

## AN ABSTRACT OF THE THESIS OF

Todd F. Kagawa for the degree of Doctor of Philosophy in Biochemistry and Biophysics  
presented on December 1, 1994. Title: Studies on the Hydrophobic Effect and Its  
Contribution to the Stability, Crystallization, and Helix Packing of Z-DNA.

Abstract Approved: \_\_\_\_\_

Redacted for Privacy

Dr. Pul Shing Ho

The studies presented here use the B- to Z-DNA transition and Z-DNA crystallization as model systems to determine the contribution of solvent interactions to macromolecular structures. The substituent groups of naturally occurring and modified nucleotide bases affect the hydration and thus the stability of right-handed B-DNA and left-handed Z-DNA. The free energy for alternating pyrimidine-purine sequences in the B- and Z-conformations are quantitated as solvent free energies (SFE) from their solvent-accessible surfaces. The effect of methylation of the C5 carbon of pyrimidine bases on the stability of Z-DNA was analyzed in comparisons of d(TA) *versus* d(UA) dinucleotides. In the minor groove, the N2 amino group of purine bases was examined by comparing the stability of d(CG) *versus* d(CI), and d(TA) *versus* d(TA') dinucleotides as Z-DNA. These comparisons correctly accounted for the observed effects of the major and minor groove substituent groups on the relative stability of Z-DNA. These predictions were confirmed by comparing the amount of salt required to crystallize various hexanucleotides as Z-DNA.

The relative contribution of solvent and steric interactions in DNA assembly were studied crystallographically using an asymmetric hexanucleotide which assumes two discrete and discernible orientations in the crystal lattice. How this sequence orients in the lattice is a direct measure of the DNA-DNA interactions at the surface of the crystal.

These interaction free energies were directly correlated with differences in solvation for 5 sequences where there was effectively no differences in steric interactions for the two orientations. In the singular case where a large steric clash was expected, the orientation was in fact determined by this interaction. The comparisons indicate that both solvent and van der Waals interactions are significant in DNA assembly with van der Waals interactions dominating in situations with large steric interactions.

The analyses of the B- to Z-DNA transition and Z-DNA crystallization based on SFEs support the significance of solvent interactions in determining the structure and assembly of macromolecules.

**Studies on the Hydrophobic Effect and Its Contribution  
to the Stability, Crystallization, and Helix Packing  
of Z-DNA**

**by**

**Todd F. Kagawa**

**A THESIS**

**Submitted to**

**Oregon State University**

**in partial fulfillment of  
the requirements for the  
degree of**

**Doctor of Philosophy**

**Completed December 1, 1994**

**Commencement June 1995**

Doctor of Philosophy thesis of Todd F. Kagawa presented on December 1, 1994

APPROVED:

Redacted for Privacy

---

Professor of Biochemistry and Biophysics in charge of major

Redacted for Privacy

---

Head of 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

---

Todd F. Kagawa, Author

---

## Table of Contents

1.	Introduction.....	1
	Statement of Purpose.....	3
	Current Views on the Hydrophobic Effect.....	8
	Introduction to the B- to Z-DNA Conformational Transition.....	11
2.	Quantitative Analysis of DNA Secondary Structure From Solvent-Accessible Surfaces: The B- to Z-DNA Transition as a Model.....	24
	Abstract.....	25
	Introduction.....	26
	Experimental Methods.....	30
	Results.....	33
	Discussion.....	56
3.	Effects of Base Substituents on the Hydration of B- and Z-DNA: Correlations to the B- to Z-DNA Transition.....	59
	Abstract.....	60
	Introduction.....	61
	Experimental Methods.....	65
	Results.....	69
	Discussion.....	87
4.	The Contribution of the Hydrophobic Effect to Z-DNA Helix Packing: A Crystallographic Study.....	93
	Introduction.....	93
	Statement of Purpose.....	96
	Introduction to the Crystal Structure of the Asymmetric Z-DNA Hexamer.....	97
	Introduction to the Crystal Packing of Z-DNA Hexamers in the <i>P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></i> Lattice.....	104
	Materials and Methods.....	117
	Results.....	131
	Discussion.....	158
5.	Discussion.....	161
6.	Bibliography.....	165

## List of Figures

1.1	Space Filling Models of B- and Z-DNA Conformations.....	15
2.1	Plot of Free Energies of Hydration ( $\Delta G^{\circ}_H$ ) versus Solvent-Accessible Surfaces (SAS) Calculated for Small Organic Molecules.....	34
2.2	Titration of poly(dG-dm <sup>5</sup> C) with NaCl.....	42
2.3	Comparison of the Calculated Differences in Hydration Energies of Z- and B-DNA with the Experimentally Determined B- to Z-DNA Transition Energies.....	44
2.4	Stacking Diagram of the d(CG), d(TA), and d(UA) Dinucleotides and Their Solvent Accessible Surfaces.....	47
2.5	Plot of Free Energies of Transition Calculated from the Hydration Free Energies versus Logarithm of Cationic Strengths for Crystallizing Synthetic Hexanucleotides as Z-DNA.....	55
3.1	Titration of poly(dG-dC) and poly(dG-dm <sup>5</sup> CG) with MgCl <sub>2</sub> , LiCl, NaCl and KCl.....	71
3.2	Solvent Accessible Surfaces of d(TA) and d(TA') Dinucleotides as B- and Z-DNA.....	75
3.3	Comparison of the log of the Cation Strength ( $\log CS$ ) for Crystallization of Z-DNA Hexanucleotides to the Calculated B- to Z-DNA Transition Free Energies ( $\Delta G^{\circ}_T(B-Z)$ ).....	81
3.4	Solvent Accessible Surfaces of d(CG) and d(CI) Dinucleotides as B- and Z-DNA.....	85
3.5	Thermodynamic Cycle Comparing the Substituent Effects at the Major and Minor Grooves of the Double Helix on the Stability of Z- versus B-DNA for Alternating Pyrimidine-Purine Dinucleotides.....	91
4.1	Space Filling Model of the Asymmetric Z-DNA Hexamer Duplex Showing the Modification Sites.....	99
4.2	Crystal Packing Diagrams of Z-DNA Hexamers in the Orthorhombic <i>P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></i> Space Group.....	105
4.3	Crystal Packing Diagrams of the Two Orientations Available to the Hexamer Duplex in the <i>P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></i> Space Group.....	107
4.4	Comparison of 2F <sub>O</sub> -F <sub>C</sub> Electron Densities of the Fully Occupied 2-11 Base Pair with the Partially Occupied 3-10 Base Pair.....	139

4.5	Comparison of $2(F_O - F_C)$ Difference Electron Density Maps of the 3-10 Base Pairs of the GG/CC, 4mCC, and GG-MC5 Crystal Structures.....	142
4.6	Plot Comparing the Difference in Solvation Free Energies Between the OG and OC Models Relative to the Reference 4mCC Structure and Observed Difference in Lattice Interaction Energies.....	147
4.7	Comparison of the Final Structures of the Fully and Hemi-methylated Structures with Root Mean Square (RMS) Deviations Calculated for Dinucleotides of OG and OC Models.....	153

## List of Tables

1.1	Selected Helical Parameters of B-DNA and Z-DNA.....	16
1.2	Factors Which Influence the Stability of Z-DNA.....	19
2.1	Calculated Solvent-Accessible Surfaces (SAS) and Free Energies of Transfer ( $\Delta G^{\circ}H^*$ ) from Octanol to Water of Aliphatic, Aromatic, and Phosphodiester Compounds.....	32
2.2	Atomic Solvation Parameters (ASP) of Hydrophilic, Hydrophobic, and Charged Phosphate Surfaces in Nucleic Acids.....	37
2.3	Calculated Hydration and Experimental B-Z Transition Energies for DNA Dinucleotides.....	38
2.4	Differences in Hydrophobic and Hydrophilic Solvation Free Energies (Z minus B) for Base and Ribose Surfaces of DNA Dinucleotides.....	45
2.5	Calculated B-Z Transition Free Energies and Crystallization Conditions for Synthetic Hexamers as Z-DNA.....	54
3.1	Atomic Solvation Parameters (ASP) of Hydrophilic, Hydrophobic, and Charged Phosphate Surfaces in Nucleic Acids.....	68
3.2	Solvent-Accessible Surface Areas of Alternating Pyrimidine-Purine Sequences as B- and Z-DNA.....	74
3.3	Calculated Solvation Free Energies and the Corresponding B- to Z-DNA Transition Free Energies of Dinucleotide Sequences.....	78
3.4	Comparison of the Calculated B- to Z-DNA Transition Free Energies to Cationic Strength in Solutions that Yield Crystals of Hexanucleotide Sequences as Z-DNA.....	80
4.1	Nomenclature for the Crystal Structures of the Six Asymmetric Hexamer Duplexes.....	102
4.2	Crystal Packing Environments for the Methyl Groups of the Asymmetric Hexamer Duplex in the $P2_12_12_1$ Space Group.....	110
4.3	Predicted Differences in Solvation Free Energies and van der Waals Lattice Interaction Energies for the Six Asymmetric Duplexes Relative to the Reference 4mCC Structure.....	116
4.4	Protocol for the Crystal Structure Refinement of the Asymmetric Z-DNA Hexamers.....	124
4.5	The Crystallization Conditions Used to Obtain Diffraction Quality Crystals of the Six Asymmetric Hexamer Duplexes.....	132



4.6	Unit Cell Parameters and Space Groups Determined for the Six Crystals of the Asymmetric Z-DNA Hexamers.....	134
4.7	Comparison of the Completeness of the Data Sets Collected for the GG/CC, 4mCC, 4mCC-MC5, GG-M1,GG-M5, and the 4mCC-MC1 Crystals.....	135
4.8	Parameters for the Refinement of the Crystal Structures of the Six Asymmetric Hexamer Duplexes.....	136
4.9	Calculated Differences in Internal and Lattice van der Waals Energies, and Solvation Free Energies Between the OG and OC Orientations for the Crystal Structures of the Six Asymmetric Hexamer Duplexes.....	149
4.10	Twist Angles Between Base Steps Calculated for the OG and OC Models of the Six Crystal Structures.....	157

# **Studies on the Hydrophobic Effect and Its Contribution to the Stability, Crystallization, and Helix Packing of Z-DNA**

## **Chapter 1**

### **Introduction**

The basic understanding of biological processes often depends on information concerning the structure and dynamics of macromolecules and their interactions. It is generally accepted that the three dimensional structure of a macromolecule is determined not only by the covalent linkages of the macromolecule, but just as significantly by the interplay of weaker noncovalent forces. Studying these forces and their role in determining macromolecular structure is therefore valuable in understanding at a very basic level how these molecules fold, interact, and carry out their designated roles in biochemical processes. The utility of this level of understanding is evident from its applicability to various areas of biological sciences for purposes such as macromolecular structural engineering (Oxender & Fox, 1987), structure-based drug design (Johnson 1994; Whittle & Blundell, 1994) as well as aid in the development of rational methods for macromolecular crystallization (Geige et al., 1994). The study of macromolecular structure for the purpose of identifying the molecular determinants of structure and structural interactions is therefore one of general interest for many in the field of structural biology.

The native structure of a macromolecule is currently thought to result from the interplay between four basic types of noncovalent weak forces. These forces can be generally categorized as being the result of electrostatic, van der Waals, and hydrogen bonding interactions, and the hydrophobic effect. Thus, macromolecular structure

reflects the balance of forces defined by intramolecular interactions and interactions with the surrounding aqueous medium. The relative significance of these forces in determining macromolecular structure has been extensively studied experimentally and theoretically in both protein (see reviews by Dill, 1990; Schellman, 1987; Rose & Wolfenden, 1993; Matthews, 1987) and DNA (Hunter, 1993; Kollman 1988; Edelhoch & Osborne, 1976) systems. The views concerning the identity of the primary force responsible for determining protein structure has changed over the years oscillating between all four weak forces (Dill 1990; Rose & Wolfenden, 1993). Since Kauzmann's observation of the similarity of the thermodynamic characteristics of the aqueous solvation of apolar solutes and protein unfolding (Kauzmann, 1959), it has largely been accepted that the hydrophobic effect is the major force responsible for the stability of the native structure of globular proteins as well as driving the formation of macromolecular assemblies such as micelles, and membranes (Tanford, 1980). The hydrophobic effect is also thought to be at least partially responsible for the stability of nucleic acid secondary structures (Saenger, 1984; Edelhoch & Osborne, 1976; Kollman, 1988).

There are a number of observations which support the significance of the contribution of the hydrophobic effect in determining the structure of macromolecules (Dill 1990). However, there remains much discussion and controversy over the phenomenological interpretation of the hydrophobic effect (solvent effects) based on comparisons of thermodynamic data for processes such as protein unfolding and apolar hydration (Dill, 1990, Rose, 1993). This is in part due to the limitations in the understanding of the structure of liquid water, water-solute interactions (Pratt & Chandler, 1977), and the process of protein folding. Thus, it is difficult to quantitatively account for the contribution of solvent interactions to the thermodynamic stabilization of macromolecules.

