (1976) Transfer RNA: molecular structure, sequence, and properties. *Annu. Rev. Biochem.*, **45**, 805-860.

Saenger, W. (1983). *Principles of nucleic acid structure.* Springer-Verlag New York Inc., New York.

Shakked, Z., Rabinovich, D., Cruse, W.B.T., Egert, E., Kennard, O., Sala, G., Salisbury, S.A., & Viswamitra, M.A. (1981) Crystalline A-DNA: the x-ray analysis of the fragment d(G-G-T-A-T-A-C-C). *Proc. R. Soc. Lond. Sect. B.*, **213**, 479-487.

Spink, N., Nunn, C. M., Vojtechovsky, J., Berman, H. M., & Neidle, S. (1995) Crystal structure of a DNA decamer showing a novel pseudo four-way helix-helix junction. *Proc. Natl. Acad. Sci.*, USA **92**, 10767-10771.

Sponer, J., Florian, J., Hobza, P., & Leszcynski, J. (1996) Nonplanar DNA base pairs. *J. Biomol. Struc. Dyn.* **13**, 827-833.

Tereshko, V., Urpi', L. Malinina, L., Huynh-Dinh, T., & Subirana, J. A. (1996) Structure of the B-DNA oligomers d(CGCTAGCG) and d(CGCTCTAGCGCG) in new crystal forms. *Biochemistry*, **35**, 11589-11595.

Van Meervelt, L, Vlieghe, D., Dautuant, A., Gallois, B., Precigoux, G. & Kennard, O. (1995) High resolution structure of a DNA helix forming (C·G)*G base triplets. *Nature*, **374**, 742-744.

Vlieghe, D., Van Meervelt, L., Dautant, A., Gallois, B. Pre'cigoux, G., & Kennard, O. (1996a) Formation of (C·G)*G triplets in a B-DNA duplex with overhanging bases. *Acta Cryst. Sect. D*, **52**, 766-775.

Vlieghe, D., Van Meervelt, L., Dautant, A., Gallois, B. Pre'cigoux, G., & Kennard, O. (1996b) Parallel and antiparallel (G·GC)$_2$ triple hleix fragments in a crystal structure. *Science*, **273**, 1702-1705.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature* , **287**, 755-758.

Yanagi, K., Privé, G. G., Goodsell, D. S., & Dickerson, R. E. (1991) Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J. Mol. Biol.*, **217**, 201-214.

Chapter 4

The Structures and Relative Stabilities of d(G·G) Reverse Hoogsteen, d(G·T) Reverse Wobble, and d(G·C) Reverse Watson-Crick Base Pairs in DNA Crystals

Blaine H. M. Mooers, Brandt F. Eichman, and P. Shing Ho*

Submited to *Journal of Molecular Biology*,
Academic Press Limited, London, England
April 1997, 12 pages, in press.

4.1. Summary

We have solved the structures of the homoduplex d(Gm^{5}CGCGCG)$_2$, and the heteroduplexes d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG) /d(CCGCGCG). The structures form six base pairs of identical Z-DNA duplexes with single nucleotides overhanging at the 5'-ends. The overhanging nucleotide from one strand remains stacked and sandwiched between the blunt-ends of two adjacent Z-DNA duplexes, while the overhanging base of the opposing strand is extra-helical. The stacked and the extra-helical bases from adjacent duplexes pair to form distorted d(G·G) reverse Hoogsteen base pairs in the d(Gm^{5}CGCGCG)$_2$ homoduplex, and d(G·T) reverse wobble and d(G·C) reverse Watson-Crick base pairs in the d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG)/ d(CCGCGCG) heteroduplexes, respectively. Interestingly, only the d(G·T) and d(G·C) base pairs were observed in the heteroduplexes, suggesting that both the d(G·T) reverse wobble and d(G·C) reverse Watson-Crick base pairs are more stable in this crystal environment than the d(G·G) reverse Hoogsteen base pair. To estimate the relative stability of the three types of reverse base pairs, crystals were grown using various mixtures of sequences and their strand compositions were analyzed by mass spectrometry. The d(G·C) reverse Watson-Crick base pair was estimated to be more stable by ~1.5 kcal/mol and the d(G·T) reverse wobble base pair more stable by ~0.5 kcal/mol than the d(G·G) reverse Hoogsteen base pair. The step during crystallization responsible for discriminating between the strands in the crystal is highly cooperative, suggesting that it occurs during the initial nucleating event of crystal growth.

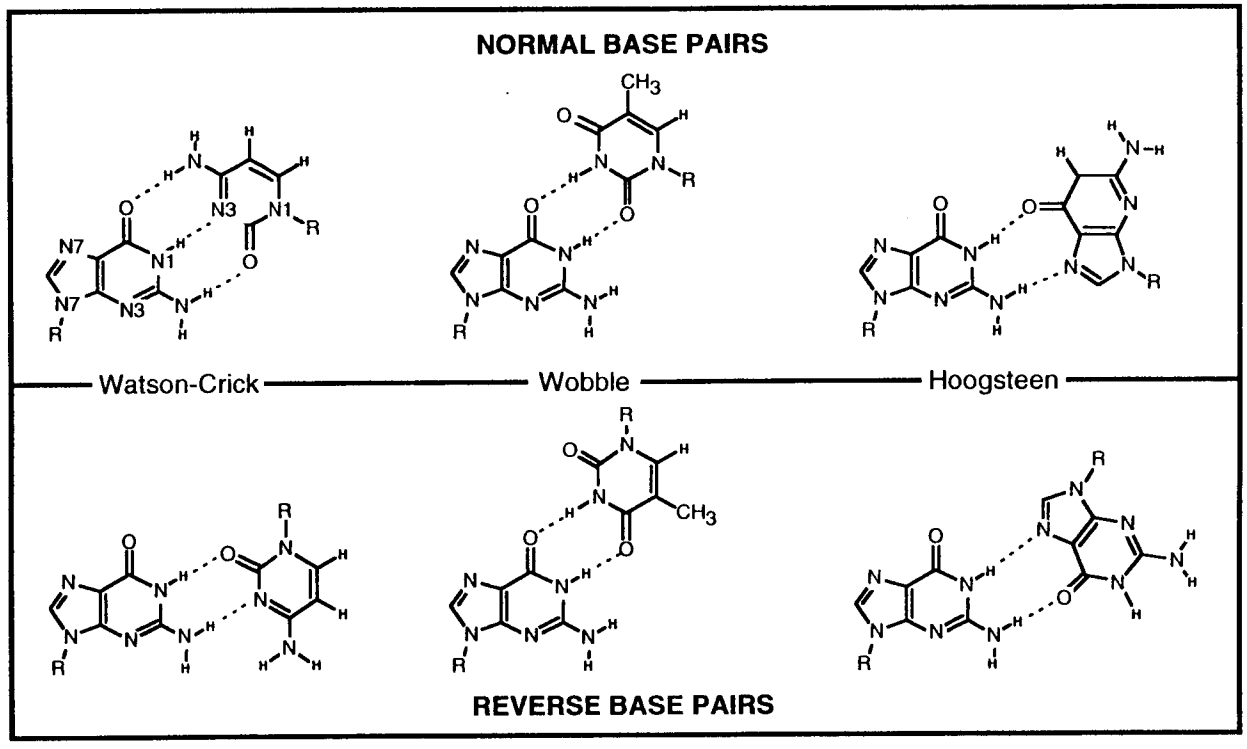
4.2. Introduction

The proper pairing of nucleotide bases ensures fidelity in replication and transcription of the genetic information in a cell. The pairing of guanine with cytosine and adenine with thymine in what is now known as standard Watson-Crick base-pairing forms the basis for the structure of antiparallel DNA and RNA duplexes. Unusual base pairs, however, also play important roles in the transmission of genetic information. Here, we study the structures of reversed base pairs formed by nucleotides that overhang at the 5'-end of the homoduplex d(Gm^{5}CGCGCG)$_2$, and of the d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG)/d(CCGCGCG) heteroduplexes. A distorted d(G·G) reverse Hoogsteen base pair is formed in the homoduplex, while d(G·T) reverse wobble and d(G·C) reverse Watson-Crick base pairs of the type observed in RNA structures form in the respective heteroduplexes.

The two strands of most DNA and RNA structures are oriented antiparallel to each other. In DNA duplexes, Watson-Crick-type base pairs are the predominant interactions that hold the two strands together. When bases are mismatched in DNA, unusual base pairing can occur, including wobble base pairs between G and T and Hoogsteen-type base pairs between two purine nucleotides (Figure 4.1). These are less stable than standard Watson-Crick-type base pairs. In RNA structures, unusual base pairing is more prevalent, and have been observed to stabilize the complex tertiary structures of large polynucleotides such as tRNA (Quigley & Rich, 1976), hammerhead ribozymes (Pley *et al.*, 1994), and the self-splicing Group I intron from *Tetrahymena thermophila* (Cate *et al.*, 1996).

Figure 4.1. Comparison of the normal and reverse Watson-Crick d(G·C) base pairs, wobble d(G·T) base pairs, and Hoogsteen-like d(G·G) base pairs. The nitrogens in the bases of the normal Watson-Crick d(G·C) base pair are numbered to orient the reader to the standard numbering of the purine and pyrimidine bases. The glycosidic bond that links the bases to the deoxyribose (R) in the reverse base pairs are oriented antiparallel to each other so that there is no distinction in terms of major and minor grooves, as there are in the normal base pairs. There is, however, a common guanine in the three structures studied here, and the major and minor grooves of this base is used as the reference in discussing the surfaces in the text. The Watson-Crick face of the bases includes the N1 base nitrogen, N2 amino nitrogen and O6 keto oxygen of the guanine and the O2 keto oxygen, N3 base nitrogen and N4 amino nitrogen of the pyrimidine bases of thymine and cytosine. The Hoogsteen face of guanine is defined by the O6 keto oxygen and the N7 base nitrogen of the purine base.

Figure 4.1.


NORMAL BASE PAIRS
Watson-Crick
Wobble
Hoogsteen
REVERSE BASE PAIRS

Watson-Crick, wobble, and Hoogsteen-type base pairs all have reverse analogues in which one base is completely inverted. Here, we refer to "reverse" base pairs as the pairing of the nucleotide bases in which the glycosidic bonds are oriented essentially antiparallel to each other (Figure 4.1). The three base pairs that we study here are asymmetric in the same manner that the normal pairings are asymmetric. The Hoogsteen and reverse Hoogsteen d(G·G) pairs match the Watson-Crick face of one purine with the Hoogsteen face of the second. The reverse analogues of the d(G·T) wobble and d(G·C) Watson-Crick base pairs match the two respective Watson-Crick faces of the bases. In contrast, truly symmetric reverse base pairs, with two identical bases related by a dyad axis perpendicular to the base pair plane, have been used in the design of synthetic parallel-stranded DNA oligomers (Rippe *et al.*, 1992; Robinson *et al.*, 1994). These are interesting structures, although their biological relevance has yet to be determined.

In large RNA structures, however, loops that fold into local secondary and tertiary structures often require the formation of unusual base pairs, including reverse base pairs even if the strands are in antiparallel orientations. For example, in the crystal structure of yeast $tRNA^{Phe}$, a reverse Watson-Crick base pair at (G15·C48) links the α region of the D arm to the variable V loop, and a reverse Hoogsteen base pair at (G22·m^7G46) links the D arm to the variable V loop (Kim, 1978). In a second example, a reverse Hoogsteen-type base pair forms between G7 and G11 at the base of a GNRA structural motif in an RNA aptamer designed to recognize and bind ATP (Jiang *et al.*, 1996). Finally, in the NMR solution structure of the hairpin formed by r(GGAC(UUCG)GUCC), a r(G·U) reverse wobble base pair stabilizes the base of a two nucleotide loop (Varani *et al.*, 1991). This hairpin structure is thought to occur frequently in ribosomal and messenger RNAs.

Thus, non-Watson-Crick base pairs are important for the proper folding of RNA molecules into the compact tertiary structure of their functional forms.

Non-Watson-Crick base pairs are also important for RNA-protein recognition. Genetic and biochemical studies have shown that protein binding sites in RNA are often associated with important non-Watson-Crick base pairs (Allmang *et al.*, 1994; Ibba *et al.*, 1996). In addition, protein binding of RNA loops can induce the formation of non-Watson-Crick base pairs. For example, the HIV Rev peptide binding to the Rev responsive element (RRE) in the *env* gene of HIV is associated with the formation of two homopurine base pairs in an internal RNA loop (Battiste *et al.*, 1994; Battiste *et al.*, 1996).

The infrequent occurrence of reverse base pairs makes it difficult to study the intrinsic stability associated with specific structures. We present here the atomic resolution structures of three different reverse base pairs formed by nucleotides that overhang the 5'-ends of DNA duplexes. By studying the structures of these base pairs and their effects on duplex formation during crystallization, we have estimated the stability of the d(G·C) reverse Watson-Crick and d(G·T) reverse wobble base pairs relative to the distorted d(G·G) reverse Hoogsteen base pair.

4.3. Results

We have solved the structures of the heptanucleotide duplexes d(GCGCGCG)$_2$ d(Gm^{5}CGCGCG)$_2$) (where m^{5}C is cytosine methylated at the C5 carbon of the base), d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG)/ d(CCGCGCG). In all four structures, the six underlined nucleotides pair to form left-handed Z-DNA d(CGCGCG) duplexes. The nucleotides within these duplexes are numbered 1-7 for each nucleotide of the common

d(GCGCGCG) strand, and 8-14 for the opposing d(NCGCGCG) strand (where N is either G, C, or T). A single nucleotide (G1 of d(GCGCGCG) and N8 of d(NCGCGCG)) is left overhanging each of the 5′-ends of the duplexes. These overhangs pair with overhangs from adjacent duplexes in the crystal lattice to form three different reverse base pairs. The overhanging dG nucleotides of the homoduplexes $d(GCGCGCG)_2$ and $d(Gm^5CGCGCG)_2$ form nearly identical reverse Hoogsteen-type d(G·G) base pairs (rhGG) (Figure 4.2a). However, only the structure of the methylated sequence will be discussed here; it provided a more reliable structure, as was evident from the final R-factors of the refined structures. The duplexes of d(GCGCGCG)/d(TCGCGCG) form reverse wobble d(G·T) base pairs (rwGT) (Figure 4.2c), while the duplexes of d(GCGCGCG)/d(CCGCGCG) form reverse Watson-Crick d(G·C) base pairs (rwcGC) (Figure 4.2b). Thus, in all the structures, the overhanging nucleotide G1 remains stacked against the Z-DNA duplex, while the overhanging nucleotide N8 is extra-helical (Figure 4.3). In the remainder of this section, we will first discuss the duplex structures and crystal lattice interactions that are common to all the sequences, followed by a more detailed description of the structure for each type of base pairing.

4.3.1 Z-DNA duplex structure

The six bases at the 3′-end of each sequence form standard Watson-Crick d(C·G) base pairs. The resulting structure is a left-handed duplex that is nearly identical to the Z-DNA structure of $d(CG)_3$ (Wang *et al.*, 1979). The sugar conformations alternate between C2′-endo for the dC and C3′-endo for the dG nucleotides. The C3′-endo sugar facilitates formation of the *syn* conformation by the purine bases. The 3′-terminal dG nucleotide of each

Figure 4.2. Electron density omit maps of the (a) reverse Hoogsteen-like d(G·G) (rhGG) (b) reverse Watson-Crick d(G·C) (rwcGC), and (c) reverse wobble d(G·T) (rwGT) base pairs. Shown are Fo-Fc maps in which the overhanging bases at the 5′-ends were excluded from the phasing information used to calculate the structure factors. The hydrogen bonds that link the two bases of each base pair and the common guanosine in *syn* to its deoxyribose sugar are indicated by the dashed lines, along with the distances for each hydrogen bond.

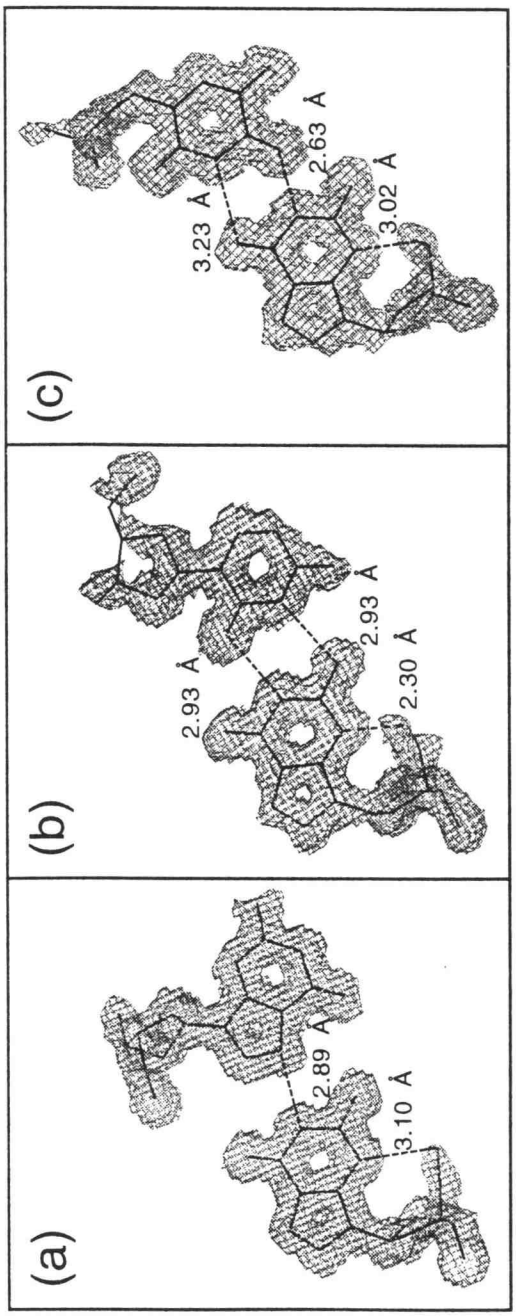


Figure 4.2.

Figure 4.3. Stereoview of the crystal lattice. All three crystals in this study are in the space group $P2_12_12_1$ with four symmetry-related heptamers in the unit cell (the contents of one unit cell of d($Gm^5CGCGCG$)$_2$ is shown). The axes of the unit cell are labeled. The one unique heptamer duplex in the unit cell is shown as a ball and stick model, while the bonds of the symmetry related hexamer duplexes are shown as simple lines. The guanosine in position 8 flips out and hydrogen bonds to the stacked guanosine in position 1 of the adjacent molecule in the same layer of duplexes. Thus, there is a single base pair interaction between adjacent duplexes within this unit cell.

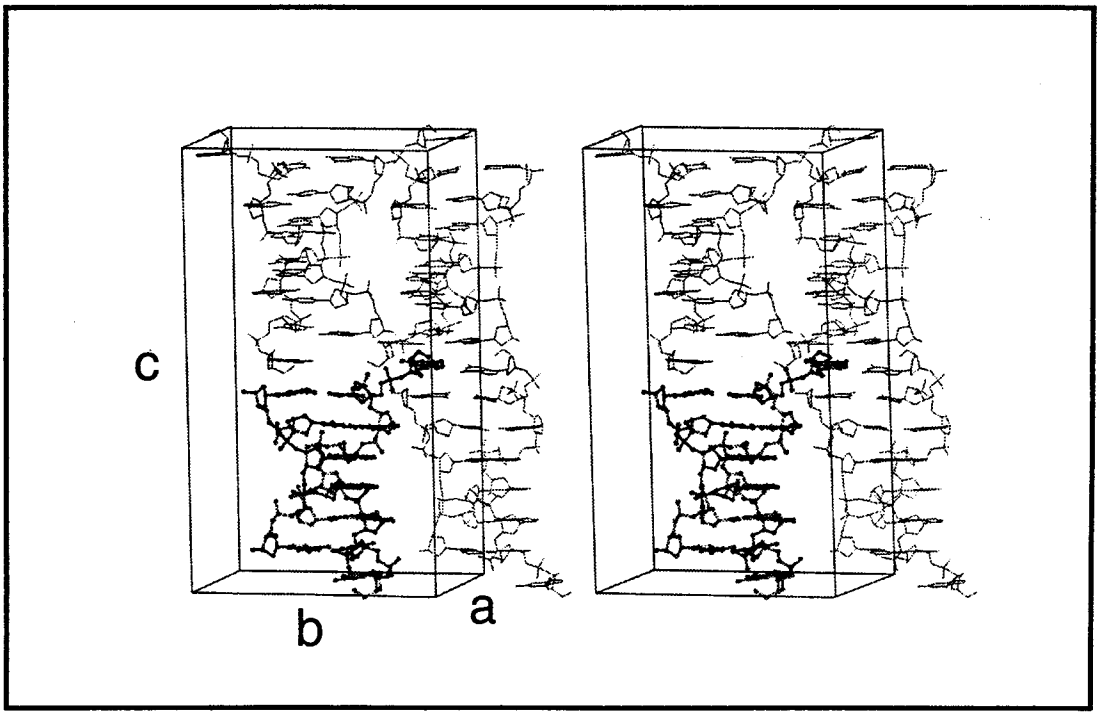


Figure 4.3.

strand, although in the *syn* conformation, adopts a C2'-endo sugar conformation, as was observed in all the standard Z-DNA structures of hexanucleotides (Ho & Mooers, 1997). Thus, the duplexes have the general features of Z-DNA as defined by the prototype structure of $d(CG)_3$.

The helical parameters of the duplex structures are compared to that of d(CGCGCG) in greater detail in Table 4.1. The helical twist alternates on average between -9.1° for the d(CpG) step and -47.7° for the d(GpC) step for all three structures, to give an average of -28.4° per base pair. This is nearly identical to the -30.1° twist per base pair observed in the structure $d(CG)_3$. When comparing the individual structures, the helical twists between the base pairs of the homoduplex are identical (to within one standard deviation) to those of the analogous base steps in standard Z-DNA. The two heteroduplexes, however, are slightly less left-handed at the d(GpC) dinucleotide steps than observed in $d(CG)_3$. This distortion in the heteroduplex structures is associated with perturbations required to pair the base of the extra-helical nucleotides with the stacked guanosine G1 of an adjacent duplex.

The other obvious difference is the large buckle (angle between base planes about the short axis of the base pair) of the base pairs at the ends of the duplex regions in all of the current structures. The Z-DNA structure of $d(CG)_3$ is very rigid, and the base pairs very planar (showing very little propeller twisting or buckling). The high buckle at the ends are again likely associated with distortions required to form the unusual base pairs here. All other helical parameters, including the rise at each base step and the propeller twist at each base pair are identical to those in the structure of $d(CG)_3$.

The solvent structures at the major groove surface and the minor

Table 4.1. Comparison of the helical parameters for base steps and base pairs of the $(CG)_3$ Z-DNA regions.

	$(CG)_3$	rhGG	rwcGC	rwGT
Twist (degree/base step)				
C2G3/C13G14	-9.2	-7.7	-6.5	-7.4
G3C4/G12C13	-48.9	-47.7	-47.5	-47.5
C4G5/C11G12	-9.4	-10.4	-10.5	-10.3
G5C6/G10C11	-50.8	-48.5	-48.0	-47.1
C6G7/C9G10	-12.2	-9.1	-9.8	-10.6
Average for CG steps	-10.3±1.7	-9.1±1.3	-8.9±2.1	-9.4±1.8
Average for GC steps	-49.9±1.3	-48.1±0.6	-47.8±0.4	-47.3±0.3
Rise (Å/base step)				
C2G3/C13G14	3.8	3.4	3.4	3.4
G3C4/G12C13	3.6	3.7	3.7	3.6
C4G5/C11G12	3.9	4.0	3.9	4.1
G5C6/G10C11	3.5	3.6	3.6	3.7
C6G7/C9G10	4.1	3.8	3.9	3.8
Average for CG steps	3.9±0.2	3.7±0.3	3.7±0.3	3.8±0.3
Average for GC steps	3.6±0.1	3.7±0.1	3.7±0.1	3.7±0.1
Propeller Twist (degree/ base pair)	1.1	-1.2	6.3	-0.7
C2·G14				
G3·C13	3.2	1.3	0.1	2.2
C4·G12	-0.9	-0.3	-0.7	-0.9
G5·C11	-1.5	-0.3	-0.9	1.2
C6·G10	0.5	-0.3	2.2	-3.6

Table 4.1. (continued)

G7•C9	2.7	3.8	0.1	5.4
Buckle (degree/base pair)				
C2•G14	1.9	10.2	10.0	10.4
G3•C13	-3.5	-6.1	-3.5	-5.9
C4•G12	3.0	0.6	0.0	0.1
G5•C11	-2.4	-1.7	-2.1	-0.7
C6•G10	2.0	3.3	3.8	8.2
G7•C9	0.0	-4.7	-1.3	-8.3

*Values of twist, rise, propeller twist, and buckle are shown for each of the three structures (rhGG, rwcGC, and rwGT) described in the text and are compared with the 1.0 Å crystal structure of $d(CG)_3$ containing only $MgCl_2$ (Gessner *et al*, 1989). Morphologies of the 5'-overhanging ends are not shown. Each structure was evaluated using the program *NASTE* (Nucleic Acid STructure Evaluation), which utilizes a global helix axis to determine each parameter. The rhGG, rwcGC, and rwGT structures were analyzed with the 5'-overhangs removed and with the remaining $d(CG)_3$ duplex juxtaposed to the reference $d(CG)_3$ structure. Residues 2-7 and 9-14 in the rhGG, rwcGC, and rwGT duplexes respectively correspond to residues 1-6 and 7-12 in $d(CG)_3$.

Table 4.2. Backbone and glycosidic torsion angles and sugar puckers of 5′-overhanging nucleotides.*

	rhGG		rwcGC		rwGT	
	G1	G8	G1	C8	G1	T8
γ (C5′-C4′)	17.90	28.01	26.91	140.15	159.91	-162.68
δ (C4′-C3′)	161.03	175.43	161.81	126.98	106.64	120.36
ε (C3′-O3′)	-156.05	-97.43	-155.20	-79.05	-58.74	-124.54
ζ (O3′-P)	-16.70	169.38	-20.17	75.36	-60.63	-134.33
χ (C1′-N)	74.62	-76.75	79.86	171.71	39.13	75.58
	(S)	(S)	(S)	(A)	(S)	(S)
sugar pucker	C2′-endo	C2′-endo	C3′-exo	C1′-exo	C3′-endo	C1′-endo

*Torsion angles (degrees) shown above were calculated using the program *Xfit* (MacRee, 1992). The conformation of the base relative to the deoxyribose ring is denoted beside χ angles as either S (*syn*) or A (*anti*).

groove crevice of the Z-DNA duplexes are for the most part nearly identical to that of the magnesium form of $d(CG)_3$ (Gessner *et al.*, 1989). For most base pairs, two waters bridge the N4 amino groups at the major groove surface of adjacent cytidines in each d(CpG) dinucleotide step. Differences in the pattern of water interactions at this surface result from disruptions caused by the binding of a $Co(NH_3)_6^{+3}$ complex at the central d(CpG) dinucleotide. The amino ligands from this complex form hydrogen bonds to the O4 oxygen and N7 nitrogen of G5, while two amino groups hydrogen bond to the phosphate oxygen of the phosphodiester linking C4 to G5. This phosphate oxygen is also hydrogen bonded to a water ligand of a hydrated magnesium complex. Waters from this $Mg(H_2O)_6^{+2}$ complex also form hydrogen bonds to the backbones of two adjacent duplexes and to the N7 of G12 on one of these duplexes. Thus, the hydrated magnesium links together three duplexes and appears to be important in stabilizing the crystal lattice.

In the minor groove, the four central d(C·G) base pairs show a spine of interconnected waters. This spine is formed by two waters lying nearly in the plane of each base pair. The disruption of this spine at the terminal base pairs is associated with the large buckling and the unusual stacking of the paired overhanging nucleotides. The DNA structure and the solvent structure around the six standard Watson-Crick d(G·C) base pairs, therefore, are very similar to that of standard Z-DNA, with some variations that are specific for the crystal lattice interactions.

4.3.2 Crystal lattice interactions

The duplexes are aligned end-to-end along the crystallographic *c*-axis, similar to the stacking of Z-DNA hexanucleotides in this same space group (Figure 4.3). In the crystal lattice of standard Z-DNA hexanucleotides, the

duplexes stack end-to-end to form quasi-continuous columns along the *c*-axis. Each adjacent column is staggered by two base pairs along this axis. In the current heptamer structures, however, the adjacent Z-DNA duplex regions are all exactly aligned. This can be envisioned as a series of discrete stacked sheets of Z-DNA.

The most noticeable feature of the crystal lattice, however, is that in these heptanucleotide sequences, there is a single base overhanging each 5′-end. The structures of the overhanging nucleotides are not identical, even in the homoduplex $d(GCGCGCG)_2$. In all cases, one of the overhanging guanosine nucleotides sits stacked against the duplex, while the other nucleotide (whether it is a guanosine, cytidine, or thymidine) is extra-helical, extending out and away from the duplex. The extra-helical base pairs with the stacked guanosine of an adjacent duplex within each layer, effectively interlinking the Z-DNA duplexes. The extra-helical base also serves to fill the gap between two stacked duplexes. Thus, the lattice consists of layers of duplexes in which each duplex is linked to two adjacent duplexes by pairing the bases that overhang the 5′-ends.

The pairing of extra-helical bases has previously been observed in the crystal structure of the Z-DNA hexamer of d(CCGCGG) (Malinina *et al.*, 1994). In this case, the bases at both ends flip out and form Watson-Crick base pairs between adjacent duplexes, leaving only four standard base pairs as Z-DNA in the center of the hexamer structure.

In the lattice of the current heptanucleotides, the reverse base pairs bring adjacent helices so close together that a direct helix to helix hydrogen bond forms between O2P of G3 and the O3' of G7 of the duplexes. This hydrogen bond is analogous to the short (2.63 Å) interhelical hydrogen bond observed between the O1P of G2 of one hexamer duplex and the O5' of G12

in an adjacent hexamer in the crystal structure of the Z-DNA hexamer $d(CG)_3$ (Wang *et al.*, 1979, 1981).

The third base step in each strand of these heptanucleotide structures (G3/C4 and G10/C11) is in the unusual Z_{II} conformation, which has been attributed to crystal packing effects in the crystal structure of $d(CG)_3$ (Wang *et al.*, 1981). When a base step is in the Z_{II} conformation, the intervening phosphate is rotated outward about 1 Å away from the minor groove. In all three heptamer crystal structures, the phosphates of nucleotides C4 and C11 are not directly hydrogen bonded to a neighboring duplex or a metal complex, as was observed in the base steps that adopt the Z_{II} conformation in the crystal structures of Z-DNA hexamers and decamers (Gessner *et al.*, 1985, Brennan *et al.*, 1986). Thus, while the Z_{II} conformation in the heptamers may be caused by crystal packing, the lattice interactions causing this distortion remain unclear.