### Statement of Purpose

The studies presented in this thesis focus on the hydrophobic effect and its contribution to the stability of deoxyribonucleic acid (DNA) secondary structure and DNA-DNA interactions. The significance of the hydrophobic effect and solvent interactions were examined in the well characterized B- to Z-DNA transition with computational, spectroscopic, and crystallographic studies. This macromolecular conformational transition was used as a model system based on the large amount of available data on the thermodynamics of the B- to Z-DNA transition, and the structural details of both the B- and Z-DNA conformations. In general, the studies attempt to quantitatively account for the observed sequence dependent stability of the alternative Z-conformation by considering the differences in water interactions between the B- and Z-conformations. The observed sequence specific behavior was analyzed by examining how DNA nucleotide base substituent groups affect potential solvent interactions. The contributions of two substituent groups, the methyl group at the C5 carbon of pyrimidine bases and the amino group at the C2 carbon of the purine bases are studied in comparisons of alternating pyrimidine-purine (APP) DNA sequences containing naturally occurring and modified DNA nucleotides. These two substituent groups influence either the major or minor groove characteristics of the B- and Z-conformation.

In chapter 2, the major groove methylation effects were studied in comparisons of differences in the relative stabilities of the d(TA) versus d(UA) dinucleotides (where U refers to uracil). This comparison focuses on determining the basis for the destabilization of Z-DNA by TA base pairs.

The second chapter first describes the development of a semi-empirical quantitative method which accounts for the solvent interactions with calculated solvation free energies (SFEs). An analogous method has been reported for the analysis of the solvation characteristics of "correctly" versus "incorrectly" folded proteins (Eisenberg &

MacLachlan, 1986). The comparison of SFEs calculated for DNA dinucleotides in the B- and Z-conformations correctly accounts for the previously observed relative thermodynamic stability of d(m<sup>5</sup>CG), d(CG), d(CA)•d(TG), and d(TA) dinucleotides as Z-DNA (m<sup>5</sup>C refers to cytosine methylated at the C5 position). The relative stabilities of DNA dinucleotides ( $\Delta G^{\circ}_{T(B-Z)}$ ) in the equilibrium between the B- and Z-conformations were obtained from experiments which measure the degree of negative supercoiling required to induce the B- to Z-DNA transition in well defined sequences within closed circular plasmids. The comparison of the observed stability of these dinucleotides as Z-DNA and the calculated differences in SFEs provided a relationship for predicting the relative stability of APP sequences as Z-DNA. In addition, the predicted stability of APP hexanucleotide sequences as Z-DNA are shown to be related to the ability for that sequence to crystallize as Z-DNA. This relationship provides a means for testing the predicted ability of APP sequences to adopt the Z-conformation.

The analysis of the calculated SFEs suggests that unlike the methyl group of d(m<sup>5</sup>CG) dinucleotides, the methyl group of the d(TA) dinucleotide destabilizes the Z-conformation. The contrary methylation effects are related to the differences in the methyl environments between d(m<sup>5</sup>CG) versus d(TA) dinucleotides in the Z-conformation. This led to a prediction that the removal of the methyl group from d(TA) dinucleotide (as with the d(UA) dinucleotide) would stabilize the Z-conformation. This prediction was tested and supported in crystallization experiments with the d(UA) containing sequence d(m<sup>5</sup>CGUAm<sup>5</sup>CG) and other Z-DNA hexamer crystals.

The third chapter continues the analysis of the B- to Z-DNA transition first with titration experiments of poly (dG - dC) and poly (dG - m<sup>5</sup>C) with salts of the lyotropic Hofmeister series. The salt dependence of the B- to Z-DNA transition of these sequences provides empirical evidence that the stabilization of Z-DNA attributed to the methylation of cytosine bases is associated with the hydrophobic effect. In addition, chapter 3 extends the SFE analysis of the stability of Z-DNA to include minor groove substituent

effects. The contribution of the N2 amino group of the purine bases to the stability of Z-DNA was examined by de-aminating the guanine base in d(CG) dinucleotides to form inosine bases in d(CI) dinucleotides, and conversely the amination of the adenine base in d(TA) dinucleotides to form 2-aminoadenine bases in d(TA') dinucleotides. The predicted stability of these dinucleotides as Z-DNA were tested with comparisons of salt conditions required to obtain diffraction quality crystals of Z-DNA hexanucleotides containing various alternating pyrimidine-purine dinucleotide sequences. In these crystallization experiments, the cation concentrations were considered to be the driving force for Z-DNA formation. Thus, hexanucleotide sequences which more readily adopt the Z-conformation would require less salt to crystallize as Z-DNA. With the information obtained from chapters 2 and 3, a thermodynamic cycle was proposed which relates substituent effects to the stability of dinucleotides as Z-DNA.

Chapter 4 describes a novel system for studying the balance of forces which determine specific DNA-DNA interactions. The crystallographic study presented in this chapter examines the balance of opposing van der Waals interactions and the hydrophobic effect in determining the orientation of Z-DNA hexamers in the  $P2_12_12_1$  crystal lattice. The parent structure for this latter study is the asymmetric Z-DNA hexamer  $d(m^5CGGGm^5CG) \cdot d(m^5CGCCm^5CG)$  reported by Schroth (1993). The asymmetry of the duplex, and the high resolution structures of the Z-DNA hexamer crystals allowed the identification of two packing orientations in the lattice. The distributions of the duplex between the two discrete and distinguishable orientations in these crystals were analyzed quantitatively. In these crystallographic experiments, the distribution is considered to be determined in the equilibrium between the Z-DNA hexamers in solution and the growing crystal surface. Therefore, the observed distributions provided a measure of the difference in the potential lattice interaction energies between the two packing orientations ( $\Delta G^\circ_{occ}$ ). The distribution of a hexamer duplex was monitored in response to five modifications to the hexamer duplex structure. The experimentally derived differences in

the stability of two packing orientations were compared to differences in calculated SFEs and van der Waals packing energies.

The remainder of the introduction will first present a brief introduction to the hydrophobic effect. The material presented in this introduction to the hydrophobic effect is not intended to be comprehensive. Rather, the brief treatment is intended to illustrate some of the current uncertainty in the understanding of the phenomenon as studied in protein systems. Much of the effort put forth in both structural and thermodynamic studies of the hydrophobic effect have used protein unfolding as a model. The parallels between the unfolding process and apolar hydration were recognized in the 1950s, and it seems logical that the protein unfolding process would therefore be the system of choice. However, it is clear that the protein folding process remains poorly understood. Although the native protein structure is assumed to reflect the free energy global minimum (Anfinsen, 1973; Dill, 1990), relatively little is known about the structure and therefore thermodynamic stability of the unfolded state. It is likely that the incomplete information on the unfolded state in combination with the complexity of water structure and water-solute interactions have contributed to the uncertainty in the current interpretations of the hydrophobic effect.

The studies presented in this thesis examine the significance of solvent interactions in a well characterized macromolecular conformational transition, and in a highly utilized crystallographic system. Therefore, the final section of the general introduction is intended to present background information describing the general features of the B- to Z-DNA transition. The B- to Z-DNA transition was used as a model system in these studies based on the large amount of data available on the thermodynamic and structural factors which affect the stability of Z-DNA. In addition, the B- to Z-DNA transition is considered to be entropy driven, and therefore likely to be dependent on DNA-solvent interactions. These factors make the B- to Z-DNA transition an excellent system for studying the balance of thermodynamic forces in determining macromolecular secondary

structure. A review of the general structural characteristics of Z-DNA is presented first followed by a brief overview of the structural and environmental factors which stabilize the Z-conformation. This section of the general introduction concludes with the presentation of computational studies which attempt to explain the sequence dependent stability of Z-DNA.



### Current Views on the Hydrophobic Effect.

Since Kauzmann's observation of the similarity in the thermodynamic characteristics of apolar solutes in aqueous solutions and protein unfolding (Kauzmann, 1959), it has largely been accepted that the hydrophobic effect is the major force responsible for the stability of the native structure of globular proteins. The hydrophobic effect has been described as the positive Gibbs free energy associated with the solvation of model apolar compounds in aqueous solutions. The aqueous solvation of model apolar molecules at room temperature is associated with a large increase in heat capacity of the system, and a negative entropy change (Kauzmann, 1959).

A commonly accepted qualitative rationalization of the thermodynamic observations on the hydrophobic effect is the formation of ice-like clathrate water structures around apolar solutes (Klotz, 1960). These ordered water structures are thought to form in response to the cohesive properties of liquid water as well as the lack of stabilizing interactions between the water and the apolar solute (Muller, 1990). The ordering of water in apolar solvation would be consistent with the observed decrease in entropy of the system. The additional heat required to 'melt' these ice-like structures would also explain the increased heat capacity increment in apolar solvation (Dill, 1990).

Similarities in the thermodynamic properties of aqueous solvation of model apolar compounds and the unfolding of proteins suggested the importance of apolar solvation in protein folding. Both processes are associated with a large characteristic increase in the heat capacity (Murphy et al., 1990). The unfavorable Gibbs free energy associated with aqueous solvation of apolar groups is thought to be the basis for the burying of protein hydrophobic groups in the interior of the protein thereby directing the folding process. The sequestering of these groups in the interior of the protein is expected to increase the randomness of the system by the release of ordered water molecules associated with the burying of hydrophobic groups. The burying of hydrophobic groups is therefore

considered to be an overall entropically favorable process. This is consistent with the finding that the hydrophobic residues are generally found buried in the interior of the protein, minimizing their solvent exposure (Perutz et. al., 1965.). Additional supporting evidence for the significance of the hydrophobic effect in the protein folding process have been cited in a review by Dill (1990).

Although widely accepted as the principal force driving the folding of the native protein structure, there remains much uncertainty and debate concerning the molecular basis for the hydrophobic effect. The uncertainties are reflected in reports which raise questions regarding the basic indices of the involvement of the hydrophobic effect in the protein folding process (Makhatadze & Privalov, 1993; Dill, 1990; Woolfson et al., 1993). In particular these questions center on the molecular basis for the increase in the heat capacity (Privalov & Makhatadze, 1992) and the significance of the water ordering in determining protein folding (Woolfson et al., 1993). The interpretation of these analyses are likely to be complicated by the sparse information on the stability of the unfolded state (Woolfson et al., 1993). Therefore, there is currently discussion concerning the validity of the commonly accepted comparisons of apolar hydration and protein unfolding.

The thermodynamic contribution of water-solute interactions to the stability of macromolecules has been estimated using semi-empirical methods. The estimation of solvent interactions based on the solvent exposure of atoms was first reported by Langmuir (Langmuir, 1925). Eisenberg and MacLachlan adopted this method in an analysis of the stability of proteins with calculated solvent accessible surface areas to estimate solvent interactions (Eisenberg and MacLachlan, 1986). The solvent accessible surfaces were weighted by thermodynamic parameters derived from experimentally determined free energies of transfer. These atomic solvation parameters (ASPs) estimate the free energies associated with solute-water interactions for various types of protein surfaces. Using this method to estimate solvation effects, they were able to correctly identify "improperly" versus "properly" folded proteins. However, this method for

examining the contribution of the hydrophobic effect to the stability of proteins was limited to a qualitative analysis due to the lack of structural information on the unfolded state.

In summary, there is evidence that the hydrophobic effect is significant in the stabilization of the folded structures of macromolecules. However, the hydrophobic effect is not well understood, and therefore remains difficult to quantitate based on the explicit descriptions of water-solute interactions. The lack of information on the unfolded state of proteins complicates both theoretical as well as experimental interpretations of the process.

### Introduction to the B- to Z-DNA Conformational Transition

The primary and secondary structure of deoxyribonucleic acids (DNA). Polymers of deoxyribonucleic acids (DNA) can be thought of as being composed of three chemical components. Each monomeric unit of the DNA polymer consists of a negatively charged phosphate group, a 2' deoxyribose sugar, and an aromatic base unit. The backbone of the DNA polymer is composed of phosphodiester linkages of the 5' and 3' hydroxyl groups of the 2' deoxyribose sugars of adjacent monomeric units. The base units are attached to the backbone by a glycosidic bond between the C1' carbon of the furanose rings of ribose sugar with the N9 of purine bases or N1 of pyrimidine bases. The four naturally occurring base types in the DNA polymer are 2 purine bases (adenine and guanine), and 2 pyrimidine bases (cytosine and thymine).

DNA exists most typically as a double-stranded duplex of two anti-parallel nucleotide strands. Early evidence for the double-stranded nature of DNA was obtained from the analyses by Chargaff (Zamenhof et al., 1952). The analysis of DNA from a number of sources showed that DNA is composed of nearly equivalent ratios of adenine to thymine residues and cytosine to guanine residues. It was concluded from these observations that regular complexes are formed between the adenine and thymine bases, and the cytosine and guanine bases within DNA. This finding in conjunction with the molecular modeling of base pairing interactions by Watson and Crick, and the fiber diffraction pattern of DNA ( Franklin & Gosling, 1953a and 1953b ) led to the initial modeling of DNA as a double-stranded helical structure (Watson & Crick, 1953).

The stability of DNA helix is due mainly to hydrogen bonding and base stacking interactions. The stability of the DNA double-helix is attributed to several structural characteristics. The location of the charged phosphate groups at the exterior of the helix allows interaction with positive cations thereby minimizing the charge repulsion. The specific hydrogen bonding interactions between the two complementary polynucleotide

strands stabilizes the double-stranded DNA complex (Edelhoch H. & Osborne 1976; Saenger, 1984). In addition, the placement of the bases at the center of the helix allows stacking interactions within the core of the duplex structure which contribute favorable van der Waals interactions as well as minimize solvent exposure of the hydrophobic base groups (Saenger, 1984). Thus, the DNA double helix structure appears to be a compromise between the geometric requirements of the interstrand hydrogen bonding interactions, the optimization of base stacking interactions, and the minimization of the charge repulsion of the phosphate backbone.

Conformational variabilities within these constraints are possible with the flexibility of the phosphoribose backbone, and variations in the base orientations. Six rotation angles in the backbone, four torsion angles in the furanose ring, and the relative orientation of the base group to the furanose ring contribute to the polymorphic nature of DNA structure. A number of DNA conformations have been observed. These conformations exhibit a variety of helical characteristics with general differences in their helical repeat (8-12bp/turn), helix pitch (28Å to 45Å), and handedness (right or left) (reviewed in Saenger, 1984; Dickerson et al., 1992).

A striking example of the intrinsic conformational flexibility of DNA is demonstrated in the conversion between the right handed B-DNA to the left handed Z-DNA conformation (for review see Jovin 1987; Rich 1984). Evidence for this DNA conformational transition was first observed by Pohl and Jovin in optical rotatory dispersion and circular dichroism studies on poly (dG-dC) (Pohl & Jovin, 1972). The conformational transition was reported to be a highly cooperative, entropy driven reversible process (Pohl & Jovin, 1972). The structural details of the conformation were reported in the first single crystal structure of Z-DNA by Wang et al. (Wang et al., 1979). Since these early observations, further studies have characterized the thermodynamic and structural requirements which stabilize Z-DNA, and have revealed that the stability of the Z-conformation is sequence dependent (Rich, 1984 ; Jovin 1972).

Crystallographic studies of Z-DNA hexamers in the  $P2_12_12_1$  space group. The left-handed Z-DNA conformation is structurally well characterized. This is due to the high resolution structures available from single crystal X-ray diffraction studies on Z-DNA hexamers. There have been a large number of Z-DNA hexamer crystal structures reported which were designed to study the sequence dependent stability of the Z-conformation (Wang, 1985; Schroth et al., 1993; Zhou & Ho, 1990; Fujii et al., 1982; Wang et al., 1984; Chevrier et al., 1986; Coll et al., 1986; Schneider et al., 1992), base pair mismatches (Coll et al., 1989; Ho et al., 1985), and metal and polyamine binding (Gao et al., 1993; Geierstanger et al., 1991; Kagawa et al., 1991; Ho et al., 1987; Gessner et al., 1989; Egli et al., 1991; Gessner et al., 1985). These Z-DNA hexamers crystallized in the  $P2_12_12_1$  space group and typically diffracted to near atomic resolution (1.0 Å to 1.5 Å). The high quality X-ray diffraction data available from these crystals has provided detailed information on both Z-DNA structure and DNA-solvent interactions.

Characteristics of the B- and Z- conformations. Figure 1.1 shows a comparison of the B-DNA and Z-DNA conformations with space filling models of a full turn of each conformation. The DNA duplex in both conformations is comprised of two anti-parallel nucleotide strands held together by standard Watson-Crick type base pairing. The most obvious difference between the B- and Z-conformations is the helical twist. The B-conformation is right handed while the Z-conformation is left-handed. Table 1.1 shows the general helical characteristics of the B- and Z-conformations. The longer helical pitch and shorter helix diameter characterizes a longer slimmer helix for the Z-conformation (Rich et al., 1984).

Figure 1.1 illustrates the general differences in the backbone and groove characteristics between B- and Z-DNA. The characteristic zig-zag pattern of the phosphates for each of the nucleotide strands are apparent in the Z-DNA backbone. The B-conformation in comparison has a smooth, regular backbone structure. The closest approach of the phosphate groups in the Z-conformation (7.7 Å) is approximately 4 Å

shorter than in the B-conformation (11.7Å) (Rich et al., 1984). The relative instability of the Z-conformation is in part due to the greater electrostatic repulsion between the phosphate groups resulting from the closer approach of the phosphates across the grooves (Rich et al., 1984).

In comparison to the B-conformation, the Z-conformation has a deep and narrow minor groove and a major groove which is no longer actually a groove, but rather a highly exposed convex surface. The groove characteristics of the B- and Z-conformations reflect the difference in the displacement of the base pairs relative to the helical axis between the two DNA conformations. In the Z-conformation, the base pairs are displaced away from the axis toward the major groove surface as opposed to the B-conformation in which the helical axis runs between the bases. The contrasting groove characteristics result in differences in the base stacking interactions and solvent interactions for the B- and Z-conformations. Thus, the relative stability of the B- and Z-DNA conformations are related to their contrasting groove and backbone structure.

Z-DNA is characterized by alternating pattern of helical parameters defining a dinucleotide repeat. The remaining parameters on table 1.1 illustrate the alternating pattern of the helical parameters for nucleotides in the Z-conformation which define a dinucleotide repeat for Z-DNA versus a single base pair repeat for B-DNA. Alternating pyrimidine-purine (APP) are able to conform to the steric restrictions imposed by the alternating pattern of *anti-syn* conformation of the glycosidic bond which characterize the Z-conformation.

In general, DNA bases are able to adopt two orientations relative to the furanose ring which are related by rotation about the glycosidic bond. In Z-DNA, the glycosidic conformation is found to alternate between the two conformations with pyrimidines adopting the *anti* conformation and purines adopting the *syn* conformation. All nucleotides in the B-conformation are in the *anti* glycosidic conformation. The stability

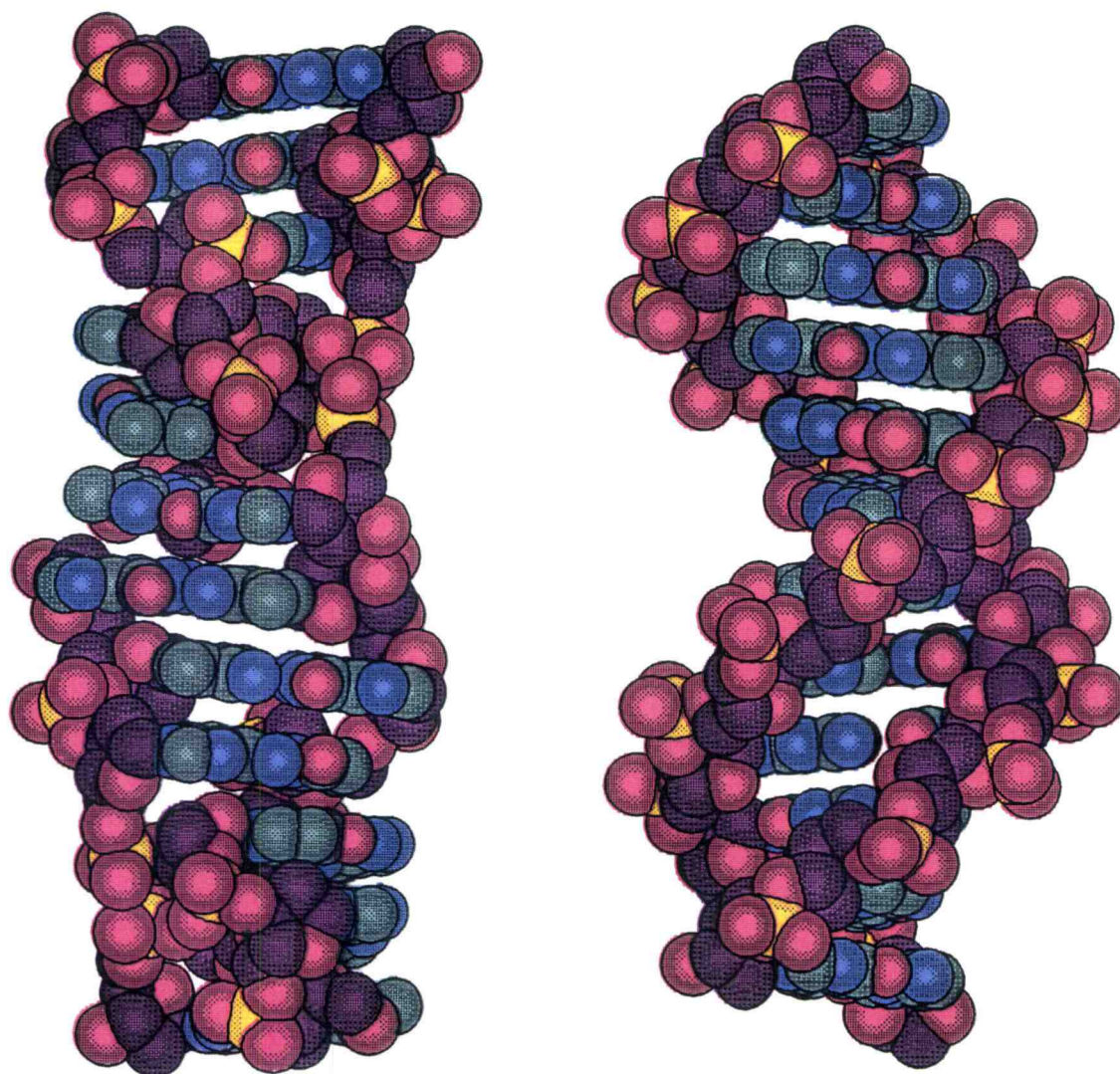


Figure 1.1. Space filling models of the B- and Z-DNA conformations. The right-handed B-conformation is shown on the right, and the left-handed Z-conformation is shown on the left. Both structures are 18 base pairs in length. Base carbon atoms are colored green, ribose carbon atoms are colored purple, oxygen atoms are colored red, nitrogen atoms are colored blue, and phosphorus atoms are colored yellow.



Table 1.1. Selected helical parameters of B- and Z-DNA.

	B-DNA <sup>1</sup>	Z-DNA <sup>1</sup>
Pitch (Å / helical turn)	34	45
Helix diameter (Å)	20	18
Hydrogen bonding		
Helical angle(°)	Watson - Crick 36	Watson - Crick -30
d(CpG)		-9
d(GpC)		-51
Helical repeat (base pairs / turn)	10.5	12
Rise (Å / residue)	3.2	3.8
Glycosidic conformation / sugar pucker		
pyrimidines	<i>anti</i> / C2' - <i>endo</i>	<i>anti</i> / C2' - <i>endo</i>
purines	<i>anti</i> / C2' - <i>endo</i>	<i>syn</i> / C3' - <i>endo</i>

## 1. Parameters from Rich et al., 1984

of purine bases is approximately equivalent in either the *anti* and the *syn* conformations. In contrast, the pyrimidine bases are sterically inhibited from adopting the *syn* conformation (Saenger, 1984). This is due to a steric collision between the C2 exocyclic group of the pyrimidine ring and the C3' hydrogen of the furanose ring (Haschemeyer & Rich, 1967 ). Thus, Z-DNA is most readily formed with alternating pyrimidine-purine (APP) sequences. Exceptions to this pattern have been reported in crystal structures of Z-DNA hexamers showing that out-of-alternation thymine and cytosine residues can be accommodated in the *syn* conformation in Z-DNA. However, the steric penalties associated with the pyrimidine base in the *syn* conformation are apparent as increased buckling of the base pair (Wang et al., 1985; Schroth et al.,1993).

The alternating features of the helical parameters in Z-DNA are also apparent in the twist angles and the sugar puckers. The twist angle is found to alternate between a small ( $-9^\circ$ ) twist angles for the d(CpG) steps and a larger ( $-51^\circ$ ) twist angle for the d(GpC) steps. The sugar pucker of pyrimidine and purine nucleotides are also found to alternate between C2' *endo* and C3' *endo* respectively.

The stability of DNA dinucleotides as Z-DNA is dependent on the base substituent groups. The B- to Z-DNA transition has been monitored with a variety of physical methods. These include circular dichroism, gel electrophoresis, ultraviolet absorption, and nuclear magnetic resonance spectroscopy (Rich et al., 1984). These methods have been used to determine the sequence specific stability of DNA dinucleotides as Z-DNA, and the factors which stabilize the Z-conformation.

A few examples of the factors which influence the stability of the Z-conformation are listed on table 1.2. Covalent modifications of the DNA substituent groups have been shown to stabilize the Z-conformation (reviewed in Rich et al., 1984; Jovin et al., 1987). These modifications alter the stability of the Z-conformation by affecting steric, electrostatic, and solvation characteristics of Z-DNA. Iodination, bromination, and methylation of the C5 carbon of cytosine bases have been shown to stabilize the Z-conformation. The bulky substituent groups are found to be in relatively inaccessible environments in the major groove of Z-DNA as opposed to highly exposed positions in the B-conformation. Therefore, the stabilizing effect of these groups have been suggested to reflect the influence of the hydrophobic effect (Fujii et al., 1982; Chevrier et al., 1986). Bromination at the C8 of the purine base stabilizes the Z-conformation by sterically restricting the purine base to the *syn* conformation (Rich et al., 1984). Methylation or platinum binding at the N7 of the guanine bases stabilizes the Z-conformation (Rich et al., 1984). These modifications are thought to introduce a positive charge to the purine ring system and therefore the stabilizing effect is considered to be

electrostatic. The destabilization of the Z-conformation by aflatoxin binding is possibly due to favorable binding to the B-conformation (Rich et al., 1984).