4.3.3 Structure of the helical stacked guanosine nucleotide

In all cases, the guanosine that remains stacked against the Z-DNA duplex is in the *syn* conformation, extending the alternating *anti-syn* pattern of nucleotides from the duplex to include this overhanging nucleotide. The *syn* conformation of this nucleotide is stabilized by a hydrogen bond between the O5′ oxygen of the terminal hydroxyl group to the N3 nitrogen of the guanine base (Figure 4.2). The Watson-Crick edge of the guanine is subsequently oriented to allow pairing with the intervening base of the extra-helical overhanging nucleotide of an adjacent duplex. Since the interduplex base pairs bring the 5′-nucleotides of adjacent duplexes together within these layers, the strands held together in this manner are necessarily parallel to each other. These base pairs are therefore the reverse type, with

the d(G·G) overhangs forming reverse Hoogsteen-type base pairs, the d(G·C) overhangs forming reverse Watson-Crick base pairs, and d(G·T) overhangs forming reverse wobble base pairs (Figures 4.1 and 4.2).

4.3.4 Structure of the reverse Hoogsteen d(G·G) base pair

The Watson-Crick edge of the stacked guanine faces the Hoogsteen edge of the extra-helical guanosine nucleotide of an adjacent duplex to form a reverse Hoogsteen-type d(G·G) base pair (rhGG, Figure 4.2a). In this case, although the stacked guanosine is in *syn*, the extra-helical guanosine base adopts the *anti* conformation. This is analogous to the two mismatched G(*anti*)·G(*syn*) Hoogsteen base pairs in the structure of $d(CGCGAATTGGCG)_2$ (Skelly *et al.*, 1993). The hydrogen bonds that hold the base pair together are shown in Figure 4.2a. The N1 to N7 distance is in the range of hydrogen bond donor to acceptor distances observed in Watson-Crick base pairs. The N2 to O6 distance, however, is significantly longer than that expected for a standard hydrogen bond. This is a result of the shift of both guanines within the plane of the bases. This shift is very noticeable when the rhGG base pair is superimposed on a rhGG base pair that is part of the r[(G·C)*G] triplet in yeast $tRNA^{Phe}$ (Westhof *et al.*, 1988) (Figure 4.4), and is likely the result of an additional hydrogen bond formed between the N2 amino nitrogen of the extra-helical base and the O2P oxygen of cytidine 9 of a third duplex. The base planes of the two guanines are almost exactly coplanar as a consequence of being sandwiched between two Watson-Crick base pairs of the stacked Z-DNA duplexes (Figure 4.5a). In contrast, the Hoogsteen d(G·G) base pairs in the antiparallel duplex of d(CGCGAATTGGCG) are propeller twisted about their long axes (Skelly *et al.*, 1993). In the current structure, several waters

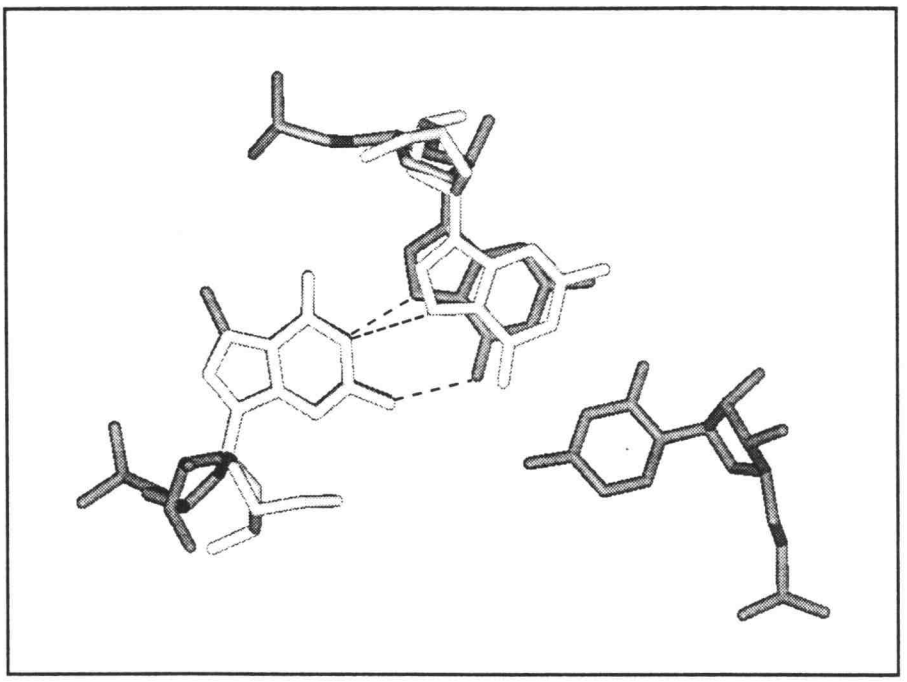


Figure 4.4. Comparison of the reverse Hoogsteen d(G·G) base pair (open bonds) with the r(G22:C13)m^{7}G46 triplet (shaded bonds) in the crystal structure of tRNA^{Phe} (Westhof *et al.*, 1988). The hydrogen bonds between the analogous bases in the two structures are drawn as dashed lines. This superposition reveals a sliding of G8 within the base plane of the Hoogsteen-like base pair in the current structure. This is likely a result of the hydrogen bond formed between the N2 amino group and a phosphate oxygen a third duplex (not shown).

Figure 4.5. Comparison of the base pair stacking and solvent interactions in the crystal structures of (a) d(Gm^{5}CGCGCG)$_2$, (b) d(GCGCGCG)/d(CCGCGCG), and (c) d(GCGCGCG)/d(TCGCGCG). The views are down the helical axes of the three base pairs in the junction between two stacked duplexes. The overall set of interactions involve three different duplexes, two in which the blunt ends of the fully duplexed Z-DNA hexamers sandwich the stacked overhanging guanine (base shown as a ball and stick model), while the third is adjacent to the lower duplex, but contributes the extrahelical base to each reverse base pair (both the base and deoxyribose of this extended nucleotide are shown as ball and stick models). The nucleotides with open bonds are closet to the viewer and include an extrahelical base that extends to form another reverse base pair with an adjacent helix which is not shown. The nucleotides with shaded bonds are farthest from the viewer and include the stacked guanosine in the 5′ terminal position. The dashed lines connect hydrogen bond donors and acceptors important in conferring specificity in the formation of each reverse base pair.

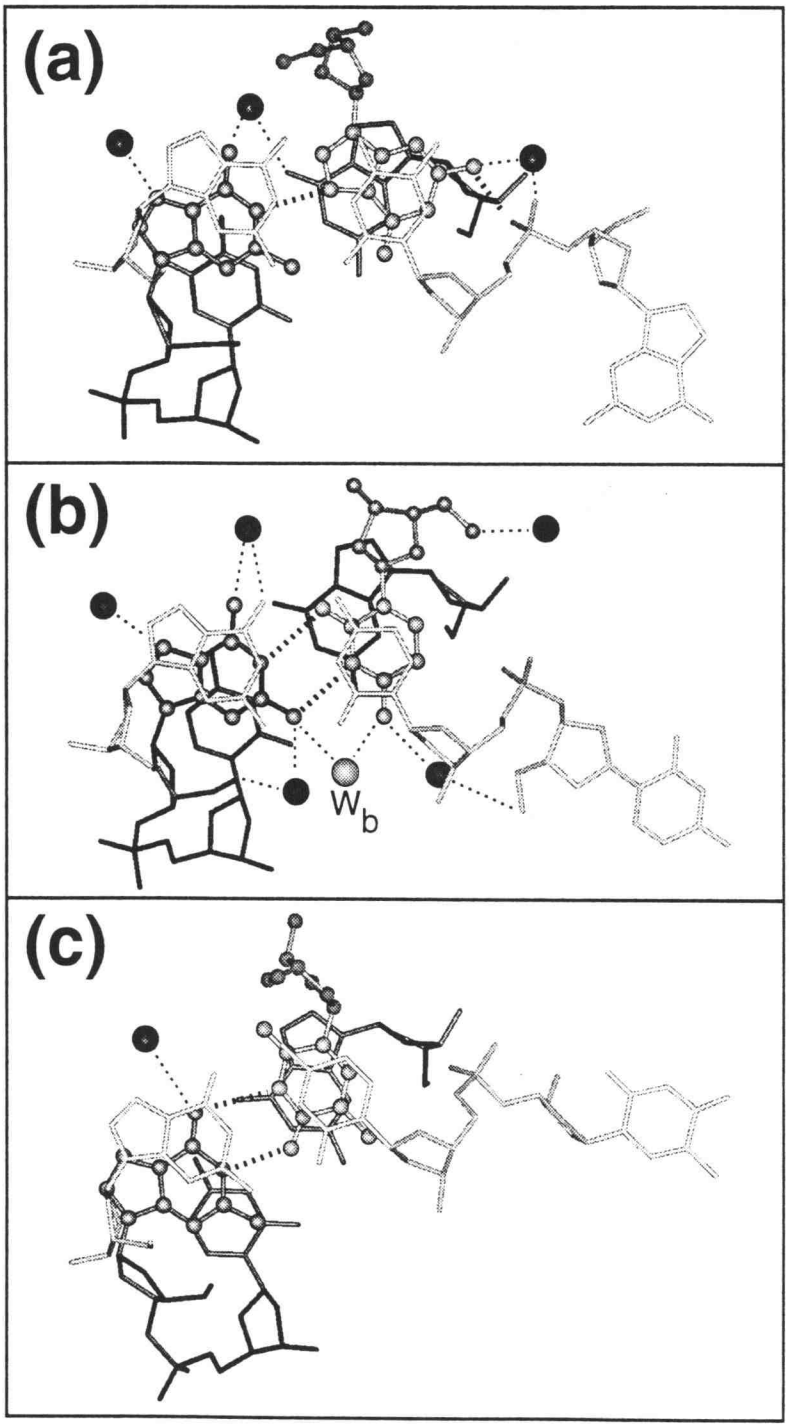


Figure 4.5.

link each overhanging guanine base to neighboring DNA atoms. There are, however, no waters that directly bridge the two guanines in this rhGG base pair. The structure of this base pair is distorted by the crystal lattice and therefore may not represent the type of rhGG base pair expected to form in RNA structures. We will still refer to this as the rhGG base pair, however, because it does conform to the geometry of this type of base pair (Figure 4.1) and, as we will see later, will be a useful reference structure for comparing the stabilities of the rwcGC and rwGT base pairs.

4.3.5 Structure of the reverse Watson-Crick d(G·C) base pair

A reverse Watson-Crick base pair is formed by pairing the stacked guanosine with the extra-helical cytidine of a d(CCGCGCG) strand, with the Watson-Crick edges of their bases facing each other. In this case, the cytidine in *anti* pairs with the guanosine in *syn*. In contrast to the normal d(G·C) base pairs, the resulting rwcGC base pair is held together by only two hydrogen bonds (from the N1 and N2 of the guanine to O2 and N3 atoms, respectively, of the cytosine base). In addition, a single water (W_b) was observed to bridge the guanine N2 to the cytosine N4 nitrogen, which may provide additional stability to this base pairing (Figure 4.5b). When taken together, the reverse Watson-Crick base pair appeared to have the greatest number of base-base and base-water-base hydrogen bonds of the three reverse base pairs presented here.

4.3.6 Structure of the reverse wobble d(G·T) base pair

A reverse wobble base pair is formed by pairing the stacked guanosine with the extra-helical thymidine of the d(TCGCGCG) strand, with the Watson-Crick edges of their bases facing each other. In this case, both

nucleotides are in the *syn*-conformation. The resulting rwGT base pair is held together by two hydrogen bonds (from the N1 and O6 atoms of the guanine to the O4 and N3 atoms, respectively, of the thymine base) (Figure 4.2c). In several crystal structures of DNA duplexes containing wobble G·T mismatches, a water or a hydrated magnesium cation bridges the guanine O6 and the thymine O4 (Ho *et al.*, 1985; Hunter *et al.*, 1986). No analogous solvent interaction was observed in the current rwGT base pair. In fact, very few water molecules were observed around this base pair (Figure 4.5c). This may be the result of slight positional disorder within the base pair plane. The average temperature factors for this base pair are ~2-fold higher than the remainder of the DNA, and 20% higher than that of the rhGG and rwcGC base pairs.

The deoxyribose 05′ oxygen of the thymidine forms a hydrogen bond to the O2 oxygen in the base of cytidine C9 from an adjacent stacked duplex. This intermolecular hydrogen bond apparently stabilizes the *syn* conformation of the thymidine nucleotide. In comparison, the deoxyribose O5′ of the *anti* cytidine in the rwcGC base pair does not show this same hydrogen bonding interaction. As in the case of the rhGG base pair, the bases in the rwGT base pair are nearly coplanar. This is similar to the normal wobble d(G·T) base pair observed in the crystal structure of a Z-DNA hexamer (Ho *et al.*, 1985). On the other hand, the normal wobble d(G·T) and d(G·U) base pairs in several A-DNA crystal structures (Kneale *et al.*, 1985; Hunter *et al.*, 1986; Vojtechovsky *et al.*, 1995) and A-RNA crystal structures (Holbrook *et al.*, 1991, Cruse *et al.*, 1994) show significant propeller twists.

4.3.7 Relative stability of reverse base pairs

We observed only the rwGT base pairs in the crystals of the heterogeneous duplexes formed by mixing the sequence d(GCGCGCG) with d(TCGCGCG). To confirm this observation from the crystal structure, the single crystal was redissolved and the DNA strand composition analyzed by MALDI mass spectrometry (Figure 4.6). The mass spectrum showed that, within experimental error, the crystal was composed of near equal ratios of the two strands. Mass spectra recorded from the four remaining crystals in the crystallization set-up were identical to that of the mounted crystal, indicating that this was not unique to the crystal that we had originally studied. We would expect that mixing these sequences would result in a 1:2:1 ratio of $d(GCGCGCG)_2$, d(GCGCGCG)/d(TCGCGCG), and $d(TCGCGCG)_2$ duplexes in solution. This mixture of homo- and heteroduplexes should also be observed when d(GCGCGCG) is paired with d(CCGCGCG). Again, only the rwcGC pairing was observed. Thus, the crystal lattice discriminates between the reverse base pairs that are formed by the overhanging nucleotides, favoring both rwGT and rwcGC over rhGG base pairs.