Table 1.2 reports various environmental conditions which stabilize Z-DNA. The Z-conformation is stabilized by aqueous solutions of alcohols, metal ions, and salts. Cations significantly affect the stability of the Z-DNA and vary in their ability to stabilize the Z-conformation. The mechanisms by which cations stabilize Z-DNA are by reducing the charge repulsion between the phosphate groups in the backbone, or through specific interactions with the DNA bases. The monovalent cations in general stabilize the Z-conformation by minimizing the phosphate repulsion (Rich 1984). The higher valent transition metal complexes such as  $[\text{Ru}(\text{NH}_3)_3]^{3+}$  and  $[\text{Co}(\text{NH}_3)_3]^{3+}$  form specific complexes with the N7 of the guanine bases in the major groove of Z-DNA (Ho et al., 1986; Gessner et al., 1985).

Specific water interactions have been shown to be different for CG and AT base pairs in the B versus Z conformations (Schneider et al., 1993). The significant differences between water interactions of the B- and Z-conformations occur in the minor groove. In the B-conformation, a spine of hydration is observed in the minor groove involving the N3 of the adenine base and the O2 of the thymine bases. This pattern is not observed in CG base pairs, and is thought to be disrupted by the N2 amino group of the guanine bases (Saenger; Drew & Dickerson, 1981). In the Z-conformation, strong water interactions in the minor groove occur at the O2 of the cytosine bases. In addition, a backbone phosphate oxygen stabilizes guanine in the *syn* conformation (Jovin et al., 1987). Ordered water molecules were not located in the minor groove of d(TA) base bridging water interaction involving the N4 amino group of the guanine base and a pairs in the Z-DNA crystal structures of d(m<sup>5</sup>CGTAm<sup>5</sup>CG) or d(CACGTG) (Wang et al., 1984; Coll et al., 1988). The differences in the adenine and thymine water interactions

Table 1.2. Factors which influence the stability of Z-DNA.

Factor
I. Covalent modifications (relative to d(CG))
A. Modification at C5 carbon of cytosine.
$I > Br > CH_3$
B. Modifications to C8 of guanine.
$Br > H$
C. Modifications to N7 of guanine.
$CH_3, \text{Platinum complexes} > \text{Aflatoxin(B-stab)}$
.....
II. Environmental
A. Ionic conditions.
1. Cations - valence dependent
$[Co(NH_3)_6]^{3+}, \text{polyamines} > Mg^{2+} > Na^{1+}, K^{1+}, Li^{1+}$
2. Anions -
$ClO_4^{-1}, \text{Acetate}$
B. Solvent.
1. Alcohols -
methanol, ethanol, ethylene glycol
2. Water interactions.
C. Negative supercoiling.
.....
III. Sequence.
A. Dinucleotide stability as Z-DNA.
$d(m^5CG) > d(CG) > d(CA) \cdot d(TG) > d(TA)$

between the two conformations may account for at least part of the noted instability of the d(TA) dinucleotide in the Z-conformation. The relationship between these differences in water interactions and DNA conformational stability is a subject of current interest (Schneider et al., 1993).

Torsional stress has also been shown to stabilize the Z-form. The torsional stress in negatively supercoiled DNA can be relaxed by the unwinding of the two strands in the duplex relative to each other. Thus, the underwound Z-DNA form is stabilized by the free energy available from negative supercoils in DNA. Negatively supercoiled DNA has been used to induce the B to Z transition in gel migration experiments (Singleton, 1987). Estimates of the free energy associated with the formation of negative supercoils (Depew & Wang, 1975), have allowed the measurement of the free energies for the transition of various sequences. Using this methodology, the order of intrinsic stability of DNA dinucleotides as Z-DNA was observed to follow the trend  $d(CG) > d(CA) \cdot d(TG) > d(TA)$  (references in Rich et al., 1984).

Theoretical studies on the relative stability B- and Z-DNA. A number of theoretical approaches have been used in attempt to explain the observed sequence specific stability of Z-DNA. Theoretical analyses of the basis for the relative stabilities of B and Z conformations have indicated that factors which stabilize the Z-DNA relative to B-DNA include improved base stacking interactions, and favorable van der Waals interactions between the base and ribose units (Kollman 1982). As expected, the Z-conformation was destabilized due to the increased electrostatic repulsion between the phosphate groups.

In addition to these general comparisons of the stability of the two DNA conformations, Kollman et al. (1982) attempted to explain the relative stabilities of d(m<sup>5</sup>CG) dinucleotides and d(TA) dinucleotides. The energy minimized structures of the B- and Z-conformations of these sequences correctly predicted the relatively greater stability of d(m<sup>5</sup>CG) versus the d(CG) dinucleotides as Z-DNA. The greater stability of the d(m<sup>5</sup>CG) dinucleotide was attributed to the destabilization of the B-conformation due

to less favorable stacking interactions of the pyrimidine. However, the calculations were not able to correctly predict the relative instability of the d(TA) dinucleotide. The energy minimization of the structures and the calculated energies included contributions from either explicit or implicit counter-ions, but did not consider solvent interactions.

Dang et. al (1990) compared the relative stabilities of CG and TA base pairs as Z-DNA using a free energy perturbation studies in which a CG base pair is mutated to a TA base pair, creating a single CA-TG dinucleotide. These molecular mechanics studies considered the enthalpic effects of solvation with the inclusion of either implicit or explicit solvent waters. They were able to quantitatively predict the relative stability of the d(CG) and d(CA) • d(TG) dinucleotides as Z-DNA. The explicit water interactions were modeled in the free energy simulations by placing the DNA structures in a box of water and optimizing the interactions, while the implicit water interactions were modeled by adjusting the dielectric constant of the surrounding medium to mimic an environment of hydrated sodium ions. In both cases, the calculations showed that the introduced mutations destabilized the B and Z conformations to differing extents. The results correctly predicted the relative instability of d(CA) • d(TG) dinucleotides in the Z-conformation as compared to d(CG) dinucleotides by approximately 3 kcal/(mol TA bp) with explicit water interactions and 2 kcal/(mol TA bp) with implicit water interactions. However, similar calculations predicted that the d(TA) dinucleotide is as stable as d(CA) • d(TG) dinucleotides, again inconsistent with the observed solution data. These calculations indicate that internal, and explicit solvent interaction energies can not explain the sequence specific stability of dinucleotides as Z-DNA.

Further analysis of the instability of d(TA) dinucleotides as Z-DNA have been reported by Singh et al. (1993). These calculations analyzed the interactions responsible for the difference in stability of the TA versus CG base pair as Z-DNA. One of the major contributions to the instability of the TA base pair as Z-DNA was reported to be a difference in the intra-base pair energies of d(CG) versus d(TA) base pairs. It was

suggested that the relative instability of adenosine base in the *syn* conformation relative guanosine was in large part responsible for the higher intra-base pair energy..

Ho et al. (1988) were able to explain the effect of methylation on the stability of d(CG) dinucleotides in the Z-conformation by comparing the differences in solvent interactions between d(CG) and d(m<sup>5</sup>CG) dinucleotides (Ho et al., 1988). The solvent interactions were modeled with calculations of solvation free energies using solvent-accessible-surface area calculations analogous to the method used in studies of protein folding (Eisenberg & MacLachlan, 1986). The greater stability of d(m<sup>5</sup>CG) dinucleotides was attributed to the differences in the solvent exposure of the methyl group in the B versus Z conformation. In the B-conformation, the methyl group is highly exposed in the major groove. The exposure of the methyl group hydrophobic surface area in the B-conformation is on average 28 Å<sup>2</sup> per methyl group. In the Z-conformation, the methyl group is less solvent exposed, packed in a tight pocket created by the adjacent guanine residues. The methyl group in the Z- conformation covers both hydrophobic and hydrophilic base groups in the pocket resulting in a net increase in hydrophobic base surface area. However, the increased exposure of hydrophobic base surface is compensated for with the burying of hydrophobic ribose surface area. Therefore, the introduced methyl group only slightly decreases the stability of the Z-conformation due to the burying of hydrophilic groups. Thus, the greater propensity of d(m<sup>5</sup>CG) dinucleotides to adopt the Z-DNA conformation over the unmethylated d(CG) dinucleotides can be explained in a difference in solvation free energies between the two conformations (Ho et al., 1988).

In summary, the structural and thermodynamic basis for the stability of the Z-conformation are well characterized. The transition between the B- and Z-conformations is considered to be entropy driven, and therefore is expected to be influenced by interactions with the aqueous environment. The experimentally determined sequence dependent stability of dinucleotides of DNA in the Z-conformation follows the order

$d(CG) > d(CA) \cdot d(TG) > d(TA)$ . Theoretical analyses have not been effective in explaining the observed sequence specific stability of DNA dinucleotides in the Z-conformation even when considering the enthalpic contributions of solvent interactions. Chapters 2 and 3 of this thesis examine the sequence dependent stability of the Z-conformation with a semi-empirical treatment of the solvent interactions.

The well characterized structural and environmental dependence of the stability of Z-DNA has allowed a more systematic approach to the crystallization of Z-DNA hexamers (Ho et al., 1991). Extension toward a more rational approach for the crystallization of DNA oligomers would be aided by the examination of the relevant forces involved in macromolecular crystallization processes. Toward this goal, the final chapter of this thesis examines helix packing interactions in the well characterized Z-DNA hexamer *P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>* crystallographic system



## Chapter 2

### **Quantitative Analysis of DNA Secondary Structure from Solvent-Accessible Surfaces: The B- to Z-DNA Transition as a Model**

**Todd F. Kagawa, Donald Stoddard, Guangwen Zhou, and Pui S. Ho**

### Abstract

Solvent structure and its interactions have been suggested to play a critical role in defining the conformation of polynucleotides and other macromolecules. In this work, we attempt to quantitate solvent effects on the well-studied conformational transition between the right-handed B- and left-handed Z-DNA. The solvent-accessible surfaces of the hexamer sequences  $d(m^5CG)_3$ ,  $d(CG)_3$ ,  $d(CA)_3$ , and  $d(TA)_3$  were calculated in their B- and Z-DNA conformations. The difference in hydration free energies between the Z and the B-conformations ( $\Delta\Delta G^\circ H(Z-B)$  or solvent free energy (SFE)) was determined from these surfaces to be  $-0.494 \text{ kcal mol}^{-1}$  for C-5 methylated  $d(CG)$ ,  $0.228 \text{ kcal mol}^{-1}$  for unmethylated  $d(CG)$ ,  $0.756 \text{ kcal mol}^{-1}$  for  $d(CA) \cdot d(TG)$ , and  $0.896 \text{ kcal mol}^{-1}$  for  $d(TA)$  dinucleotides. These  $\Delta\Delta G^\circ H(Z-B)$  values were compared to the experimental B- to Z-DNA transition energies of  $-0.56 \text{ kcal mol}^{-1}$  that we measured for C-5 methylated  $d(CG)$ ,  $0.69\text{-}1.3 \text{ kcal mol}^{-1}$  for unmethylated  $d(CG)$ ,  $1.32\text{-}1.48 \text{ kcal mol}^{-1}$  reported for  $d(CA) \cdot d(TG)$ , and  $2.3\text{-}2.4 \text{ kcal mol}^{-1}$  for  $d(TA)$  dinucleotides. From this comparison, we found that the calculated  $\Delta\Delta G^\circ H(Z-B)$  of these dinucleotides could account for the previous observation that the dinucleotides were ordered as  $d(m^5CG) > d(CG) > d(CA) \cdot d(TG) > d(TA)$  in stability as Z-DNA. Furthermore, we predicted that one of the primary reasons for the inability of  $d(TA)$  sequences to form Z-DNA results from a decrease in exposed hydrophilic surfaces of adjacent base pairs due to the C-5 methyl group of thymine; thus,  $d(UA)$  dinucleotides should be more stable as Z-DNA than the analogous  $d(TA)$  dinucleotides. This prediction was tested and confirmed by the finding that the hexamer sequence  $d(m^5CGUAm^5CG)$  crystallized as Z-DNA in 2-fold lower  $MgCl_2$  concentrations than the analogous  $d(m^5CGTAm^5CG)$  sequence.

## Introduction

Since the first spectroscopic evidence for its existence (Jovin & Pohl, 1972) and the subsequent determination of its structure from single-crystal X-ray diffraction (Wang et al., 1979), left-handed Z-DNA has been an interesting and useful structure for physical studies on macromolecular stability. Solution and high-resolution single-crystal studies of Z-DNA sequences have shown that a variety of environmental conditions stabilize DNA in the Z-conformation, including the presence of high salt concentrations, transition-metal complexes, base alkylation, and negative supercoiling [reviewed by Rich et al., (1984) and Jovin et al., (1987)]. Underlying these effects is an intrinsic sequence dependence for the stability of Z-DNA.