In order to gain additional insight into the mechanism for this discrimination and the free energy differences between the rhGG versus either rwGT or rwcGC, we studied the crystallization of these duplexes in solutions containing increasing ratios of the d(GCGCGCG) strand (G_S) relative to either the d(TCGCGCG) or d(CCGCGCG) strands, redissolved the DNA in the crystals, and quantitated the strand composition within the crystals by mass spectrometry. The mass spectra showed equal quantities of d(GCGCGCG) and d(TCGCGCG) when the two strands were added at a 1:1 ratio, but showed only the d(GCGCGCG) strand at ratios $\geq$ 2:1. For the d(CCGCGCG) containing crystals, equal proportions of both strands persisted

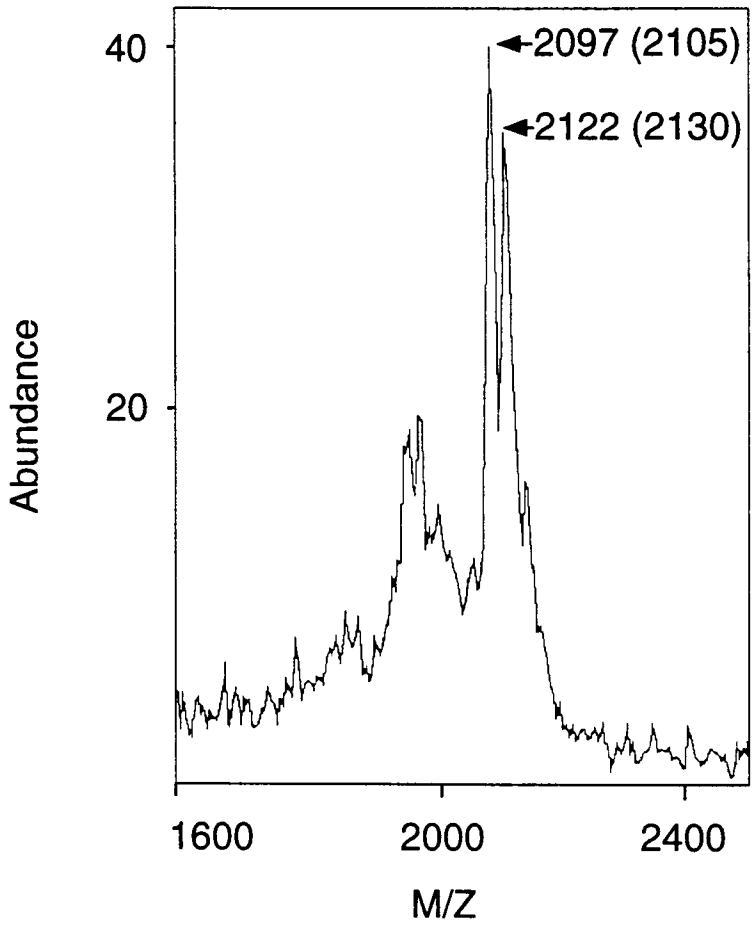


Figure 4.6. MALDI mass spectrumetry analysis of the DNA strand composition of d(GCGCGCG)/d(TCGCGCG) single crystals. The horizontal axis indicates the mass to charge ratio (M/Z) of the observed fragments, while the vertical axis is the abundance of each fragment. The calculated mass of each strand is shown in parentheses next to the measured mass. The mass determined for the d(GCGCGCG) single-strand is 2122 gm/mol (calculated to be 2130 gm/mol), while that of the d(TCGCGCG) single-strand is 2097 gm/mol (calculated to be 2105 gm/mol).

to a ratio of 3:1. At a 4:1 ratio of d(GCGCGCG) added to d(CCGCGCG), however, the spectrum showed predominantly (>90%) the d(GCGCGCG) strand. Thus, the crystals convert from the heteroduplexes to the homoduplex of $d(GCGCGCG)_2$ as the ratio of d(GCGCGCG) added was increased. The order of stability for the reverse base pairs can thus be defined as rwcGC > rwGT > rhGG. This was confirmed in an experiment where crystals were grown with all three sequences added in equal proportions. In this case, the mass spectrum of the dissolved crystals showed only the d(GCGCGCG) and d(CCGCGCG) strands, indicating that this was the preferred pairing of the DNA.

To estimate the stability of the rwcGC and rwGT base pairs relative to the rhGG base pair, we derived a thermodynamic model to simulate these titration results. The sharp transition from hetero- to homoduplexes in the crystals suggests that discrimination between the various reverse base pairs occurs at a highly cooperative step during crystallization. This is most likely during the nucleating event that initiates the formation of the crystals. In this model, we consider only two different duplexes that can crystallize, the homoduplex (GG) formed by the d(GCGCGCG) strands in solution (G_S) and the heteroduplex (GY) formed by G_S and Y_S in solution (where Y represents either the d(CCGCGCG) or d(TCGCGCG) strand). The homoduplexes (YY) are not considered in the model because these have not been observed to crystallize in these studies. Qualitatively, we can think of this mechanism as one in which the initiation step of crystallization is the formation of a lattice in a solution consisting of either the homo- or heteroduplexes of the DNA. Once formed, this lattice allows the addition of duplexes that are identical to those already in the lattice, excluding all others. Thus, discrimination between base pairs results from the probability of bringing n number of

identical duplexes together to form the initial nucleating lattice in solution (Eq. 1 and 2). Subsequent growth of the crystal is then dependent on the composition of this initial lattice.

$$n\mathrm{GG_S} <-> \mathrm{GG_L} \qquad \text{Eq. 1}$$

$$n\mathrm{GY_S} <-> \mathrm{GY_L} \qquad \text{Eq. 2}$$

The duplexes in the lattices (GG_L and GY_L) are considered to be in equilibrium with the free duplexes in solution (GG_S and GY_S) in this model. Thus, the probability for forming GY_L versus GG_L is related by Eq. 3 (where K_{GYL} and K_{GGL} are the equilibrium constants for formation of the nucleating lattices).

$$[\mathrm{GG_L}]/[\mathrm{GY_L}] = ([\mathrm{GG_S}]^n K_{GGL})/([\mathrm{GY_S}]^n K_{GYL}) = ([\mathrm{GG_S}]/[\mathrm{GY_S}])^n (K_{GGL}/K_{GYL}) \qquad \text{Eq. 3}$$

The ratios of the homo- and heteroduplexes in solution are dependent on the ratios of the single-strands added to the solutions (Eq. 4-6).

$$2\mathrm{G_S} <-> \mathrm{GG_s}, \text{ and } K_{GGS} = [\mathrm{GG_S}]/[\mathrm{G_S}]^2 \qquad \text{Eq. 4}$$

$$\mathrm{G_S} + \mathrm{Y_S} <-> \mathrm{GY_s}, \text{ and } K_{GYS} = [\mathrm{GY_S}]/([\mathrm{G_S}]\,[\mathrm{Y_S}]) \qquad \text{Eq. 5}$$

$$[\mathrm{GG_S}]/[\mathrm{GY_S}] = (K_{GGS}[\mathrm{G_S}]^2)/(K_{GYS}[\mathrm{G_S}][\mathrm{Y_S}]) = (K_{GGS}[\mathrm{G_S}])/(K_{GYS}[\mathrm{Y_S}]) \qquad \text{Eq. 6}$$

Since the base pairing and structure of the resulting duplex regions are identical between the homo- and heteroduplexes, we can assume that $K_{GGS} = K_{GYS}$. Thus, $[\mathrm{GG_S}]/[\mathrm{GY_S}] = [\mathrm{G_S}]/[\mathrm{Y_S}]$. During the slow nucleation steps, the DNA is predominantly in solution, so that $[\mathrm{G_S}]/[\mathrm{Y_S}]$ can be assumed to be the ratio of the two strands added to solution. The relative probabilities for initiating the GG_L and GY_L lattices are related to the strand compositions of the crystallization setup and to the difference in free energy between the lattices of the two types of duplexes ($\Delta\Delta G^\circ$) (Eq. 7).

$$[\mathrm{GG_L}]/[\mathrm{GY_L}] = ([\mathrm{G_S}]/[\mathrm{Y_S}])^n K_{GGL}/K_{GYL}) = ([\mathrm{G_S}]/[\mathrm{Y_S}])^n e^{-\Delta\Delta G^\circ/RT} \qquad \text{Eq. 7}$$

The mass spectrometry analyses provide the compositions of the crystals as the types of single-strands and not of DNA duplexes. Thus, the observed quantities are the ratios of the individual strands that are associated with each type of duplex that is potentially found in each crystal. The GG_L species contributes two strands of d(GCGCGCG) while GY_L contributes one such strand. The observed quantity of d(GCGCGCG) in the lattice (G_{obs}) is thus 2 GG_L + GY_L. The amount of Y strand observed from the mass spectra (Y_{obs}) is simply GY_L, since this is the only species that contributes. Thus, the ratio $G_{obs}/Y_{obs} = 2(GG_L/GY_L) + 1$. The data we obtained showed the complete conversion from the hetero- to the homoduplex; we therefore represent the observed data as the fraction of Y_{obs} (f_Y, from 0.5 to 0.0). Finally, this can be related to the ratio of the two strands in solution by Eq. 8.

$$f_Y = 0.5/\{([G_S]/[Y_S])^n e^{-\Delta\Delta G^\circ/RT} + 1\} \qquad \text{Eq. 8}$$

Using Eq. 8, we can simulate titration curves for $[G_S]/[Y_S]$ from 1:1 to 5:1 and values for n = 1 to 16, in which the transition from the heteroduplex ($f_Y = 0.5$) to the homoduplex ($f_Y = 0$) occurs at $[G_S]/[Y_S] \approx 3.5$ (Fig. 7a). The simulated curve at n = 16 reproduces the sharp transition observed between the two lattice types (assuming a 10% accuracy in defining f_Y from the mass spectrometry data). This suggests that the discrimination between rwcGC and rhGG base pairs in the crystal lattice occurs when a minimum of 16 duplexes (the contents of four complete unit cells) associate to form an initiation complex. The assembly of four complete unit cells in the lattice, therefore, appears to be the defining step for the composition of the crystals. The assembly of a minimum of four unit cells produces an environment in

Figure 4.7. Comparison of the fractions of the d(YCGCGCG) strand (f_Y = Y / (Y+G), Y = observed quantity of d(YCGCGCG) and G = quantity of d(GCGCGCG) in the crystal) as the ratio of the d(GCGCGCG) strand relative to the d(YCGCGCG) strand ($[G_S]/[Y_S]$) is increased from 1:1 to 4:1 in the crystallization setups, as observed in the crystals and calculated using the model and Eq. 8 in the text. (a) The fraction of the strand d(CCGCGCG) (f_C) in the crystal of the heteroduplex d(GCGCGCG)/d(CCGCGCG) were determined at each $[G_S]/[Y_S]$ ratio by MALDI mass spectrometry (closed circles, with errors approximated at 10%). The simulated curves were calculated with the number of duplexes in the initiation complex (n) set at 1 to 16 in Eq. 8, and $\Delta\Delta G^\circ$ set to a value that places the midpoint of the transition at $[G_S]/[Y_S]$ = 3.5 for each curve. (b) The fraction of the d(YCGCGCG) strand in crystals of the heteroduplexes d(GCGCGCG) /d(YCGCGCG) are compared for Y = cytidine (closed cicles) and Y = thymidine (open boxes). The values for f_Y were calculated using Eq. 8, with n = 16 and the difference in free energy between the rhGG and the rwcGC and rwGT reverse base pairs normalized to each interaction expected within and between four unit cells ($\Delta\Delta G^\circ$/interaction) set at 0.0, 0.5, 1.0, and 1.5 kcal/mol.

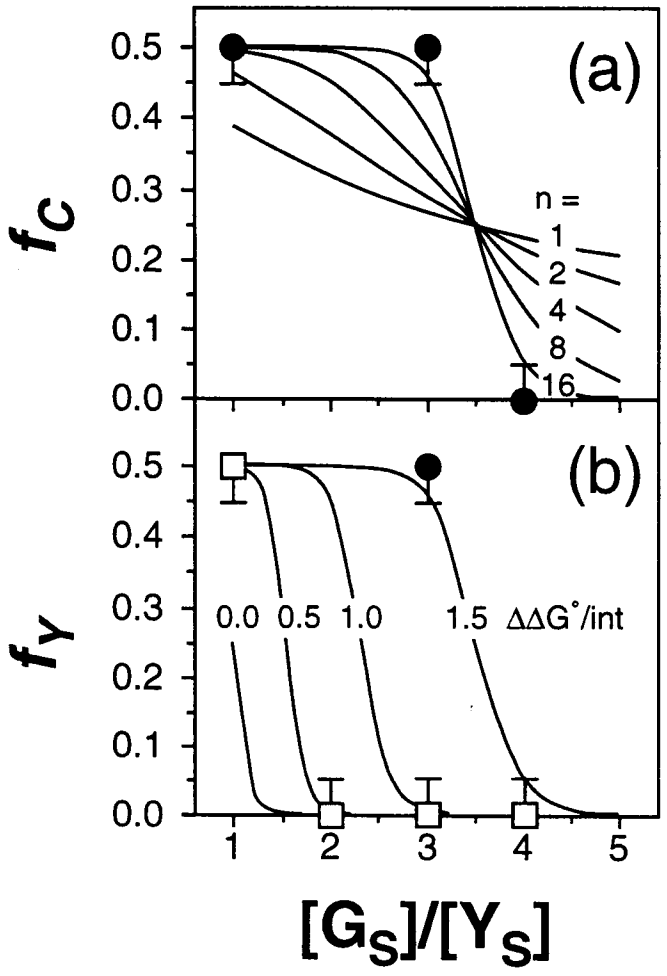


Figure 4.7.

which all the possible intermolecular interactions (within the unit cell and between unit cells) are represented.

With the value of n = 16, or four unit cells defining the minimum size of the initiation complex, we can now estimate the free energy differences between rhGG and the other two types of base pairs. In this initiation complex, there is one interduplex base pair interaction within each unit cell and one between each unit cell. For the most compact assembly, there are a total of 4 base pairs within and 4 between the four unit cells, yielding a total of 8 identical base pairing interactions. If we now consider the total difference in free energy of interaction as the sum of 8 individual interactions ($\Delta\Delta G^\circ$/int), the titration curves can be simulated, again using Eq. 8 (Figure 4.7b). This assumes that only the differences in the free energy for pairing bases between the duplexes are important. A comparison of the simulated and the mass spectrometry results show the rwcGC base pair to be ~1.5 kcal/mol more stable than rhGG, and rwGT to be ~0.5 kcal/mol more stable than rhGG, within this crystal system.

4.4. Discussion

We present here the structures of reverse base pairs formed by pairing the 5'-overhanging nucleotide of d(Gm^{5}CGCGCG) with either the 5'-guanosine of d(Gm^{5}CGCGCG) to form a d(G·G) reverse Hoogsteen, the 5'-thymidine of d(TCGCGCG) to form a d(G·T) reverse wobble, or the 5'-cytidine of d(CCGCGCG) to form a d(G·C) reverse Watson-Crick base pair (Figure 4.2). The common guanosine nucleotide is stacked against the standard Z-DNA and is relatively inflexible in its conformation (Figure 4.3).

In all three structures, this guanosine is in the *syn* conformation, extending the alternating *anti-syn* character of the Z-DNA duplex.

The extra-helical nucleotide, which distinguishes each type of base pair, adopts either the *anti* conformation (in rhGG and rwcGC) or the *syn* conformation (in rwGT). This is likely determined by the requirements for orienting the base so that the proper face is presented to the stacked guanine to form the base pair. For example, the thymine of the rwGT base pair must lie towards the major groove of the guanine in order to properly pair its N3 and O4 atoms with the N1 and O6 of the guanine base (Figure 4.2c). This pushes the pyrimidine base closer to the duplex in the lattice and thus requires that the nucleotide adopt the more compact *syn* conformation (Figure 4.5c). The cytosine and guanine bases of the rwcGC and rhGG base pairs, on the other hand, are pulled towards the minor groove, and thus can adopt the more extended *anti* conformation (Figure 4.2a & 4.2b). Indeed, the rhGG is an extreme case, where the extra-helical guanosine is extended to the point where it now has only a single hydrogen bond within the base pair. A second hydrogen bonding interaction occurs between the N2 amino nitrogen of this extended guanosine and the O2P oxygen of cytidine 9 in a third duplex (Figure 4.5a). This feature of the purine-purine base pair is not observed in either of the base pairs formed with the pyrimidines.

The structure of the rwcGC base pair is identical to the analogous base pairs observed in tRNA (Westhof *et al.*, 1988). The root-mean-square (rms) deviation between the atoms of the bases in this structure and that of the $tRNA^{Phe}$ is 0.160 Å. In comparison, the average rms deviation of the d(G·C) Watson-Crick base pairs within the structure of rwcGC is 0.124 Å, and relative to the d(G·C) Watson-Crick base pair in the structure of B-DNA at 2.5 Å resolution is 0.184 Å. Likewise, the bases of the rwGT base pair are similar

to the G·U reverse wobble base pair observed in the solution structure of the unusually stable RNA hairpin formed by the sequence r(GGAC(UUCG)GUCC) (Varani *et al.*, 1991).

In all of these crystal structures, only one well-defined type of reverse base pair was observed, even though we attempted to solve the structures with all possible combinations of base pairs. It is easy to rationalize the common stacked guanosine nucleotide, since this is an extension of the highly invariant Z-DNA duplex. It is less obvious, however, why the extra-helical base of the heteroduplexes should always be the pyrimidine, since the pairing two guanines of the homoduplex can obviously be accommodated by this same lattice. Mass spectrometry analysis of these heteroduplex showed that crystals grown with equal ratios of each strand are composed only of the heteroduplexes. In addition, when looking at the population of crystals, all the crystals analyzed in this way had this same composition. We concluded that the difference must result from the greater stability of the rwcGC and rwGT base pairs over the rhGG in this environment.

Since these base pairs result from crystal lattice interactions, we were able to estimate the relative stability of each type of base pair within nearly identical environments by analyzing the compositions of crystals grown with various ratios of the parent strand (d(GCGCGCG)) with the paired pyrimidine strand (d(CCGCGCG) or d(TCGCGCG)). In the case of the rwcGC structure, the transition from all heteroduplexes to all homoduplexes in the crystals was highly cooperative. The simplest model for this transition defines the discriminating step as the initial nucleation event of crystal growth. This is not the only interpretation of the cooperativity. However, it seems reasonable that once a particular lattice type is established, the growth of the crystal can proceed only by extending this same lattice. The minimum

cooperativity coefficient that fits the data requires that 16 duplexes be involved in the nucleation of the crystal. This is the content of four complete unit cells, suggesting that this is the minimum size of a regular lattice. In this model, all the interactions between molecules within the unit cell and between unit cells are established within this minimum lattice. Thus, once this initial lattice is formed, the structure of the pockets in which the extra-helical bases must fit becomes defined. These pockets then allow only a single type of duplex to add to the crystal lattice.

The free energies determined here are specific for the base pairing and lattice interactions observed in these crystals, with the rwcGC > rwGT > rhGG base pairs. All of these form two hydrogen bonds within the lattice, either directly to the stacked guanine, as in the rwcGC and rwGT structures, or one with the stacked guanine and one with the backbone of a third duplex as in the rhGG structure. With both the atomic resolution structures and the thermodynamic data in hand, we can ask what accounts for the differences in stability. The rwcGC base pair places the pyrimidine nucleotide in the *anti* conformation. In addition, waters in the plane of the base pairs accommodate the unfulfilled hydrogen bonding groups of the bases and thus may contribute to the overall stability of the rwcGC base pairing. The relative instability of the rwGT base pair likely results from the disfavored *syn* conformation adopted by this pyrimidine base (Haschemeyer & Rich, 1967; Neumann *et al.*, 1979). Furthermore, no waters were observed bridging the guanine and thymine bases, as has been observed in the structures of "normal" d(G·T) wobble base pairs (Ho *et al.*, 1985; Kneale *et al.* 1985; Hunter *et al.*, 1986).

The relative instability of the rhGG base pair is likely associated with the sliding of the extra-helical guanine away from the stacked guanine,

leaving only a single distorted hydrogen bond between the two bases. Still, one would expect the hydrogen bond interactions between the N2 amino nitrogen and the phosphate of a third duplex (there is one direct hydrogen bond and a second water mediated interaction) to compensate for the hydrogen bond lost between the two bases. However, the coordination of three duplexes to form all the observed hydrogen bond interactions may not occur during the nucleation event. It is not clear whether a more standard rhGG base pair forms at these initial stages, which is then distorted by subsequent lattice interactions. To test these possibilities, we crystallized a reverse d(A·A) base pair, but its structure is significantly different from the d(G·G) so that they are not comparable at this time. The structure and thermodynamics of a d(G·I) base pair in this crystal system should resolve this problem. Thus, although the rhGG base pair observed here is highly distorted by the crystal lattice, and thus may not represent the structure expected in RNA structures, it does serve as a reference for comparing the stability of rwcGC and rwGT base pairs.

4.5. Materials and Methods

4.5.1 Synthesis, purification, & crystallization

The seven base oligonucleotides d(Gm^{5}CGCGCG), d(GCGCGCG), d(TCGCGCG), and d(CCGCGCG) were synthesized using phosphoramidite chemistry on an Applied Biosystems DNA synthesizer in the Center for Gene Research and Biotechnology at Oregon State University. Size exclusion chromatography on a Sephadex G-25 column was used to remove salts, blocking groups, and prematurely terminated oligonucleotides. The oligonucleotides were lyophilized, redissolved in 30 mM sodium cacodylate

buffer (pH 7.0), and used for crystallization without further purification. The oligonucleotides d(GCGCGCG) and d(Gm^{5}CGCGCG) produce homoduplexes with two G overhangs. The oligonucleotides d(GCGCGCG) and d(TCGCGCG) were mixed in an equimolar ratio to yield duplexes with G and T overhangs. Likewise, d(GCGCGCG) and d(CCGCGCG) were similarly mixed to produce duplexes with C and G overhangs.

Crystals of the duplexes were grown at room temperature by vapor diffusion in sitting drop setups. All sequences crystallized from initial solutions containing 0.5 mM DNA (single-strands), 50 mM sodium cacodylate (pH 7.0), 1 mM $MgCl_2$, 2.5 mM cobalt hexaammine (Aldrich), and 5% (v/v) 2-methyl-2,4-pentanediol (MPD), equilibrated against a reservoir of 17% MPD. Blocky, amber-colored plates appeared within one week and reached dimensions of up to 0.4 mm x 0.4 mm x 0.1 mm within two weeks.

4.5.2 X-ray diffraction data collection

X-ray diffraction data for the crystals were collected at room temperature using a Siemens P4 diffractometer with a Siemens HI-STAR area detector (Cu-K_α radiation from a sealed tube source). The raw data were integrated and scaled using the software package SAINT (Siemens, Inc.). All crystals were isomorphous (in the space group $P2_12_12_1$ with nearly identical unit cell dimensions (Table 4.3)), and diffracted to high resolution (1.68 to 1.9 Å).

4.5.3 Structure solution and refinement

The structure of d(Gm^{5}CGCGCG) was solved first using features of the diffraction data to construct an appropriate model for molecular replacement. The dimensions of the unit cell suggested that the heptamer

was in the Z-DNA form and aligned along the crystallographic *c*-axis. The space group was the same as that of most previously crystallized Z-DNA hexamers, and the *a* and *b* unit cell axial lengths were very similar to those of the archetypal Z-DNA hexamer d(CGCGCG) (Wang *et al.*, 1979, 1981). The length of the *c* axis (~52Å) could accommodate 14 base pairs with a helical rise of 3.7 Å, suggesting that the helical axes of two stacked heptamers were aligned parallel to the *c*-axis. This was confirmed by the Patterson map, which showed base-pair cross vectors spaced 3.7 Å apart along the *c*-axis.

The alternating purine-pyrimidine heptamer sequence could pair to form d(Gm^{5}CGCGC) hexamer duplexes with guanosines overhanging the 3'-ends or d(m^{5}CGCGCG) hexamer duplexes with guanosines overhanging the 5'-ends. This latter case seemed more likely. Sequences of the type d(CGCGCG) typically crystallize as Z-DNA, while d(Gm^{5}CGCGC) crystallize as A-DNA (Mooers *et al.*, 1995). This is consistent with studies of Quadrifoglio *et al.*, (1984) showing that short alternating d(CG) sequences, but not alternating d(GC) sequences form Z-DNA in solution. Both possibilities, however, were tested. Models of both types of structures were constructed using the program InsightII (Biosym/MSI Corp.) with standard helical parameters for Z-DNA. The initial 3'-overhang model was generated by removing the cytidine nucleotide at the 5'-end of the duplex structure of d(CGCGCGCG), while the 5'-overhang model was constructed by removing the nucleotide at the 3'-end of the duplex structure of d(GCGCGCGC).
For each initial model, the best orientation and position of the heptamer in the unit cell was located using the rotation and translation search functions of the program AMORE (Nazava, 1994). The R-values for the best initial solutions were 44.7% for the 5'-overhang model and 47.7% for the

Table 4.3. Diffraction data from the crystal structures of the sequences d(Gm^{5}CGCGCG)$_2$, d(GCGCGCG)/d(CCGCGCG), and d(GCGCGCG)/(TCGCGCG) which crystallized in the space group $P2_12_12_1$. The crystals are represented by their overhangs as rhGG, rwcGC, and rwGT, respectively.

	rhGG	rwcGC	rwGT
Unit Cell Dimensions (Å):			
a	20.34	20.32	20.28
b	29.62	29.54	29.41
c	51.93	51.84	51.89
Measured Reflections	28,961	20,832	5,784
Unique Reflections	5,324	5,340	2,601
Resolution Range (Å)	29.60-1.46	14.20-1.48	14.00-1.88
$R_{sym}(I)$* for I > 0 (%)	8.6	8.3	7.2

*$R_{sym}(I) = 100 \times (\Sigma hkl\ (|\ I - \langle I \rangle\ |\ /\ |\ \langle I \rangle\ |)) / n$ where I is the integrated intensity of a reflection, <I> is the average of all observations of the reflection and its symmetry equivalents, and n is the number of unique reflections. All positive, non-zero reflections were merged.

3'-overhang model. Subsequent refinement of these models demonstrated that the 5'-overhang structure was correct. The model was refined to an R-value of 41.6% using the rigid body and rigid parts (with the bases, deoxyriboses, and phosphates treated as independent groups) refinement functions in X-PLOR (Brünger, 1992) using a new parameter file for the DNA (Parkinson *et al.*, 1996) and data from 8-3.5 Å. After simulated annealing (with a starting temperature of 3000°K) the R-factor was reduced to 30.6% for data from 8-2.2 Å.

The actual conformations of the overhanging nucleotides were determined from electron density maps calculated using only the phasing information from the six base pairs of the d(CGCGCG) duplex. Difference maps generated using XtalView (MacRee, 1992) showed that only one of the 5'-terminal guanines was stacked. The other overhanging nucleotide was flipped out and extended so that it base paired with the stacked overhanging guanine of a neighboring duplex (Figure 4.2). The refinement converged to a final R of 20.7% (R_{free} = 27.8%) at 1.68 Å resolution, with 49 waters added, including one cobalt hexaammine and one hydrated magnesium complex. The coordinate error of less than 0.2 Å was estimated from a Luzzati plot (Luzzati, 1952).

The final structures of $d(GCGCGCG)_2$ and $d(Gm^5CGCGCG)_2$ were identical in all respects. The structures of the heteroduplexes of d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG)/d(CCGCGCG) were solved in a similar fashion, using the d(CGCGCG) duplex region of the d(GCGCGCG) structure as the starting model, and defining the conformations of the overhanging bases from difference maps. The statistics for the refined

Table 4.4. Refinement results for the crystal structures of d(Gm^{5}CGCGCG)$_2$, (rhGG), d(GCGCGCG)/d(CCGCGCG) (rwGC), and d(GCGCGCG)/d(TCGCGCG) (rwGT).

	rhGG	rwGC	rwGT
R-working (%)	20.7	20.9	19.1
R-free (%)	28.7	27.2	28.6
Resolution Range (Å)	8-1.68	8-1.80	8-1.90
Data Completeness (%)[1]	82.2	84.5	76.3
Number of Reflections	3,538	3,127	2,230
No. of nonhydrogen DNA atoms	286	281	283
No. of water molecules[2]	49	64	47
Ave. B-factors			
DNA atoms	15.9	13.3	17.5
water atoms	31.1	31.5	29.6
r.m.s. deviation from ideality			
bond lengths (Å)	0.007	0.008	0.008
bond angles (degrees)	1.477	1.304	1.311

[1]Exclusive of the reflections sequestered in the test set to calculate R_{free}. Refinements were made with a three sigma on F cutoff on each dataset.
[2]Each structure was refined with a cobalt hexaammine and a hexaaquomagnesium complex in addition to these waters.

structures of d(Gm^{5}CGCGCG)$_2$, d(GCGCGCG)/d(TCGCGCG), and d(GCGCGCG)/d(CCGCGCG) are listed in Table 4.4.

In the crystallization solution of the heteroduplexes, homoduplexes of d(GCGCGCG) may have been present. As an independent check on the composition of the crystals of the heteroduplexes, four large crystals of the duplex d(GCGCGCG)/ d(TCGCGCG) were isolated, carefully washed with cold crystallization solution lacking DNA, dissolved at 90°C in 50 μL of deionized and distilled water, and individually analyzed by mass spectrometry as described below. The spectra from each crystal showed that the d(GCGCGCG) and d(TCGCGCG) strands were present in a 1:1 ratio (Fig. 4.6).

Helical parameters for the Z-DNA duplex regions were analyzed using the program *NASTE* (Nucleic Acid STructure Evaluation), a program developed in this laboratory for analysis of the helical parameters in Z-DNA structures. The final coordinates and structure factors for the structures of d(Gm^{5}CGCGCG)$_2$, d(GCGCGCG)/(CCGCGCG), and d(GCGCGCG)/ d(TCGCGCG) have been deposited in the Nucleic Acid Database (Berman *et al.*, 1992). Their reference codes are ZDGB55, ZDG054, and ZDG056, respectively.

4.5.4 Crystallization and mass spectrometry analyses of crystals grown with different strand compositions

To estimate the relative stability of the different duplex pairings, crystals were grown from solutions in which d(GCGCGCG) was mixed with either d(CCGCGCG) or d(TCGCGCG) in molar ratios of 1:1, 2:1, 3:1, and 4:1, with the total concentration held constant at 1.4 mM. The crystallization solutions contained the identical buffers, salts, and precipitants as those that yielded the original crystals. After two weeks, crystals were isolated from the

setups, washed several times with deionized water, and then dissolved into deionized water for analysis by mass spectrometry.

Dialyzed DNA from the dissolved crystals were analyzed by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry using a custom-built time-of-flight instrument, as previously described (Jensen *et al.*, 1993). All samples were analyzed with a matrix of 10 mg/ml of 2,4,6-trihydroxyacetophenone (Aldrich) in a 50 mM diammonium hydrogen citrate/50% acetonitrile solution. For each mass analysis, 0.5 μL of DNA analyte was mixed in a 1:2 ratio with the matrix solution and 0.5 μL of this mixture was placed on the sample stage. At the first sign of crystal formation (generally 10-15 seconds after deposition when viewed with a stereo microscope), the droplet was gently wiped with a lab tissue, leaving a seed layer of crystallites on the surface of the sample stage. Another 0.5 μL of the analyte/matrix mixture was then deposited on top of the seed layer and then gently rinsed with cold (4°C) Millipore-filtered water. Each mass spectrum was recorded as the sum of 30 consecutive spectra, each produced by a single pulse of 355 nm photons from a Nd:YAG laser (Spectra Physics). Mass spectra were calibrated using ion-signals from the matrix.