Z-DNA is characterized not only by the left-handed twist of the double helix but also by an alternating *anti-syn* configuration of its base pairs. Since purine bases are sterically able to adopt the *syn* conformation more readily than are pyrimidines, sequences that are alternating pyrimidine and purine have been reported to be more stable as Z-DNA than sequences that do not follow this alternation rule (Wang et al., 1981; Drew & Dickerson, 1981; Jovin et al., 1983). This alternation in the configuration of the base pairs also requires that the basic repeat unit of the Z-DNA be the dinucleotide, or two base pairs, rather than single base pairs as is found in the known right-handed conformations (Jovin et al., 1987). Of the sequences that follow this alternation rule, there exists a hierarchy of pyrimidine-purine dinucleotides that adopt the left-handed conformation, with  $d(CG) > d(CA) \cdot d(TG) > d(TA)$  in stability as Z-DNA (Jovin et al., 1983). This difference in stability of dinucleotides can be quantitatively related to experimental B- to Z-DNA transition energies. The transition energies for various types of dinucleotides have been calculated by determining the degree of negative superhelicity required to induce Z-DNA within well-defined sequences inserted in closed circular plasmid DNAs. The B- to Z-DNA transition energies for  $d(CG)$ ,  $d(CA) \cdot d(TG)$ , and  $d(TA)$  dinucleotides

have been measured at 0.66-1.2 kcal mol<sup>-1</sup> (Peck & Wang, 1983; Nordheim et al., 1982; Frank-Kamenetskii & Vologodskii, 1984; Vologodskii & Frank-Kamenetskii, 1984; Mirkin et al., 1987), 1.32-1.48 kcal mol<sup>-1</sup> (Vologodskii & Frank-Kamenetskii, 1984; Mirkin et al., 1987), and 2.3-2.4 kcal mol<sup>-1</sup> (Ellison et al., 1986; Mirkin et al., 1987), respectively.

The availability of this type of experimentally derived structural and thermodynamic information on the stabilities of various sequences as B- versus Z-DNA makes this transition an excellent system for studying the contribution of various thermodynamic forces on macromolecular conformations. A previous calculation of the differences in intramolecular energy for DNA sequences in the B- and Z-conformations using the energy minimization program AMBER could account for the stabilizing effect of methylating cytosines of d(CG) dinucleotides in enthalpic terms but was unable to account for the relative inability of d(TA) dinucleotides to form Z-DNA (Kollman et al., 1982). Similarly, Karplus and co-workers (Tidor et al., 1983; Irikura et al., 1985), using related calculations, were unable to adequately model the salt-induced B-Z transition (Jovin et al. 1987). The shortcomings of these methods fall primarily in their inability to incorporate interactions of DNA with water networks, bulk water, and ions (Jovin et al., 1987). Thus, the stability of Z-DNA and DNA conformations in general may be better understood by considering differences in solvent interactions with their structures.

The contribution of solvent interactions to the stability of alternative DNA conformations has been previously discussed in qualitative terms. It was shown that the stability of right-handed A-DNA under dehydrating conditions can be explained in terms of the difference in the solvent-accessible surfaces of this conformation relative to B-DNA (Alden & Kim, 1979). Saenger et al. (1986) have attributed the propensity of a sequence to adopt the B versus alternative helical forms to the concept of "economy of hydration". They suggest that the more economical hydration of the phosphates in A-DNA and Z-

DNA is responsible for their enhanced stabilities under dehydrating conditions. The question that we address here is whether the effect of bulk water structure and interactions on DNA conformation can be described in quantitative terms that account for the actual differences in abilities of sequences to adopt non-B conformations. A useful semi-empirical approach to estimating the contribution of solvent interactions on molecular structures, first developed by Langmuir (1925), has been to calculate the overall hydration energy of the molecule and of its component parts. This approach was extended by MacLachlan and Eisenberg (1986) to estimate the effect of solvent interactions on "properly" versus "improperly" folded proteins. The lack of experimental data on the stabilities of the various conformational states being considered, however, limited these previous studies to only qualitative evaluations of how the calculated solvation energies contribute to the thermodynamics of macromolecular stability. Recently, we have calculated the difference in hydration energies of methylated and unmethylated d(CG) dinucleotides in the B and Z conformations to study the Z-DNA stabilizing effect due to methylation of the cytosine base (Ho et al., 1988). The stabilization due to methylation was found to result from an overall increase in the hydrophobic character of B-DNA, while the exposed hydrophobic and hydrophilic surfaces of the dinucleotide were essentially identical for both the methylated and unmethylated bases in the Z form.

In this study, we set forth to quantitate the contribution of solvent interactions on the stability of various sequences as Z-DNA by calculating the hydration energies of DNA sequences in the B and Z conformations from their solvent-accessible surfaces (SAS). To circumvent any problems that may result from steric inhibitions to Z-DNA formation, we limited our studies to alternating pyrimidine-purine sequences. The differences in the hydration energies between the Z and B conformations of d(m<sup>5</sup>CG), d(CG), d(CA)•d(TG), and d(TA) were compared to the experimental transition energies for the dinucleotides. From this, we examined the molecular basis for the inability of d(TA) to

form Z-DNA in terms of hydrated surfaces. A comparison of d(TA) with d(UA) dinucleotides as Z-DNA showed that the C-5 methyl group of the thymine base contributes significantly to the inability of d(TA) dinucleoties to adopt the left-handed conformation. We therefore predicted that d(UA) dinucleotides should be more stable in the Z conformation than d(TA). This prediction was tested by comparing the conditions for obtaining single crystals of a hexamer sequence that contains d(UA) base pairs as Z-DNA to published conditions for analogous hexamer sequences that contain d(m<sup>5</sup>CG), d(TA), d(CA)•d(TG), and d(CG) dinucleotides.

### Experimental Methods

The atomic coordinates for each atom, including hydrogens, of the sequences d(CGCGCG), d(CACACA)•d(TGTGTG), d(UAUUAUA), and d(TATATA) were generated in their right-handed B conformation by using generalized helical parameters (Arnott, 1976) and standard distances and geometries. The coordinates of the sequence d(CGCGCG) as Z-DNA were taken from its single-crystal magnesium structure (Gessner et al., 1985). The structures of d(m<sup>5</sup>CGm<sup>5</sup>CGm<sup>5</sup>CG) in the B and the Z conformations were obtained as previously described (Ho et al., 1988). The structures of d(CACACA)•d(TGTGTG) and d(TATATA) were generated by combining dinucleotides from the single crystal structures of d(CGCGCG) (Gessner et al., 1985) and d(m<sup>5</sup>CGTAm<sup>5</sup>CG) (Wang et al., 1984) and optimizing the bond length, dihedral and torsional angles, and van der Waal's contacts. The solvent-accessible surfaces (SAS) of the DNA models were calculated by using a rolling ball method (Connolly, 1983) as previously described (Ho et al., 1988). The surfaces of each atom in the molecules were assigned to different chemical groups (as either carbon, nitrogen, phosphorus, or oxygen and their associated hydrogen surfaces at the aromatic bases, the ribose sugars, or the charged phosphate group).

A set of atomic solvation parameters (ASP) (Eisenberg & MacLachlan, 1986) associated with solvating each surface type of DNA molecules was derived from the partition coefficients of a set of small organic molecules (Table 2.1) that represent the aromatic base, ribose sugar, and phosphate components of nucleic acids. The atomic coordinates of these molecules were generated by standard distances and geometries and their solvent-exposed surfaces calculated by the rolling ball method (Connolly, 1983). The calculated surfaces were related to their free energies of hydration ( $\Delta G^{\circ}_H$ ) as derived from experimentally determined partition coefficients of the molecules in aqueous solution and octanol (Hansch & Leo, 1979). From the slope of these relationships, the hydration

energy per exposed area (ASP values) was calculated for each atom type for the base, sugar, and phosphate groups of nucleic acids (Table 2.1).

The hydration free energies ( $\Delta G^{\circ}_H$ ) of the dinucleotides were calculated by applying the ASP values to the calculated areas of each surface type ( $SAS_i$ , where  $i$  represents a surface type listed in Table I) of the dinucleotides in the B and Z conformations and summing the resultant energies (Eisenberg & MacLachlan, 1986).

$$\Delta G^{\circ}_H = \sum_i (SAS_i \times ASP_i) \quad \text{Eq. 1}$$

The surfaces of the two terminal base pairs of each hexamer sequence were excluded from the calculations to avoid errors due to end effects.

The B- to Z-DNA "transition energy" of d(m<sup>5</sup>CG) dinucleotides was estimated by measuring the ratio of DNA in the left- versus the right-handed conformation of poly(dG-dm<sup>5</sup>C). Polymeric d(m<sup>5</sup>CG) from Pharmacia was dialyzed against several 2-L changes of glass-distilled, deionized water. The sample was titrated with 0.1 M NaCl and then to 5.0 M NaCl in 1 M increments. The ratio of B- and Z-DNA present in solution at each titration point was calculated from the ratio of 295- versus 260 nm absorption as recorded on an HP-8451 spectrophotometer.



Table 2.1. Calculated Solvent-Accessible Surfaces (SAS) and Free Energies of Transfer ( $\Delta G^{\circ}H^*$ ) from Octanol to Water of Aliphatic, Aromatic, and Phosphodiester Compounds.

Compound	C	SAS (Å <sup>2</sup> )			ΔG°H* (kcal mol <sup>-1</sup> )
		O	N	PO <sub>2</sub>	
aliphatic surfaces					
ethanol	58.8	36.1			-0.43
1-propanol	79.3	36.1			0.37
1-butanol	96.8	36.1			1.21
1-pentanol	117.2	36.1			2.02
1-hexanol	134.9	36.1			3.14
methoxyethanol	67.9	61.6			-1.04
ethane-1,2-diol	39.3	72.1			-2.63
aromatic surfaces					
benzene	111.5				2.86
naphthalene	160.8				4.57
anthracene	204.7				6.07
pyrrole	75.2		15.0		1.02
pyridine	91.6		13.7		0.88
pyrimidine	51.5		26.2		-0.54
imidazole	64.1		26.6		-0.11
aniline	98.3		24.6		1.23
o-phenylenediamine	86.6		43.8		0.20
benzophenone	177.2		13.8		4.61
quinone	82.5		29.9		0.27
phosphodiester					
diethyl phosphate	121.2			46.9	0.48
dibutyl phosphate	195.2			52.7	3.01

## Results

Derivation of atomic solvation parameters (ASP) for DNA surfaces. The thermodynamic weights associated with solvating the various surface types in DNA were derived by first dissecting the nucleotide components of DNA into its substituent groups. The primary components include the ribose sugar, an aromatic base, and a negatively charge phosphodiester. Each component can be further simplified to a sum of surface types, and each surface type can be mimicked by a series of small organic molecules whose energies of hydration can be calculated from their partition coefficients in an octanol/water solvent system (Table 2.1). The slope of the relationship between the solvent-accessible surface (SAS) of these molecules and their respective hydration energies defines the ASP value for that surface type (Figure 2.1).

The ribose component consists of two surface types: (1) aliphatic carbon surfaces and (2) aliphatic oxygen surfaces. These surfaces were mimicked by using a series of linear alcohols. The relationship between the SAS for a set of *n*-alcohols (for carbon lengths from 2 to 6) and their  $\Delta G^\circ_H$  was essentially linear with a slope of  $+0.43 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  and y intercept of  $-2.86 \text{ kcal mol}^{-1}$ . The very negative y intercept is due to negative free energy for hydrating a hydroxyl group having no carbon surface associated with it. The actual ASP value for the aliphatic hydroxyl group was derived from the  $\Delta G^\circ_H$  and SAS of ethanol, methoxyethanol, and ethane-1,2-diol. The  $\Delta G^\circ_H$  value of each compound was corrected for the contributions from the surfaces from the carbon chains by using the ASP value for aliphatic carbons and eq 1. The linearity of this relationship with the inclusion of methoxyethanol shows that the ring oxygen (O1') of the ribose can be treated as any other hydroxyl surface.

The aromatic bases found in naturally occurring in DNA are heterocyclic ring structures composed of surfaces from four atom types: (1) unsaturated carbons, (2) ring

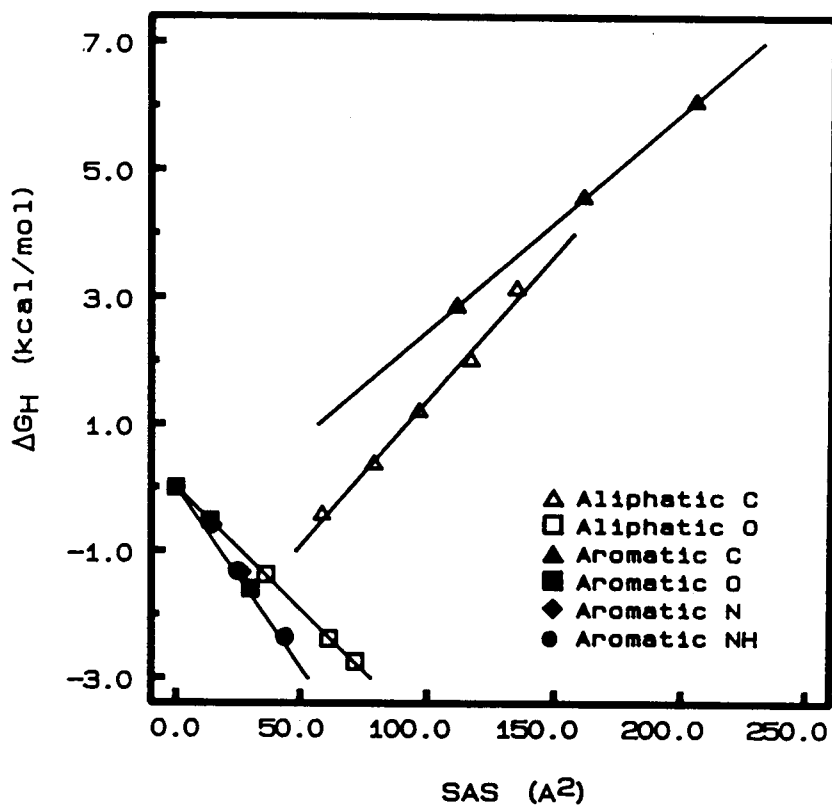


Figure 2.1. Plot of free energies of hydration ( $\Delta G^*_H$ ) versus solvent-accessible surfaces (SAS) calculated for small organic molecules.  $\Delta G_H$  in the figure is equivalent to  $\Delta G^*_H$  in the text. The open symbols represent aliphatic surfaces and the solid symbols aromatic surfaces. The symbols for each surface type listed in Table 2.1 are defined in the figure.

nitrogens, (3) conjugated amides, and (4) conjugated carbonyl oxygens. The ASP values of the aromatic carbon surface were calculated from the SAS and the  $\Delta G^\circ_H$  of benzene, naphthalene, and anthracene (Figure 2.1). This relationship yielded a slope of  $0.034 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  and a y intercept of  $-0.98 \text{ kcal mol}^{-1}$ . The significance of this negative y intercept is unclear. It may be related to an intrinsic solubility of aromatic compounds in water or to a nonlinear relationship for aromatic compounds with less than six atoms. In either case, we will be comparing differences in free energies of hydrations calculated by applying the ASP values to SAS of different DNA surfaces. The y intercept, being an additive value to the surfaces of all aromatic surfaces, will be negated in these differences. The ASP values for the ring nitrogens, amides, and carbonyl oxygens were treated as a single hydrophilic surface type. The relationship between  $\Delta G^\circ_H$  and SAS for the aromatic hydrophilic surfaces is linear (slope =  $-0.055 \text{ kcal mol}^{-1}$ ) with a y intercept of 0. The linearity of the relationship is evidence that these hydrophilic surfaces can indeed be treated as a group, and the y intercept at 0 demonstrates that the ASP value derived for the aromatic carbon surfaces is valid for aromatic molecules that are as small as five non-hydrogen atoms and consisting of only three carbon atoms.

The ASP values for the phosphodiester were derived from diethyl and dibutyl phosphate. Because of the negative charge associated with the phosphate group, these compounds are very insoluble in octanol. The  $\Delta G^\circ_H$  values were thus calculated for a partitioning of these compounds between *sec*-pentanol and 0.1M NaClO<sub>4</sub> aqueous solution. The  $\Delta G^\circ_H$  value of these two compounds was corrected for the contributions from the aliphatic carbon surfaces, averaged, and divided by the SAS of the phosphodiester group. The phosphorus and oxygen surfaces were treated as a group rather than individually. Since the aqueous solvent is ionic, some counterions are likely to be carried into the organic phase. The partition  $\Delta G^\circ_H$  is likely to be underestimated in this calculation. As we will discuss later, the contribution of the phosphate surface to the differences in the  $\Delta G^\circ_H$  values of the DNA conformations is not dramatic, and the overall

conclusions from these studies is not strongly dependent on the precise ASP value of the phosphate groups. The ASP values for all the surface types represented in DNA are listed in Table 2.2.

Free energies of hydration ( $\Delta G^{\circ}_H$ ) of dinucleotides. The calculated solvent-accessible surfaces (SAS) of the dinucleotides d(CG), d(CA)•d(TG), and d(TA) in the B- and Z-DNA conformations show that, overall, the hydration energies of unmodified dinucleotides are more positive in the left-handed Z conformation (Table 2.3). This is consistent with findings that Z-DNA is stabilized in solution by dehydrating conditions such as high salt (Pohl & Jovin, 1972; Wang et al., 1981; Drew & Dickerson, 1981) or alcohol concentrations (Pohl, 1976) and is accounted for by the overall more positive energy of hydration for the phosphates in Z-DNA than in B-DNA. The hydration energies of alternating pyrimidine-purine dinucleotides increase as the ratio of TA base pairs in the dinucleotides increases, regardless of their conformation. These results indicate that alternating CG base pairs have a lower hydration energy than do TA base pairs. This base-pair dependence of the hydration energies agrees with the report that the partition coefficient (aqueous to organic solvent) is more positive for polyribo(AU) than for polyribo(CG) (Garel et al., 1972).

The hydration energies ( $\Delta G^{\circ}_H$ ) calculated for d(CG) base pairs generated in the B-DNA conformation by using standard helical parameters were compared to the hydration energies calculated for the analogous base pairs in the B-DNA crystal structure of the self-complementary sequence d(CGCGAATTCGCG) (Dickerson & Drew, 1981). The  $\Delta G^{\circ}_H$  calculated for individual CG base pairs within the crystal structure varied by as much as 20%. The average  $\Delta G^{\circ}_H$  of the model CG base pairs ( $-5.73 \text{ kcal mol}^{-1}$  of base pair), however, differed by less than 1% from the average  $\Delta G^{\circ}_H$  calculated for the CG base pairs from the crystal structure ( $-5.69 \text{ kcal mol}^{-1}$  of base pair). Thus, even though the hydration of individual base pairs may deviate from the average value, we believe that the structures generated for each sequence are valid representations of B-DNA in solution

Table 2.2. Atomic solvation parameters (ASP) of hydrophilic, hydrophobic, and charged phosphate surfaces in nucleic acids. The parameters are derived from the partition coefficients and calculated solvent-accessible surfaces of small organic molecules.

group	surface type	ASP (kcal mol <sup>-1</sup> Å <sup>-2</sup> )
ribose	hydrophobic (C)	0.043
	hydrophilic (O)	-0.038
base	hydrophobic (C)	0.034
	methyl (C)	0.043
	hydrophilic (O/N)	-0.055
phosphate	charged (O/P)	-0.100

and that the SAS calculated provide reasonable values for determining the hydration energies of representative base pairs of each sequence in solution.

Differences in hydration free energies between Z- and B-DNA ( $\Delta\Delta G^{\circ}_H(Z-B)$ ). If we accept that the difference in stability between B- and Z-DNA is dependent on solvent interactions, then the difference in hydration energies between the conformations ( $\Delta\Delta G^{\circ}_H(Z-B)$ ) should be more negative for sequences that favor Z-DNA formation and more positive for sequences that favor the B conformation. Thus,  $\Delta\Delta G^{\circ}_H(Z-B)$  values should increase according to the hierarchy  $d(TA) > d(CA) \cdot d(TG) > d(CG)$  established for the instability of these dinucleotides as Z-DNA. To test this, we would need to

Table 2.3. Calculated hydration and experimental B-Z transition energies (kcal mol<sup>-1</sup>) for DNA dinucleotides.

dinucleotide sequence	hydration energy			B-Z transition energy	
	$\Delta G^{\circ}H(B)$	$\Delta G^{\circ}H(Z)$	$\Delta G^{\circ}H(Z-B)$	$\Delta G^{\circ}T(B-Z)$	ref
d(m <sup>5</sup> CG) <sup>a</sup>	-4.344	-4.838	-0.494	-0.6 ± 0.3	this work
d(CG) <sup>a</sup>	-5.730	-5.502	0.228	0.66	Peck and Wang (1983)
				0.90	Nordheim et al. (1982)
				1.12	Frank-Kamenetskii and Vologodskii (1984)
				1.20	Mirkin et al. (1987)
d(CG) <sup>b</sup>	-5.689	NA	NA	NA	NA
d(CA)•d(TG)	-5.916	-5.160	0.756	1.32	Mirkin et al. (1987)
				1.38	Frank-Kamenetskii and Vologodskii (1984)
				1.40	Vologodskii and Frank-Kamenetskii (1984)
d(TA)	-5.242	-4.346	0.896	2.3	Mirkin et al. (1987)
				2.4	Ellison et al. (1986)
d(UA)	-5.032	-4.532	0.500	NA	NA

<sup>a</sup> The thermodynamic ASP values used to calculate the hydration energies of d(m<sup>5</sup>CG) and d(CG) (Ho et al., 1988) were those derived for proteins from Eisenberg and MacLachlan (1986). The values reported in this study have been updated by using ASP values that are more applicable to the methods used to calculate the solvent-accessible surfaces and the functional groups of nucleic acids. Although this affects the details of the results from the previous study, it does not alter the overall results of conclusions drawn from that study. <sup>b</sup> Average hydration energies calculated from the 2nd, 3rd, 10th, and 11th CG base pairs of the d(CGCGAATTCGCG) B-DNA crystal structure (Dickerson & Drew, 1981).



compare the calculated  $\Delta\Delta G^\circ_{\text{H}}(\text{Z-B})$  values to their stabilities as Z-DNA. The stabilities of the dinucleotides d(CG), d(CA)•d(TG), and d(TA) in the right- versus left-handed conformations at low ionic strengths have been determined by various groups as B-Z transitions free energies,  $\Delta G^\circ_{\text{T}}(\text{Z-B})$ . The values of  $\Delta G^\circ_{\text{T}}(\text{Z-B})$  are a measure of the energy available from negative supercoiling required to induce Z-DNA formation from B-DNA within well-defined sequences inserted into closed circular plasmids. These values should thus be more positive for sequences that do not readily adopt the Z conformation. In all these studies, the degree of negative superhelicity that is required to induce the formation of Z-DNA was determined for well-defined inserts in closed circular plasmids. Using a two-state zipper model to analyze the transition, several laboratories have derived a set of energies for initiating the B-Z transition and for propagating Z-DNA formation through the various inserts (see Table 2.3). Since these measurements were made at low ionic strengths ( $[\text{Na}^+] < 3\text{mM}$ ), the transitions observed are primarily supercoil induced. The propagation energies are thus a measure of the relative intrinsic stabilities of the dinucleotides as Z-DNA under low ionic strength conditions.

Stability of poly(dG-dm<sup>5</sup>C) as Z-DNA. Aside from the base-sequence dependence of Z-DNA stability, methylation of cytosine bases at the C-5 has been shown to stabilize poly(dG-dC) in the left-handed form (Behe & Felsenfeld, 1981). The B- to Z-DNA transition energy of d(m<sup>5</sup>CG) dinucleotides, however, has not been previously determined. To estimate the relative stability of this dinucleotide as B- and Z-DNA under conditions similar to that of the supercoil-induced transitions, the equilibrium ratio of the two conformations present in polymeric d(m<sup>5</sup>CG) was spectroscopically determined after extensive dialysis against deionized water. The polymer under these low ionic strength conditions was observed to exist as a 0.65:0.35 mixture of Z- and B-DNA (Figure 2.2), equivalent to an equilibrium constant ( $K_{\text{Z/B}}$ ) of 1.86 or a difference in stability of  $-0.6 \pm 0.3 \text{ kcal mol}^{-1}$  of dinucleotide. The addition of 0.1M sodium chloride to the solution

shifted the equilibrium to an entirely B conformation. The transition to Z-DNA was not observed again until greater than 2 M NaCl had been added.

These titration results show that poly(dG-m<sup>5</sup>dC) is stable as Z-DNA under low salt conditions but is not entirely in the left-handed conformation as observed in several previous papers (Narasimhan & Bryan, 1984; Kruger & Prairie, 1985; Latha & Brahmachari, 1985; Fuerstein et al., 1985). Devarajan and Shafer suggested that the Z-DNA observed under these conditions is due to contaminating divalent cations such as Mg<sup>2+</sup> and Ca<sup>2+</sup> (Devarajan & Shafer, 1986). They reported that under conditions of extremely low divalent cation concentrations poly(dG-m<sup>5</sup>dC) is in the B conformation. However, a different group (Feuerstein et al., 1985) reported that, under similar dialysis conditions, poly(dG-m<sup>5</sup>dC) is in the left-handed conformation. Because we attempted in our titrations to mimic the conditions of the supercoil-induced transitions, divalent cations were not rigorously depleted from the polymer. It is thus not unreasonable that this polymer is poised at an equilibrium point between the two conformations under our conditions.