4.6. Acknowledgments

This work has been supported by grants from the National Science Foundation (MCB 9304467), the National Institutes of Health (R5GM54538A) and the Environmental Health Sciences Center (EHSC) at Oregon State University (NIEHS ES00210). We thank Dr. Victor Hsu for sharing his computer resources, Dr. Herman Schreuder of Marion Merrell Dow Research Institute for his program *Luzzati*, Dr. Lilo Barofsky in the

Analytical Core unit of the EHSC for mass spectrometry analyses, Beth Basham for her *NASTE* program, and Cyndi Thompson for helpful discussions concerning non-canonical base pairs in RNA.

4.7. References

Allmang, C., Mougel, M., Westhof, E., Ehresmann, B., & C. Ehresmann. (1994). Role of conserved nucleotides in building the 16S rRNA binding site of *E. coli* ribosomal protein S8. *Nucleic Acids Res.* **22**, 3708-3714.

Battiste, J. L., Tan, R., Frankel, A. D., & Williamson, J. R. (1994). Binding of an HIV Rev responsive element RNA induces formation of purine-purine base pairs. *Biochemistry*, **33**, 2741-2747.

Battiste, J. L., Moa, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., & Williamson, J. R. (1996). α-Helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science*, **273**, 1547-1551.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Geblin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R., & Schneider, B. (1992). The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751-759.

Brennan, R. G., Westhof, E., & Sundaralingam, M. (1986). Structure of a Z-DNA with two different backbone chain conformations. Stabilization of the decadeoxyoligonucleotide d(CGTACGTACG) by $[Co(NH_3)_6]^{3+}$ binding to the guanine. *J. Biomol. Struc. & Dyn.* **3**, 649-665.

Brünger, A. (1992). *X-PLOR Manual, Version 3.1.* Yale University, New Haven, USA.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., & Doudna, J. A. (1996). Crystal structure of a Group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.

Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., PrangÇ, T., & Kennard, O. (1994). Structure of a mispaired RNA double helix at 1.6-Å

resolution and implications for the prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **91**, 4160-4164.

Gessner, R.V., Frederick, C.A., Quigley, G.J., Rich, A., Wang, A.H.-J. (1989). The molecular structure of the left-handed Z-DNA double helix at 1.0 angstrom atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). *J. Biol. Chem.*, **264**, 7921-7935.

Gessner, R. V., Quigley, G. J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H. & Rich, A. (1985). Structural basis for stabilization of Z-DNA by cobalt hexaammine and magnesium cations. *Biochemistry*, **24**, 237-240.

Haschemeyer, A. E. V. & Rich, A. (1967). Nucleoside conformations: an analysis of steric barriers to rotation about the glycosidic bond. *J. Mol. Biol.* **27**, 369-384.

Ho, P. S., Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., & Rich, A. (1985). G·T wobble base-pairing in Z-DNA at 1.0 Å atomic resolution: the crystal structure of d(CGCGTG). *EMBO J.* **4**, 3617-3623.

Ho, P. S. & Mooers, B. H. M. (1997). Z-DNA crystallography. *Biopolymers (Nucleic Acid Science)* **44**, 65-90.

Holbrook, S. R., Cheong, C., Tinoco Jr., I., & Kim, S.-H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature*, **353**, 579-581.

Hunter, W. N., Kneale, G., Brown, T., Rabinovich, D., & Kennard, O. (1986). Refined crystal structure of an octanucleotide duplex with G·T mismatched base pairs. *J. Mol. Biol.* **190**, 605-618.

Ibba, M., Hong, K.-W., Sherman, J. M., Sever, S., & Sîll, D. (1996). Interactions between tRNA identity nucleotides and their recognition sites in glutaminyl-tRNA synthetase determine the cognate amino acid affinity of the enzyme. *Proc. Nat. Acad. Sci., USA* **93**, 6953-6958.

Jensen, O. N., Barofsky, D. F., Young, M. C., von Hippel, P. H., Swenson, S., and Seifried, S. E. (1993). Direct observation of UV-crosslinked protein-nucleic acid complexes by matrix-assisted laser desorption ionization mass spectroscopy. *Rapid Commun. Mass Spectrom.* **7**, 496-501.

Jiang, F., Kumar, R. A., Jones, R. A., & Patel, D. J. (1996). Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature*, **382**, 183-186.

Kim, S.-H. (1978). The three dimensional structure of transfer RNA and its functional implications. *Adv. Enzymol.* 46, 279-315.

Kneale, G., Brown, T., & Kennard, O. (1985). G·T base-pairs in a DNA helix: the crystal structure of d(G-G-G-G-T-C-C-C). *J. Mol. Biol.* **186**, 805-814.

Luzzati, P. V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr.* **5**, 802-810.

MacRee, D. (1992). A visual protein crystallographic software system for X11/Xview. *J. Mol. Graphics* **10**, 44-46.

Malinina, L., Urpi, L., Salas, X., Huynh-Dinh, T., & Subirana, J. A. (1994). Recombination-like structure of d(CCGCGG). *J. Mol. Biol.* **243**, 484-493.

Mooers, B. H. M., Schroth, G. P., Baxter, W. W., & Ho, P. S. (1995). Alternating and non-alternating dG-dC hexanucleotides crystallize as canonical *A*-DNA. *J. Mol. Biol.* **249**, 772-784.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. sect. A*, **50**, 157-163.

Neumann, J.-M., Guschlbauer, W., & Tran-Dinh, S. (1979). Conformation and flexibility of GpC and CpG in neutral aqueous solution using ^{1}H nuclear-magnetic-resonance and spin-lattice-relaxation time measurements. *Eur. J. Biochem.* **100**, 141-148.

Parkinson, G. N., Vojtechovsky, J., Clowney, L., & Berman, H. M. (1996). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr., Sect. D.*, **52**, 57-64.

Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68-74.

Quadrifoglio, F., Manzini, G., & Yathindra, N. (1984). Short oligonucleotides with d(G-C)*n* sequences do not assume left-handed conformation in high salt conditions. *J. Mol. Biol.* **175**, 419-423.

Quigley, G. J. & Rich, A. (1976). Structural domains of transfer RNA molecules. *Science*, **194**, 796-806.

Rippe, K., Fritsch, V., Westhof, E., & Jovin, T. M. (1992). Alternating d(G-A) sequences form a parallel-stranded DNA homoduplex. *EMBO J.* **11**, 3777-3786.

Robinson, H., van Boom, J. H., & Wang, A. H.-J. (1994). 5'-CGA motif induces other sequences to form homo base-paired parallel-stranded DNA duplex: the structure of (G-A)*n* derived from four DNA oligomers containing $(G\text{-}A)_3$ sequence. *J. Am. Chem. Soc.* **116**, 1565-1566.

Skelly, J. V., Edwards, K. J., Jenkins, T. C., & Neidle, S. (1993). Crystal structure of an oligonucleotide duplex containing G·G base-pairs: influence of mispairing on DNA backbone conformation. *Proc .Nat. Acad. Sci., USA* **90**, 804-808.

Varani, G., Cheong, C. & Tinoco, Jr., I. (1991). Structure of an unusually stable RNA hairpin. *Biochemistry*, **30**, 3280-3289.

Vojtechovsky, J., Eaton, M. D., Gaffney, B., Jones, R., & Berman, H. M. (1995). Structure of a new crystal form of a DNA dodecamer containing T·(O^6Me)G base pairs. *Biochemistry*, **34**, 16632-16640.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A. & Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, **282**, 680-686.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H. & Rich, A. (1981). Left-handed double helical DNA: variations in the backbone conformation. *Science*, **211**, 171-176.

Westhof, E., Dumas, P., & Moras, D. (1988). Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallogr., sect. A*, **44**, 112-123.

Chapter 5

Discussion

The packing of DNA in crystals is controlled by modifying the duplex length and sequence. These modifications are made to reduce the distortions of crystal structures, to crystallize two helical forms of identical sequences, and to make the two stands of a heteroduplex distinguishable in the lattice. In Chapter 2, the crystallization of shorter sequences yield four crystal structures of A-DNA that are less distorted from fiber model A-DNA than longer A-DNA crystal structures. In Chapter 3, the use of overhangs to crystallize identical sequences in different helical conformations is introduced by crystallizing the sequence d(GCGTACGC) as A-DNA and as B-DNA. In Chapter 4, the use of two different overhangs is introduced to crystallize heteroduplexes without orientational disorder in crystals.

Sequence length, an important crystal packing variable, does not appear to limit the helical form found in crystals. For example, A-DNA has been crystallized in the tetramer, octamer, decamer, and dodecamer length classes in the past (Dickerson, 1992). In Chapter 2, four crystal structures of the first A-DNA hexamers are presented. Since the publication of our report (Mooers *et al.*, 1995), the crystal structures of the first B-DNA hexamer and B-DNA octamer have been published (Tari & Secco, 1995; Tereshko *et al.*, 1996). Now, almost every length class and helical form combination has at least one representative crystal structure.

Changing the length of a sequence of interest may be one way of crystallizing it in a lattice that distorts the structure to a small degree. For

example, the A-DNA octamers crystal structures of A-DNA ocatmers are the most numerous representatives of A-DNA crystal structures. Unfortunately, they are also the most distorted by crystal packing (Ramakrishnan & Sundaralingam, 1993a). When extended to twelve base pairs in length, a dodecamer sequence crystallizes with a conformation that is close to that of fiber model A-DNA (Verdaguer *et al.*, 1991). When reduced to a hexamer length, four A-DNA forming duplexes crystallize in a conformation that is closer to that of fiber model A-DNA than longer A-DNA crystal structures, including the A-DNA dodecamer. The hexamers did not have the base-to-base interduplex hydrogen bonds that distort the octamers, which partially explains this difference. Instead, the two crystal structures of the alternating dG-dC sequence each have four symmetric base-to-backbone hydrogen bonds, which alter the base stacking slightly because the backbone extends in response to this interduplex interaction. The non-alternating hexamer duplexes do not have any direct interduplex hydrogen bonds.

The reduced distortion of the crystal structures of the hexamers was somewhat unexpected. Longer DNA oligonucleotides are thought to experience fewer "end effects" than shorter oligonucleotides because the longer oligonucleotides have more base pairs protected between their ends. In A-DNA at least, longer DNA oligonucleotides are not always less distorted.

It does not follow, however, that all A-DNA hexamers will crystallize in a conformation like that fiber A-DNA. The hexamer d(Gm^{5}CTAGC) crystallizes in a slightly different lattice in an extended conformation not like fiber A-DNA because the two base-to-base hydrogen bonds to one strand distort this structure asymmetrically (Mooers *et al.*, unpublished results). Likewise, two additional dodecamer sequences crystallize with large

distortions in lattices different from the above-mentioned dodecamer (Bingman *et al.*, 1992a,b). Thus, strand length is not the sole determinant of the distortion that a DNA duplex experiences in the crystalline state.

The four crystal structures presented in Chapter 2 demonstrate for the first time a clear correlation between base sequence and structural variations in A-DNA. The two crystal structures of the alternating dG-dC sequence have alternating patterns of variation in the local helical parameters roll, helical twist, and helical rise. In contrast, the two crystal structures of the non-alternating dG-dC sequence have non-alternating patterns of variation in the same local helical parameters. There were no significant differences in structural variations due to differing numbers of methylated cytosines in the alternating dG-dC crystal structures and due to the presence or absence of methylated cytosines in the non-alternating dG-dC crystal structures.

The base sequence dependent features and the fiber model-like global shape of the A-DNA hexamer crystal structures suggests that they could be used to separate crystal packing effects from base sequence effects in the octamers of related sequences. A correction function is being derived from the hexamer crystal structures that removes most of the crystal packing induced distortions in octamer crystal structures (Mooers *et al.*, unpublished results).

Self-complementary sequences extended by one nucleotide have an odd length, so they have one nucleotide overhanging each end of the duplex when they form an antiparallel duplex. Overhanging bases have been used in the past to stabilize the crystal packing of protein*DNA complexes (Schultz *et al.*, 1990) and to study base triplets formed by the overhangs and neighboring duplexes in the crystal (Van Meervelt *et al.*, 1995; Vlieghe *et al.*, 1996a,b). However, no successful attempt at using overhanging bases to

stabilize a duplex in a second helical form has been published. In Chapter 3, the sequence d(GCGTACGC), which crystallizes as A-DNA, is extended the 3'-direction by one guanosine and crystallizes as B-DNA. The crystal structures of a duplex in two helical forms may provide valuable structural data for theoretical studies of how the A- and B-DNA helical forms inter-convert.

Also in Chapter 3, the concept of "packing driving boxes", sequence elements that supposedly direct the orientation of duplexes during crystallization by forming specific interduplex hydrogen bonds (Timsit & Moras, 1992), is extended to the termini of duplexes. The sequence d(GCAATTGCG) is designed to crystallize with the same end-on-end stacking and base triplet formation as d(GCGTACGCG). The sequence elements GC were retained at the 5' and 3' ends of the octamer duplex d(GCxxxxGC), where x is any nucleotide, because the 5' terminal guanines participate in base triplet formation. In addition, the GC base step has helical parameters which are very similar to those of the GC base steps from the Drew-Dickerson type dodecamers. In the dodecamers, these base steps are involved in the formation of interduplex d(G·G) base pairs similar to the type seen in the crystal structures of the nonamers. This suggests that the template sequence for coaxial stacking and base triplet formation in the nonamers is d(GCxxxxGCG). The nucleotides in the first and ninth positions are obviously required for the formation of the interduplex d(G·G) base pair. The nucleotides in the seventh position may need to be purines because the major groove face of each guanine in the seventh position interacts directly with neighboring duplexes via two hydrogen bonds with bridging aquomagnesium complexes. Pyrimidines have only one hydrogen bond acceptor or donor one their major groove faces, so they may not be able to

form such a stable interaction. Thus, it may be possible to generalize the template to d(GYxxxxRCG), where Y is a pyrimidine and R is a purine. In any event, our extension of the "packing driving box" concept succeeds in controlling the crystal packing of the sequence d(GCAATTGCG) to a limited extent. This sequence crystallizes with the same coaxial stacking of duplexes and the same base triplet, although in two different lattices. The packing driving boxes fail to control the orientation of the columns of duplexes since they crystallize in a 62° or 63° cross in a monoclinic lattice rather than in the 90° cross in the tetragonal lattice of the crystal structure of the nonamer d(GCGTACGCG). Thus, some packing driving boxes do not always lead to a unique packing arrangement. The extra helical guanosine at both ends of the duplex form identical interduplex interactions which makes it impossible to for the lattice to distinguish between the two strands of non-self-complementary sequences.

Non-self-complementary sequences are notorious for crystallizing with end-to-end disorder because both of the backbones of such heteroduplexes are almost identical although the sequence in the strand is not (DiGrabriele and Steitz, 1993; Schroth *et al.*, 1993). This problem is generally avoided by crystallizing self-complementary sequences, which greatly reduces the variety of sequences that can be studied. In Chapter 3, a Z-DNA heptamer system is presented in which two heteroduplexes crystallize without end-on-end disorder. In addition, a thermodynamic model of lattice nucleation is used to explain the absence of homoduplexes in the lattices of heteroduplexes. The results suggest that it should be possible to crystallize non-self-complementary sequences without end-on-end disorder in this system as long as the overhanging bases are different. The uniqueness of the two ends of the duplex is the consequence of the differences in the crystalline

environments surrounding each overhang. This is a unique feature of the *P2$_1$2$_1$2$_1$* lattice that could be exploited in other DNA crystal systems in this space group. For example, the decamer d(GGCCAATTGG) crystallizes in space group *P2$_1$2$_1$2$_1$* with a overhanging guanosine dinucleotide at one end of the duplex remaining stacked and a overhanging guanosine dinucleotide at the other end flipping out of the helical stack (Vlieghe *et al.*, 1996b). It may be possible to also crystallize non-self-complementary sequences in this system if each strand has distinguishable overhanging ends.

While the double helical structure of DNA can lead to difficulties in studying DNA structure in crystals, it also holds the solutions of these problems. We have demonstrated how modifications of strand sequence and length can be used to control the packing of DNA duplexes in crystals. We have also shown how these modifications can be used to solve three problems that have limited the study of DNA in crystals.

Bibliography

Allmang, C., Mougel, M., Westhof, E., Ehresmann, B., & C. Ehresmann. (1994) Role of conserved nucleotides in building the 16S rRNA binding site of *E. coli* ribosomal protein S8. *Nucleic Acids Res.* **22**, 3708-3714.

Arnott, S. & Hukins, D. W. L. (1972) Optimized parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* **47**, 1504-1509.

Babcock, M. S., Pednault, E. P. D. & Olson, W. K. (1994) Nucleic acid structure analysis: mathematics for local cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* **237**, 125-156.

Ban, C., Ramakrishnan, B., & Sundaralingam, M. (1996) Crystal structure of the self-complementary 5'-purine start decamer d(GCGCGCGCGC) in the Z-DNA conformation. I. *Biophysical J.* **71**, 1215-1221.

Basham, B., Schroth, G. P., & Ho, P. S. (1995) An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA. *Proc. Natl. Acad. Sci. USA*, **92**, 6464-6468.

Battiste, J. L., Tan, R., Frankel, A. D., & Williamson, J. R. (1994). Binding of an HIV Rev Responsive Element RNA induces formation of purine-purine base pairs. *Biochemistry*, **33**, 2741-2747.

Battiste, J. L., Moa, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., & Williamson, J. R. (1996). α-Helix-RNA major groove recognition in an HIV-1 Rev Peptide-RRE RNA complex. *Science*, **273**, 1547-1551.

Berman, H. M. (1994) Hydration of DNA: Take 2. *Curr. Opin. Struct. Biol.*, **4**, 345-350.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, D., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992) The nucleic acid database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751-759.

Berstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. F., Brice, M. D., Rodgers, J. D., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) The Protein Data Bank: A computer-based archival file for macromolecular structures." *J. Mol. Biol.* **112**, 535-542.

Bingman, C., Li, X., Zon, G. & Sundaralingam, M. (1992) Crystal and molecular structure of d(GTGCGCAC): investigation of the effects of base sequence on the conformation of octamer duplexes. *Biochemistry*, **31**, 12803-12812.

Bingman, C. A., Zon, G., & Sundaralingam, M. (1992a). Crystal and molecular structure of the A-DNA dodecamer d(CCGTACGTACGG). *J. Mol. Biol.*, **227**, 738-756.

Bingman, C. A., Jain, S., Zon, G., & Sundaralingam, M. (1992b). Crystal and molecular structure of the alternating dodecamer d(GCGTACGTACGC) in the A-DNA form: Comparison with the isomorphous non-alternating dodecamer d(CCGTACGTACGG). *Nucleic Acids Res.*, **20**, 6637-6647.

Brennan, R. G., Westhof, E., & Sundaralingam, M. (1986). Structure of a Z-DNA with two different backbone chain conformations. Stabilization of the decadeoxyoligonucleotide d(CGTACGTACG) by $[Co(NH_3)_6]^{3+}$ binding to the guanine." *J. Biomol. Struc. & Dyn.* **3**, 649-665.

Broomhead, J. M. (1951) The structures of pyrimidines and purines. IV. The crystal structure of guanine hydrochloride and its relation to the adenine hydrochloride. *Acta Crystallogr.*, **4**, 92-100.

Brown, T. & Hunter, W. N. (1997) Non-Watson-Crick base associations in DNA and RNA revealed by single crystal x-ray diffraction methods: mismatches, modified bases, and nonduplex DNA. *Biopolymers (Nucleic Acid Sciences)*, **44**, 91-103.

Brünger, A. (1992) *X-PLOR Manual, Version 3.1.* Yale University, New Haven, USA.

Bugg, C. E., Thomas, J. M., Sundaralingam, M., & Roa, S. T. (1971) Stereochemistry of nucleic acids and thier constituents. X. Solid-state base-stacking patterns in nucleic acid constituents and polynicleotides". *Biopolymers*, **10**, 175-219.

Calladine, C. R. (1982) Mechanism of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.* **161**, 343-352.

Calladine, C. R. & Drew, H. R. (1984). A base-centered explanation of the B-to-A transition in DNA. *J. Mol. Biol.* **178**, 773-782.

Cate, J. H. and Doudna, J. A. (1996) Metal-binding sites in the major groove of a large ribozyme domain. *Structure*, **4**, 1221-1229.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., & Doudna, J. A. (1996). Crystal structure of a Group I Ribozyme Domain: Principles of RNA packing. *Science*, **273**, 1678-1685.

Chandrasekaran, R., Wang, M., He, R.-G., Puigianer, L. C., Byler, M. A., Millane, R. P. & Arnott, S. (1989) A re-examination of the crystal structure of A-DNA using fiber diffraction data. *J. Biomol. Struct. Dynam.* **6**, 1189-1202.

Coll, M., Solans, X., Font-Altaba, M., and Subriana, J. A. (1987) Crystal and molecular structure of the sodium salt of the dinucleotide duplex d(CpG)". *J. Biomol. Struct. Dyn.*, **4**,797-811.

Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

Crick, F. H. C. & Waton, J. D. (1954) The complementary structure of deoxyribonuicleic acid. *Proc. Roy. Soc. (London) Ser. A* **223**, 80-96.

Cruse, W. B. T., Egert, E., Kennard, O., Sala, G. B., Salisbury, S. A., and Viswamitra, M. A. (1983) Self base pairing in a complementary deoxydinucleotide monophosphate duplex: crystal and molecular structure of deoxycytidylyl-(3'-5')-deoxyguanosine. *Biochemistry*, **22**, 1833-1839.

Cruse, W. B. T. , Salisbury, S. A., Brown, T., Cosstick, R., Eckstein, F. & Kennard, O. (1986) Chiral phosphorothioate analogues of B-DNA, the crystal structure of Rp-d | Gp(S)CpGp(S)CpGp(S)C | . *J. Mol. Biol.* **192**, 891-905.

Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., PrangÇ, T., & Kennard, O. (1994) Structure of a mispaired RNA double helix at 1.6-Å resolution and implications for the prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **91**, 4160-4164.

Dickerson, R. E. (1992) DNA structure from A to Z. In *Methods in Enzymology, Vol. 211* (Lilley, D. M. J. & Dahlberg, J. E. eds.), pp. 67-111.

Dickerson, R. E., Goodsell, D. S., Kopka, M. L., and Pjura, P. E. (1987) The effect of crystal packing on oligonucleotide double helix structure. *J. Biomol. Struct. Dyn.* **5**, 557-579.

Dickerson, R. E., Goodsell, D. S. & S. Neidle. (1994) '...the tyranny of the lattice...'. *Proc. Nat. Acad. Sci., U.S.A.* **91**, 3579-3583.

Dickerson, R. E., Goodsell, D. S., & Kopka, M. L. (1996) MPD and DNA bending in crystals and in solution. *J. Mol. Biol.*,**256**, 108-125.

DiGrabriele, A. D. and Steitz, T. A. (1993) A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024-1039.

Drew, H. R. & Dickerson, R. E. (1981) Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **151**, 535-556.

Drew, H. R., McCall, M. J., & Calladine. (1988) Recent studies of DNA in the crystal. *Ann. Rev. Cell. Biol.*, **4**, 1-20.

Duocet, J., Beoit, J. P. Cruse, W. B. T., Prange, T. & Kennard, O. (1989) Coexistence of A- and B-form DNA in a single crystal lattice. *Nature (London)*, **337**, 190-192.

Egli, M., Gessner, R. V., Williams, L. D., Quigley, G. J., van der Marel, G., van Boom, J. H., Rich, A., & Frederick, C. A. (1990). Atomic-resolution structure of the cellulose synthetase regulator cyclic diguanylic acid. *Proc. Natl. Acad. Sci. USA* **87**, 3235-3239.

El Hassan, M. A. & Calladine, C. R. (1996) Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.*,**259**, 95-103.

Franklin, R. E., & Gosling, R. G. (1953a) The structure of sodium thymonucleate fibers I: The influence of water content. *Acta Cryst.* **6**, 673-677.

Franklin, R. E., & Gosling, R. G. (1953b) The structure of sodium thymonucleate fibers II: the cylindrical symmetrical Patterson function. *Acta Cryst.* **6**, 678-685.

Franklin, R. E., & Gosling, R. G. (1953c) Molecular configuraton of sodium thymonucleate. *Nature (London)*, **171**, 740-741.

Frederick, C. A., Saal, D., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. & Rich, A. (1987) The crystal structure of d(GGm^{5}CCGGCC): the effect of methylation on A-DNA structure and stability. *Biopolymers* , **26**, S145-S160.

Frederick, C. A., Quigley, G. J., Teng, M.-K., Coll, M., van der Marel, G. A., van Boom, J. H., Rich, A. & Wang, A. H.-J. (1989) Molecular structure

of an A-DNA decamer d(ACCGGCCGGT). *Eur. J. Biochem.* **181**, 295-307.

Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J.H. & Rich, A. (1982) Molecular structure of $(m^5CG)_3$: the role of the methyl group on 5-methyl-cytosine in stabilizing Z-DNA. *Nucleic Acids Res,* **10**, 7879-7892.

Fuller, W., Wilkins, M. H. F., Wilson, H,. R., & Hamilton, L. D. (1965) The molecular configuration of deoxyribonucleic acid IV. X-ray Diffraction study of the *A* form. *J. Mol. Biol.* **12**, 60-80.

Gessner, R.V., Frederick, C.A., Quigley, G.J., Rich, A., Wang, A.H.-J. (1989) The molecular structure of the left-handed Z-DNA double helix at 1.0 Ångstrom atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). *J. Biol. Chem.*, **264**, 7921-7935.

Gessner, R. V., Quigley, G. J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H. & Rich, A. (1985) Structural basis for stabilization of Z-DNA by cobalt hexaammine and magnesium cations. *Biochemistry,* **24**, 237-240.

Goodsell, D. S., Kaczor-Grzeskowiak, M., and Dickerson, R. E. (1994) The crystal structure of C-G-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps. *J. Mol. Biol.*, **239**, 79-96.

Guzikevich-Guerstein, G. & Shakked, Z. (1996) A novel form of the DNA double helix imposed on the TATA box by the TATA-binding protein. *Nat. Struct. Biol.* **3**, 32-37.

Hagerman, P. J. (1985) Sequence dependence of curvature of DNA: A test of the phasing hypothesis. *Biochemistry,* **24**, 7033-7037.

Hagerman, P. J. (1986) Sequence-directed curvature of DNA. *Nature,* **321**, 449-450.

Haran, T. E., Shakked, Z., Wang, A. H.-J., & Rich, A. (1987) The crystal structure of d(CCCCGGGGG): A new A-form variant with an extended backbone conformation. *J. Biomol. Struct. Dynam.*, **5**, 199-217.

Haschemeyer, A. E. V. & Rich, A. (1967). Nucleoside conformations: An analysis of steric barriers to rotation about the glycosidic bond. *J. Mol. Biol.* **27**, 369-384.

Heinemann, U. and Alings, C. (1989) Crystallographic study of one turn of G/C rich B-DNA. *J. Mol. Biol.* **210**, 369-381.

Heinemann, U., Alings, C. & Hahn, M. (1994) Crystallographic studies of DNA helix structure. *Biophysical Chem.* **50**, 157-167.

Herbert, A. & Rich, A. (1996). The biology of left-handed Z-DNA. *J. Biol. Chem.*, **271**, 11595-11598.

Ho, P. S., Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., & Rich, A. (1985). G·T wobble base-pairing in Z-DNA at 1.0 Å atomic resolution: the crystal structure of d(CGCGTG). *EMBO J.* **4**, 3617-3623.

Ho, P. S., Kagawa, T. F., Tseng, K., Schroth, G. P. & Zhou, G. (1991) Prediction of a crystallization pathway for Z-DNA hexanucleotides. *Science*, **254**, 1003-1006.

Ho, P. S. & Mooers, B. H. M. (1997) Z-DNA crystallography. *Biopolymers (Nucleic Acid Sciences)* **44**, 65-90.

Holbrook, S. R., Cheong, C., Tinoco Jr., I., & Kim, S.-H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature (London)*, **353**, 579-581.

Holbrook, S. R., Sussman, J. L., Warrant, R. W., Church, G. M., and Kim, S.-H. (1977) RNA-ligand interactions: (I) Magnesium binding sites in yeast $tRNA^{Phe}$. *Nuc. Acid. Res.* **4**, 2811-2820.

Hou, Y.-M., Westhof, E., and Giege, R. (1993) An unusual RNA tertiary interaction has a role for the specific aminoacylation of a transfer RNA. *Proc. Natl. Acad. Sci. USA*, **90**, 6776-6780.

Hunter, C. A. and Lu, X.-J. (1997) DNA base-stacking interactions: A comparison of theoretical calculations with oligonucleotide X-ray crystal structure. *J. Mol. Biol.* **265**, 603-613.

Hunter, W. N., Kneale, G., Brown, T., Rabinovich, D., & Kennard, O. (1986). Refined crystal structure of an octanucleotide duplex with G·T mismatched base pairs". *J. Mol. Biol.* **190**, 605-618.

Ibba, M., Hong, K.-W., Sherman, J. M., Sever, S., & Sîll, D. (1996). Interactions between tRNA identity nucleotides and their recognition sites in glutaminyl-tRNA synthetase determine the cognate amino Acid affinity of the enzyme". *Proc. Nat. Acad. Sci., USA* **93**, 6953-6958.

Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D. & Schyolkina, A. K. (1974) The B to A transition of DNA in solution. *J. Mol. Biol.*, **87**, 817-833.

Ivanov, V. I., Minchenkova, L. E., Chernov, B. K., McPhie, P., Ryu, S., Garges, S., Barber, A. M., Zhurkin, V. B., Adhya, S. (1995) CRP-DNA complexes: inducing the A-like form in the binding sites with an extended central spacer." *J. Mol. Biol.*, **245**, 228-240.

Jain, S., Zon, G. & Sundaralingam, M. (1987) The potentially Z-DNA forming sequence d(GTGTACAC) crystallizes as A-DNA. *J. Mol. Biol.* **197**, 141-145.

Jain, S. & Sundaralingam, M. (1989) Effect of the crystal environment on conformation of the DNA duplex. *J. Biol. Chem.* **264**, 12780-12784.

Jain, S., Zon, G. & Sundaralingam, M. (1989) Base only binding of spermine in the deep groove of the A-DNA octamer d(GTGTACAC). *Biochemsitry*, **28**, 2360-2364.