Relationship of  $\Delta\Delta G^{\circ}H(Z-B)$  to Z-DNA stability. Table 2.3 is a comparison of  $\Delta\Delta G^{\circ}H(Z-B)$  calculated for the dinucleotides d(m<sup>5</sup>CG), d(CG), d(CA)•d(TG), and d(TA) to their respective thermodynamic stabilities of Z- versus B-DNA, as defined experimentally from the B-Z transition energies ( $\Delta G^{\circ}T(Z-B)$ ). From the small values of  $\Delta G^{\circ}T(Z-B)$ , we would expect that the differences in the hydration free energies ( $\Delta\Delta G^{\circ}H(Z-B)$  or solvent free energies (SFE)) should also be small, and they were. These small values result from the difference of rather large numbers and represent at most a 16% change in the overall hydration energies calculated for each dinucleotide. One can ask then whether these small perturbations are relevant. Hydration of the phosphate groups contributes significantly to the magnitude of the negative  $\Delta G^{\circ}H$  values, owing to

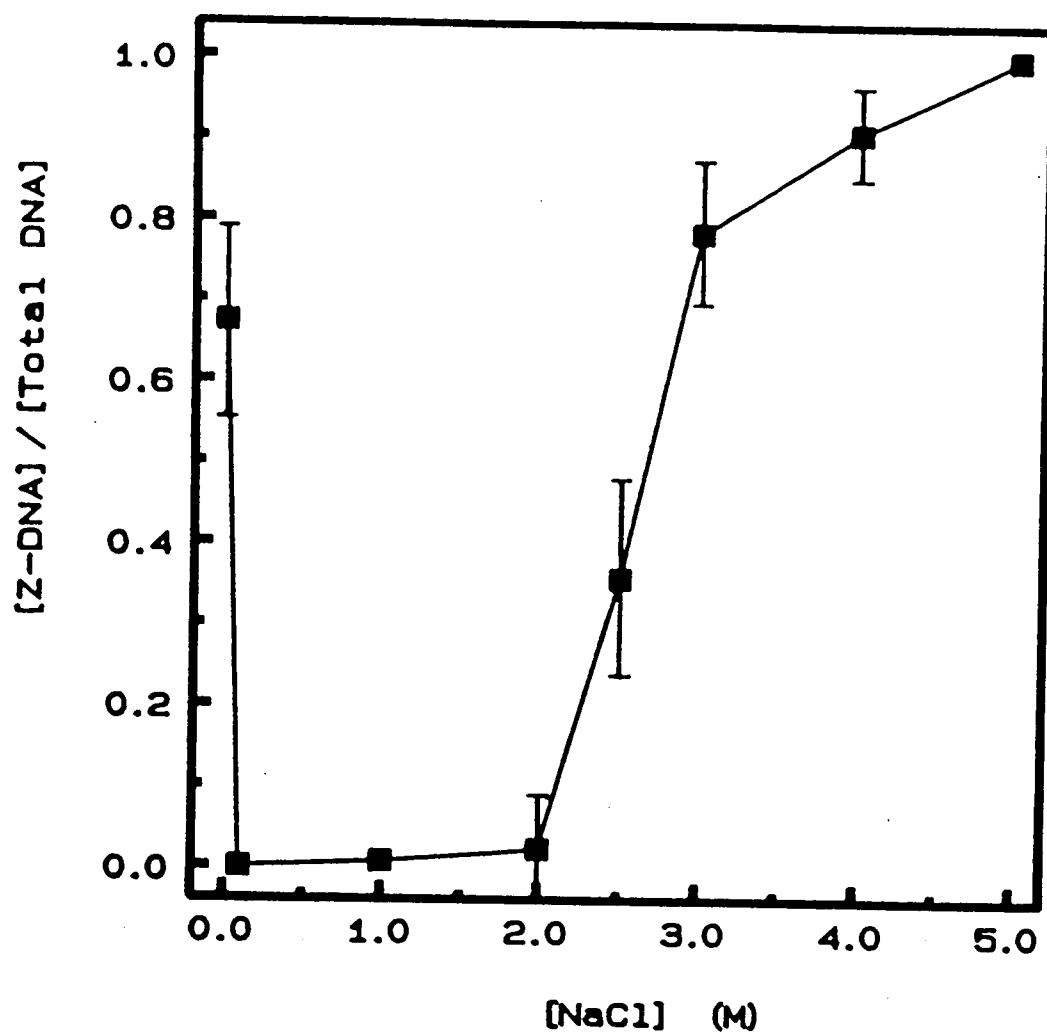


Figure 2.2. Titration of poly(dG-dm<sup>5</sup>C) with NaCl. The fraction of Z-DNA to the total DNA in solution was calculated from the 295 to 260 nm absorption ratios at each titration point. The data represent three independent sets of titrations.

the large negative ASP for hydration of phosphate surfaces. The differences in the overall exposure of the phosphates between a dinucleotide in the B- versus the Z-conformation is  $2.6 \pm 2.3 \text{ \AA}^2$ . This translates into an overall free energy of hydration difference of  $0.26 \pm 0.23 \text{ kcal mol}^{-1}$  from d(m<sup>5</sup>CG) to the d(TA) dinucleotides. The deviation across these differences in phosphate exposure is small. This would be expected since the overall backbones are nearly identical for alternating purine and pyrimidine sequences in each conformation. The  $\Delta\Delta G^{\circ}H(Z-B)$  ranges from -0.49 to 0.90, or a  $1.39 \text{ kcal mol}^{-1}$  of dinucleotide difference from d(m<sup>5</sup>CG) to d(TA). The contribution of the  $\Delta\Delta G^{\circ}H(Z-B)$  variation resulting from differences in the hydration free energies of the phosphates is thus approximately 17%. The remainder is due to differences in the exposure of the bases. A 20% error in the phosphate ASP values would increase the contribution of the phosphates from 17% to 20% toward the  $\Delta\Delta G^{\circ}H(Z-B)$  values.

In Figure 2.3,  $\Delta G^{\circ}T(B-Z)$  is plotted against  $\Delta\Delta G^{\circ}H(Z-B)$ . Within the variation of the reported  $\Delta G^{\circ}T(B-Z)$  values, this relationship is linear. A linear least-squares fit of this plot yields a relationship having a slope of 0.491 and a y-intercept at  $-0.130 \text{ kcal mol}^{-1}$ , with a correlation coefficient (R) of 0.96. The slope  $< 1$  of this relationship suggests that the hydration energies calculated consistently underestimate the experimental transition energies. Both the calculated  $\Delta\Delta G^{\circ}H(Z-B)$  and the experimental  $\Delta G^{\circ}T(B-Z)$  values are measures of the stability of Z-DNA versus B-DNA. The observation of a one-to-one relationship between stability and hydration energies with a positive slope and a y intercept approximately equal to zero and the calculated and experimental energies being of the same orders of magnitude suggest that indeed the transition of a base sequence from B- to Z-DNA is at least partially accounted for by the difference in hydration of the sequences in the two conformations.

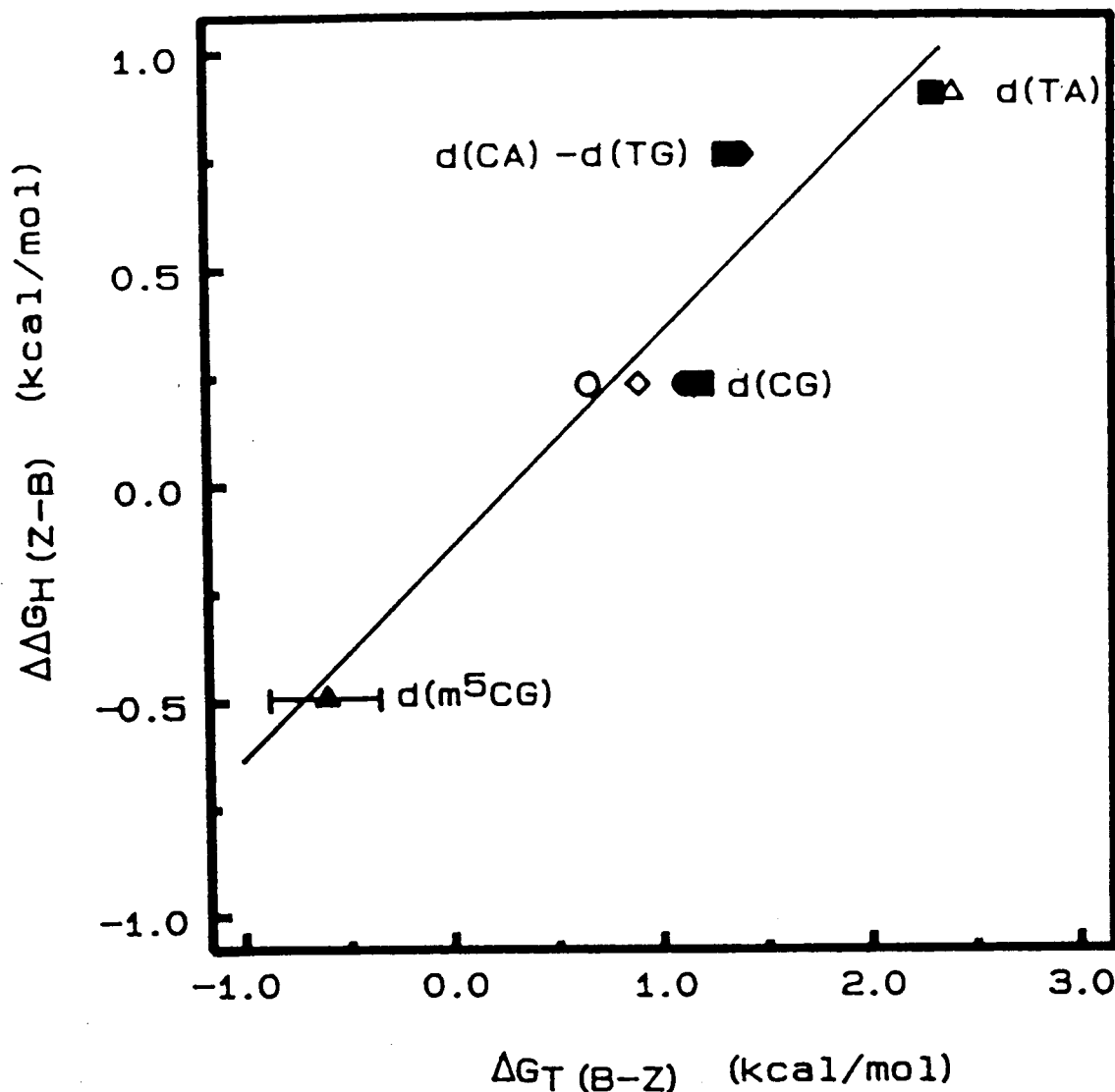


Figure 2.3. Comparison of the calculated differences in hydration energies of Z- and B-DNA with the experimentally determined B- to Z-DNA transition energies.  $\Delta\Delta G_H(Z-B)$  and  $\Delta G^*T(B-Z)$  in the figure are equivalent to  $\Delta\Delta G^*H(Z-B)$  and  $\Delta G^*T(B-Z)$  respectively in the text. A linear least-squares fit of these points yielded the relationship  $\Delta\Delta G^*H(Z-B) = 0.491\Delta G^*T(B-Z) - 0.130$ , with a correlation coefficient ( $R$ ) of 0.96. ( $\blacktriangle$ ) Data from this work; ( $\circ$ ) data from Peck and Wang (1983); ( $\diamond$ ) data from Nordheim et al. (1982); ( $\bullet$ ) data from Frank-Kamenetskii and Vologodskii (1984); ( $\blacksquare$ ) data from Mirkin et al. (1987); ( $\blacklozenge$ ) data from Vologodskii and Frank-Kamenetskii (1984); ( $\blacktriangle$ ) data from Ellison et al. (1986).

Destabilization of Z-DNA by d(TA) dinucleotides. A better understanding of the molecular basis for the differences in the hydration of these sequences and their dependence on conformation can be attained by comparing the contributions of each component surface to the overall hydration energy calculated. The difference between solvation of the hydrophobic and of the hydrophilic surfaces of the aromatic base and ribose groups for the various dinucleotides in their B and Z conformations is summarized in Table 2.4. By analyzing the difference in exposed surfaces for Z- versus B-DNA of d(CG), d(CA)•d(TG), and d(TA), we can start to understand why TA base pairs destabilize Z-DNA. The average total difference in exposed hydrophobic surfaces of the sequences in Z- versus B-DNA is higher for d(TA) than for d(CG) dinucleotides, while that for d(CA)•d(TG) is intermediate between the two. This is observed as resulting from a markedly greater exposure of the hydrophobic carbon surfaces and lower exposure of

Table 2.4 Differences in hydrophobic and hydrophilic solvation free energies ( $\Delta\Delta G^\circ H(Z-B)$ ) for base and ribose surfaces of DNA dinucleotides. The solvation free energies are reported in units of kcal mol<sup>-1</sup>.

dinucleotide sequence	$\Delta\Delta G^\circ H(Z-B)$		$\Delta\Delta G^\circ H(Z-B)$	
	<u>hydrophilic</u>		<u>hydrophobic</u>	
	base	ribose	base	ribose
d(m <sup>5</sup> CG)	0.08	-0.26	-0.06	-0.78
d(CG)	0.02	-0.03	0.08	-0.19
d(CA)•d(TG)	0.13	-0.12	0.08	-0.04
d(TA)	0.20	0.06	0.04	0.08
d(UA)	-0.08	0.06	0.05	0.05

the hydrophilic oxygen surface of the ribose sugars in TA dinucleotides. The difference in exposure of both the hydrophobic and hydrophilic surfaces of the aromatic bases is lower in the d(TA) dinucleotides when compared to d(CG) base pairs.

These results can be attributed to the difference in the stacking arrangement of TA base pairs as compared to CG base pairs. In the purine-pyrimidine stacking of base pairs in the Z conformation, the purine bases stack directly above the ribose of adjacent base pairs. A comparison of the stacking arrangements of CG:CG to TA:TA dinucleotides as Z-DNA is shown in Figure 2.4 A,B. In Figure 2.4, the larger solid circles represent the contact points of solvent molecules with hydrophobic surfaces, while the smaller open circles represent the contact points with hydrophilic surfaces as calculated by the rolling ball method (Connolly, 1983). The absence of the N2 amino group of adenine exposes the surface of its adjacent thymine ribose in the d(TA) dinucleotides relative to the d(CG) dinucleotides, resulting in a greater exposure of the hydrophobic surface in the minor groove surface of d(TA). The overall exposure of the carbon surfaces of the bases, particularly of adjacent base pairs, is actually lower in the TA dinucleotides than in the CG dinucleotides, even though the methyl group itself contributes to the local exposure of hydrophobic surfaces at the major groove surface of the TA base pairs. The higher overall hydrophobicity of the d(TA) dinucleotides in the Z conformation results from the greater increase in the exposed carbon surface of ribose as compared to the decrease in the exposed carbon surfaces of the bases.