Jensen, O. N., Barofsky, D. F., Young, M. C., von Hippel, P. H., Swenson, S., and Seifried, S. E. (1993). Direct observation of UV-crosslinked protein-nucleic acid complexes by matrix-assisted laser desorption ionization mass spectroscopy. *Rapid Commun. Mass Spectrom.* **7**, 496-501.

Jiang, F., Kumar, R. A., Jones, R. A., & Patel, D. J. (1996). Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature*, **382**, 183-186.

Kagawa, T. F., Stoddard, D., Zhou, G., & Ho, P. S. (1989) Quantitative analysis of DNA secondary structure from solvent-accessible surfaces: The B- to Z-DNA transtion as a model. *Biochemistry*, **28**, 6642-6651.

Kagawa, T. F. (1994) Studies of the hydrophobic effect and its contribution to the stability, crystallization, and helix packing of Z-DNA. Ph. D. Thesis, Oregon State Univ., Corvallis, Oregon.

Kennard, O., Cruse, W. B. T., Nachman, J., Prange, T., Shakked, Z., & Rabinovich, D. (1986) Ordered water in an A-DNA octamer at 1.7Å resolution. *J. Biomol. Struct. Dynam.* **3**, 623-647.

Kennard, O. & Salisbury, S.A. (1993) Oligonucleotide X-ray structures in the study of conformations and interactions of nucleic acids. *J. Biol. Chem.* **268**, 10701-10704.

Kim, S.-H. (1978). The three dimensional structure of transfer RNA and its functional implications. *Adv. Enzymol.* **46**, 279-315.

Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z., & Steits, T. A. (1979) A hypothesis on a specific sequence dependent conformation of DNA and its relation to binding of the the *lac* repressor protein. *J. Mol. Biol.*, **131**, 669-80.

Kneale, G., Brown, T., & Kennard, O. (1985). G·T base-pairs in a DNA helix: the crystal structure of d(G-G-G-G-T-C-C-C). *J. Mol. Biol.* **186**, 805-814.

Langridge, R., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F., & Hamilton, L. D. (1957). Molecular structure of deoxyribonucleic acid (DNA). *J. Biophys. Biochem. Cytol.*, **3**, 767-778.

Levitt, M. (1978) How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc. Natl. Acad. Sci USA* **75**, 640-644.

Liu, L. F. & Wang, J. C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **84**, 7024-7027.

Lomonossoff, G. P., Butler, P. J. G., & Klug, A. (1981) Sequence-dependent variation in the conformation of DNA. *J. Mol. Biol.*, **149**, 745-60.

Luzzati, P. V. (1952). "Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Cryst.* **5**, 802-810.

MacRee, D. (1992). A Visual protein crystallographic software system for X11/Xview. *J. Mol. Graphics* **10**, 44-46.

Malinina, L., Urpi, L., Salas, X., Huynh-Dinh, T., & Subirana, J. A. (1994) Recombination-like structure of d(CCGCGG). *J. Mol. Biol.* **243**, 484-493.

Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. (1982) Bent helical structure in kinetoplast DNA. *Proc. Natl. Acad. Sci. USA*, **79**, 1025-1029.

McCall, M., Brown, T. & Kennard, O. (1985) The crystal structure of d(GGGGCCCC): A model for poly (dG)·poly(dC). *J. Mol. Biol.* **183**, 385-396.

Mohr, S. C., Sokolov, N. V. H. A., He, C., & Setlow, P. (1991) Binding of small acid-soluble spore proteins from *Bacillus subtilis* changes the conformation of DNA from B to A. *Proc. Natl. Acad. Sci. USA* **88**, 77-81.

Mooers, B. H. M., Schroth, G. P., Baxter, W. B., & Ho, P. S. (1995). Alternating and non-alternating A-DNA dG-dC hexanucleotides crystallize as canonical *A*-DNA. *J. Mol. Biol.* **249**, 772-784.

Mooers, B. H. M., Eichman, B. F., & Ho, P. S. (1997) The structures and stabilities of d(G·G) reverse Hoogsteen, d(G·T) reverse wobble, and d(G·C) reverse Watson-Crick base pairs in DNA crystals. (in press) *J. Mol. Biol.*

Navaza, J. (1994) *AMoRe*: an automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.

Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987) The structure of oligo(dA)·oligo(dT) tract and its biological implications. *Nature (London)*, **330**, 221-226.

Neumann, J.-M., Guschlbauer, W., & Tran-Dinh, S. (1979). Conformation and flexibility of GpC and CpG in neutral aqueous solution using ^{1}H nuclear-magnetic-resonance and spin-lattice-relaxation time measurements. *Eur. J. Biochem.* **100**, 141-148.

O'Brien, E. J. (1967) Crystal structures of two complexes containing guanine and cytosine derivatives. *Acta Crystallogr.* **23**, 92-106.

Pabo, C. O. & Sauer, R. T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053-1095.

Parkinson, G. N., Vojtechovsky, J., Clowney, L., & Berman, H. M. (1996) New parameters for the refinement of nucleic acid-containing structures. *Acta Cryst., Sect. D.*, **52**, 57-64.

Peck, L. J. & Wang, J. C. (1981) Sequence dependence of the helical repeat of DNA in solution. *Nature (London)*, **292**, 375-378.

Peticolas, W. L., Wang, Y., & Thomas, G. A. (1988) Some rules for predicting the base-sequence dependence of DNA conformation. *Proc. Natl. Acad. Sci. USA*, **85**, 2579-2583.

Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature (London)*, **372**, 68-74.

Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Model for an RNA tertirary interaction from the structure of an intermolecular complex

between a GAAA tetraloop and an RNA helix. *Nature (London)*, **372**, 111-113.

Pohl, F., & Jovin, T. (1972) Salt-induced cooperative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.*, **6**, 375-396.

Privé, G., Hienneman, U., Chandrasegaran, S., Kopka, L. K. M., & Dickerson, R. E. (1987) Helix geometry, hydration, and G·A mismatch in a *B*-DNA decamer. *Science* **238**, 498-504.

Quadrifoglio, F., Manzini, G., & Yathindra, N. (1984). Short oligonucleotides with d(G-C)*n* sequences do not assume left-handed conformation in high salt conditions. *J. Mol. Biol.* **175**, 419-423.

Quigley, G. J. (1991) Understanding the crystal packing of Z-DNA. *J. Crystal Growth*, **110**, 131-136.

Quigley, G. J. & Rich, A. (1976). Structural domains of transfer RNA molecules. *Science*, **194**, 796-806.

Ramakrishnan, B. & Sundaralingam, M. (1993a) Crystal packing effects on A-DNA helix parameters: a comparative study of the isofroms of the tetragonal and hexagonal family of octamers with differing base sequences. *J. Biomol. Struct. Dynam.* **11**, 11-26.

Ramakrishnan, B. and Sundaralingam, M. (1993b) High resolution crystal structure of the A-DNA decamer d(CCCGGCCGGG). Novel intermolecular base-paired G*(G·C) triplets. *J. Mol. Biol.*, **231**, 431-444.

Rhodes, D. & Klug, A. (1981) Sequence-dependenct helical periodicity of DNA. *Nature (London)*, **292**, 378-380.

Rich, A. and RajBhandary, U. L. (1976) Transfer RNA: molecular structure, sequence, and properties. *Annu. Rev. Biochem.*, **45**, 805-860.

Rippe, K., Fritsch, V., Westhof, E., & Jovin, T. M. (1992). Alternating d(G-A) sequences form a parallel-stranded DNA homoduplex. *EMBO J.* **11**, 3777-3786.

Robinson, H., van Boom, J. H., & Wang, A. H.-J. (1994). 5'-CGA motif induces other sequences to form homo base-paired parallel-stranded DNA duplex: the structure of (G-A)*n* derived from four DNA oligomers containing $(G\text{-}A)_3$ sequence. *J. Am. Chem. Soc.* **116**, 1565-1566.

Saenger, W. (1983) *Principles of Nucleic Acid Structure.* Springer-Verlag New York Inc., New York.

Saenger, W., Hunter, W. N. & Kennard, O. (1986) DNA conformation is determined by the economics in the hydration of phosphate groups. *Nature (London),* **324,** 385-388.

Satchwell, S. C., Drew, H. R. & Travers, A. A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.,***191,** 659-675.

Schneider, B., Cohen, D. M., Schleifer, L., Srinivasan, A. R., Olson, W. K. & Berman, H. M. (1993) A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophysical J.* **65,** 2291-2303.

Schroth, G. P., Kagawa, T. F., & Ho, P. S. (1993) Structure and thermodynamics of nonalternating C·G base pairs in Z-DNA: The 1.3-Å crystal structure of the asymmetric hexanucleotide d(m^5CGGCm^5CG)·d(m^5CGCCm^5CG). *Biochemistry,* **32,** 13381-13392.

Schultz, S.C., Shields, G.C., & Steitz, T.A. (1990). Crystallization of the *Escherichia coli* catabolite gene activator protein with its DNA binding site. The use of modular DNA. *J. Mol. Biol.* **213,** 159-166.

Schultz, S.C., Shields, G.C., & Steitz, T.A. (1991) Crystal structures of a CAP-DNA complex: the DNA is bent by 90°. *Science* **253,** 1001-1007

Shakked, Z., Guerstein-Guzikevih, G., Eisentsein, M., Frolow, F., & Rabinovich, D. (1989) The conformation of the DNA double helix in the crystal is dependent on its environment. *Nature* **342,**456-460.

Skelly, J. V., Edwards, K. J., Jenkins, T. C., & Neidle, S. (1993). Crystal structure of an oligonucleotide duplex containing G·G base-pairs: influence of mispairing on DNA backbone conformation. *Proc .Nat. Acad. Sci., USA* **90,** 804-808.

Spink, N., Nunn, C. M., Vojtechovsky, J., Berman, H. M., and Neidle, S. (1995) "Crystal structure of a DNA decamer showing a novel pseudo four-way helix-helix Junction. *Proc. Natl. Acad. Sci.,* USA **92,** 10767-10771.

Sponer, J., Florian, J., Hobza, P., and Leszcynski, J. (1996) Nonplanar DNA base pairs. *J. Biomol. Struc. Dyn.* **13,** 827-833.

Suzuki, M. & Yagi, N. (1995) Stereochemical basis of DNA bending by transcription factors. *Nuc. Acid Res.* **23,** 2083-2091.

Tari, L. W. & Secco, A. S. (1995) Base-pair opening and spermine binding--B-DNA features displayed in the crystal structure of a *gal* operon fragment: Implications for protein-DNA recognition. *Nuc. Acid Res.*, **23**, 2065-2073.

Tereshko, V., Urpí, L. Malinina, L., Huynh-Dinh, T., and Subirana, J. A. (1996) Structure of the B-DNA oligomers d(CGCTAGCG) and d(CGCTCTAGCGCG) in new crystal forms. *Biochemistry*, **35**, 11589-11595.

Timsit, Y., & Moras, D. (1991) Groove-backbone interactions in B-DNA--implications for DNA condensation and recombination. *J. Mol. Biol.* **221**:919-940.

Timsit, Y. & Moras, D. (1992) Crystallization of DNA. In *Methods in Enzymology, Vol. 211* (Lilley, D. M. J. & Dahlberg, J. E. eds.), pp. 409-429.

Timsit, Y. & Moras, D. (1994) DNA self-fitting: the double helix direct the geometry of its supramolecular assembly. *EMBO J.* , **13**, 2737-2746.

Timsit, Y. & Moras, D. (1995) Self-fitting and self-modifying properties of the B-DNA molecule. *J. Mol. Biol.*, **251**, 629-647.

Tippin, D. B. & Sundaralingam, M. (1996) Structure of d(CCCTAGGG): comparisons with nine isomorphous octamer sequences reveals four distinct patterns of sequence dependent intermolecular interactons. *Acta Cryst. Sec. D*, **52**, 997-1003.

Urpí, L., Tereshko, V. Malinina, L., Huynh-Dinh, T., & Subirana, J. A. (1996) Structural comparion between the d(CTAG) sequence in oligonucleotides and *trp* and *met* repressor-operator complexes. *Nature Struct. Biol.* **3**, 325-328.

Van Meervelt, L, Vlieghe, D., Dautuant, A., Gallois, B., Precigoux, G. and Kennard, O. (1995) High resolution structure of a DNA helix forming (C·G)*G base triplets. *Nature (London)*, **374**, 742-744.

Varani, G., Cheong, C. & Tinoco, Jr., I. (1991) Structure of an unusually stable RNA hairpin. *Biochemistry*, **30**, 3280-3289.

Verdaguer, N., Aymamí, J., Fernández-Forner, D., Fita, I., Coll, M., Huynh-Dinh, T., Igolen, J. & Subirana, J. A. (1991). Molecular structure of a complete turn of A-DNA. *J. Mol. Biol.* **221**, 623-635.

Vlieghe, D., Van Meervelt, L., Dautant, A., Gallois, B. Pre'cigoux, G. and Kennard, O. (1996a) Formation of (C·G)*G triplets in a B-DNA duplex with overhanging bases. *Acta Cryst. Sect. D*, **52**, 766-775.

Vlieghe, D., Van Meervelt, L., Dautant, A., Gallois, B. Pre'cigoux, G. and Kennard, O. (1996b) Parallel and antiparallel $(G{\cdot}GC)_2$ triple helix fragments in a crystal structure. *Science*, **273**, 1702-1705.

Vojtechovsky, J., Eaton, M. D., Gaffney, B., Jones, R., & Berman, H. M. (1995) Structure of a new crystal form of a DNA dodecamer containing T·(O^6Me)G base pairs". *Biochemistry*, **34**, 16632-16640.

Wahl, M. C. & Sundaralingam, M. (1997) Crystal structures of A-DNA duplexes. *Biopolymers (Nucleic Acid Sciences)* **44**, 45-63.

Wang, A. H.-J., Quigley, G.J., Kolpak, F. J., Crawford, J. L., van der Marel, G. A., van Boom, J. H. & Rich, A. (1979). Molecular structure of a left handed double helical DNA fragment at atomic resolution. *Nature (London)*, **282**, 680-686.

Wang, A. J.-H., Fujii, S., van Boom, J. H. & Rich, A. (1982) Molecular structure of the octamer d(G-G-C-C-G-G-C-C): Modified A-DNA. *Proc. Nat. Acad. Sci., U.S.A.* **89**, 534-538.

Wang, A. H.-J. & Teng. (1988) Crystallization and crystal packing analysis of DNA oligonucleotides. *J. Cryst. Growth*, **90**, 295-310.

Watson, J. D. & Crick, F. H. C. (1953) Molecular structure of nucleic acids. *Nature*, **171**, 737.

Westhof, E. (1988) "Water: An integral part of nucleic acid structure." *Ann. Rev. Biophys. Biophys. Chem.* **17**, 125-144.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) "Crystal structure analysis of a complete turn of DNA. *Nature (London)*, **287**, 755-758.

Wittig, B., Wolfl, S., Dorbic, T., Vahrson, W., & Rich, A. (1992) Transcription of human c-*myc* in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *J. Biol. Chem.*, **267**, 11846-11855.

Yanagi, K., Privé, G. G., Goodsell, D. S., and Dickerson, R. E. (1991) Analysis of local helix geometry in three B-DNA decamer and eight dodecamers. *J. Mol. Biol.*, **217**, 201-214.

Young, M. A., Ravishanker, G. , Beveridge, D. L., & Berman, H. M. (1995) Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes. *Biophysical J.*, **68**, 2454-2468.