The surface areas of the accessible hydrophilic groups are slightly higher for d(CG) base pairs as Z-DNA than as B-DNA, resulting in negative  $\Delta\Delta G^{\circ}H(Z-B)$  values for the oxygens of the riboses and the oxygens and nitrogens of the aromatic bases. The hydrophilic areas of the TA base pairs are less accessible in the Z versus the B conformations.  $\Delta\Delta G^{\circ}H(Z-B)$  becomes significantly more positive as the ratio of TA base

Figure 2.4. Stacking diagram of d(CG), d(TA) and d(UA) dinucleotides and their solvent-accessible surfaces. Panel A shows a d(CG) dinucleotide, panel B a d(TA) dinucleotide and panel C a d(UA) dinucleotide. The large solid circles represent the calculated hydrophobic surfaces of carbons, and the smaller open circles represent the calculated hydrophilic surfaces of the oxygens and nitrogens. The surfaces were calculated with  $1.0 \text{ \AA}^{-2}$  are shown and the hydrogen atoms removed from the base pair models for clarity. The SAS values used to determine the  $\Delta G^\circ_H$  values were calculated with a density of  $5 \text{ \AA}^{-2}$ .



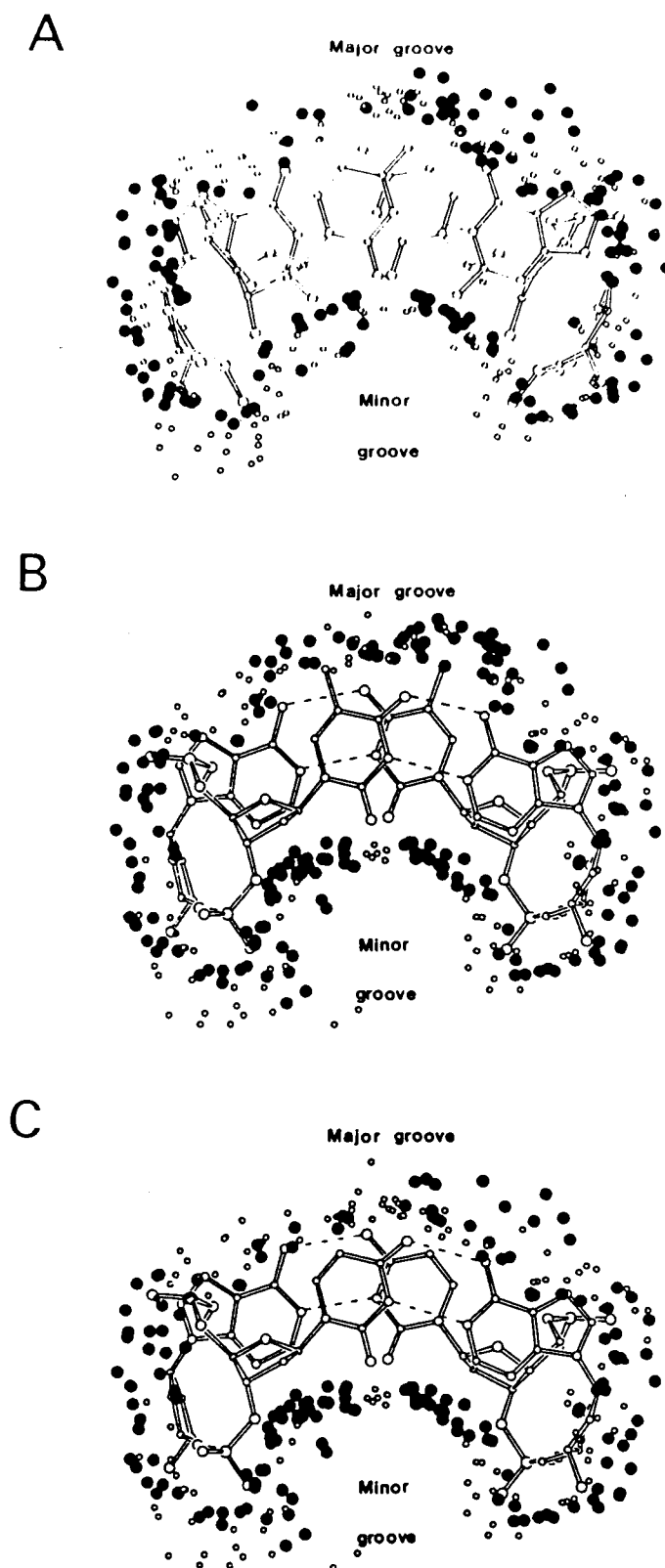


Figure 2.4

pairs is increased from 0 to 1 to 2 per dinucleotide. Thus, overall, the addition of TA base pairs to a DNA sequence increases the difference in the hydration free energies of that sequence as Z- and as B-DNA.

Stability of d(UA) dinucleotides as Z-DNA. To determine the extent to which the thymine C-5 methyl contributes to these differences in hydration free energies of d(TA) base pairs in the left-handed conformation, these same hydration energy calculations were repeated for the sequence d(UA)<sub>3</sub> as B- and Z-DNA (Table 2.4). This sequence is identical to d(TA)<sub>3</sub>, except that the methyl groups of the pyrimidines have been removed to form the analogous deoxyuridine. The exposure of hydrophobic surfaces in the minor groove crevice are nearly identical for the d(TA) and d(UA) dinucleotides (Figure 2.4 B,C). The accessibility of the ribose carbons is, thus, not significantly affected by the methyl group, confirming the earlier suggestion that the lack of an N2 amino group on the adenine bases is responsible for the greater exposure of the ribose to solvent in d(TA) dinucleotides. Alternatively, the major groove surface of d(UA) appears much more hydrophilic overall than hydrophobic as compared to the comparable surface of d(TA) and even the major groove surface of d(CG) (Figure 2.4). The  $\Delta\Delta G^{\circ}_H(Z-B)$  values in Table 2.4 show that removing this methyl group results in an increase in the accessibility of both the hydrophobic and hydrophilic groups of the aromatic bases and a decrease in the hydrophobic surface of the ribose. This is consistent with previous reports that the C-5 methyl group of pyrimidine bases sits in a pocket formed by the ribose and the stacked purines of adjacent base pairs (Ho et al., 1988; Fujii et al., 1982). Contrary to previous reports (Jovin et al., 1983), however, the overall effect of the C-5 methyl appears to be different for thymine as compared to cytosine.

The C-5 methyl group of the pyrimidine bases effectively increases the overall exposed surface of B-DNA. The surface in Z-DNA, however, is decreased because the methyl group sits in a pocket. The effect of the C-5 methyl on the stability of Z-DNA, therefore, is dependent on the atom types that line the surface that make up this pocket.

For cytosine bases in alternating d(CG) dinucleotides, the C-5 pocket is lined equally by hydrophobic and hydrophilic atoms from the ribose and bases of adjacent base pairs. Thus, the decrease in exposed hydrophilic surface is accompanied by a near equivalent decrease in the hydrophobic surfaces, including the additional methyl surface (Ho et al., 1988). In the B conformation, this methyl group sits exposed in the minor groove. This accounts for the stabilizing effect of methylation of d(CG) dinucleotides as Z-DNA.

The C-5 pocket of d(TA) dinucleotides, as with d(CG), in the Z conformation is lined by both hydrophobic and hydrophilic surfaces. One difference in the effect of the methyl group on the stability of Z-DNA lies in how this group sits within the pocket. The C-5 methyl of d(TA) is more exposed in the major groove surface than that of d(m<sup>5</sup>CG) by more than 1 Å<sup>2</sup> per methyl group. This is equivalent to a 0.2 kcal mol<sup>-1</sup> destabilization of d(TA) dinucleotides relative to d(CG) by the methyl group. If we consider only the exposed surface of the methyl, however, we cannot account for the entire effect of the addition of one carbon. This methyl also affects the exposure of the neighboring atoms to solvent. Methylation decreases the exposure of all atoms in this pocket, but differently for the two types of dinucleotides. Methylation decreases the hydrophobicity of the d(CG) due to exposure of carbon atoms in the bases by 0.14 kcal mol<sup>-1</sup>, as compared to 0.02 kcal mol<sup>-1</sup> for d(TA) dinucleotides. Alternatively, methylation of d(CG) increases the contribution of the hydrophilic oxygen surfaces to the hydration of the riboses by 0.23 kcal mol<sup>-1</sup>, while the methyl group of d(TA) buries a hydrophilic surface equivalent to 0.26 kcal mol<sup>-1</sup> for each dinucleotide. These effects are thus dependent on the positioning of the methyl group relative to the adjacent base pairs and riboses resulting from the stacking arrangements in Z-DNA.

Relationship of  $\Delta\Delta G^\circ H(Z-B)$  to crystallization of hexamers as Z-DNA. From these observations, we predict that the propensity of d(UA) dinucleotides to form Z-DNA is greater than that of the analogous C-5 methylated d(TA) dinucleotides. The calculated  $\Delta\Delta G^\circ H(Z-B)$  value for d(UA) dinucleotides (0.500 kcal mol<sup>-1</sup>) is 0.396 kcal mol<sup>-1</sup>

lower than that of d(TA). This translates into a  $0.81 \text{ kcal mol}^{-1}$  lower  $\Delta G^{\circ}_T(\text{B-Z})$ , as calculated by from the relationship from Figure 2.3, that would result by removing the methyl group of d(TA) dinucleotides. The d(UA) dinucleotide is still  $0.55 \text{ kcal mol}^{-1}$  less stable as Z-DNA than d(CG) and  $2.02 \text{ kcal mol}^{-1}$  less stable than the d(m<sup>5</sup>CG) dinucleotides, but should have the equivalent ability to form Z-DNA as d(CA)•d(TG). To test this prediction, the hexamer sequence d(m<sup>5</sup>CGUAm<sup>5</sup>CG) was synthesized and crystallized. We have solved the structure of this sequence to better than  $1.3\text{\AA}$  resolution and found it to be in the Z-conformation (Zhou & Ho, 1990). This hexamer sequence is analogous to the sequence d(m<sup>5</sup>CGTAm<sup>5</sup>CG) and d(m<sup>5</sup>CGm<sup>5</sup>CGm<sup>5</sup>CG), both of which have been previously crystallized in the Z-conformation (Wang et al., 1984; Fujii et al., 1982). A comparison of the crystallization conditions of d(m<sup>5</sup>CGUAm<sup>5</sup>CG) with those for d(m<sup>5</sup>CGm<sup>5</sup>CGm<sup>5</sup>CG) and d(m<sup>5</sup>CGTAm<sup>5</sup>CG) would allow us to determine whether in fact d(UA) dinucleotides are more stable in the Z conformation than d(TA) and less stable than d(m<sup>5</sup>CG) dinucleotides. We assume in this comparison that there is a critical equilibrium concentration of the hexamer sequences in the Z conformation that is required for initiating and propagating the crystal packing of Z-DNA in its crystal lattice and that this equilibrium concentration is strongly dependent on the salt concentration during crystallization.

Z-DNA is stabilized by both mono- and divalent cations (Pohl & Jovin, 1972; Wang et al., 1981; Drew & Dickerson, 1981). It is therefore not surprising that the conditions required to crystallize hexamer sequences as Z-DNA include high concentrations of both Na<sup>+</sup> and Mg<sup>2+</sup>. The effectiveness of each type of cation at stabilizing Z-DNA in solution is dependent on the charge as well as its concentration. As expected, Mg<sup>2+</sup> is much more effective at stabilizing Z-DNA than Na<sup>+</sup> (Chen et al., 1984). The ionic conditions for stabilizing Z-DNA can thus be related to the ionic strength of the solution, which is dependent on both the concentration and the square of

the ionic charge ( $Z^2$ ). Even though ionic strengths may not be a valid representation of specific interactions of cations with the DNA, it is a useful measure of the ionic conditions of the bulk solution. Since anions do not appear to play a significant role in Z-DNA stability, the Z-DNA crystallization conditions can be translated into cationic strengths (CS) as defined by equation 2.

$$CS = Z^2Na^+[Na^+] + Z^2Mg^{2+}[Mg^{2+}] \quad \text{Eq. 2.}$$

We would expect that sequences more readily stabilized as Z-DNA would require lower concentrations of  $Na^+$  and  $Mg^{2+}$  to be stable in the Z conformation and, therefore, lower concentrations of these cations to form crystals as Z-DNA. Synthetic hexanucleotide sequences that crystallize as Z-DNA have been found to be isomorphous and, therefore, provide us with a consistent data set to compare the abilities of sequences to form Z-DNA. The hexamers in these crystals are packed end-to-end to form essentially continuous strands of Z-DNA aligned along the crystallographic  $c$  axis. The stability of the dinucleotides in the hexamers as the Z versus B conformations can thus be assessed as if the dinucleotides were all within continuous strands of Z-DNA without the need to consider end effects. The average difference in solvation of the dinucleotides in each hexamer was estimated by summing the hydration energies of each component dinucleotide and dividing this by the number of dinucleotides in the hexamer. This value can be translated into a term reflecting the average stability if the dinucleotides  $\Delta G^{\circ}T(B-Z)$  in the hexamers as Z-DNA by using the relationship derived from Figure 2.3.

The initial crystallization conditions and the equilibrium cationic strength calculated for crystallization of the hexamer sequences  $d(m^5CG)_3$ ,  $d(m^5CGTAm^5CG)$ ,  $d(CG)_3$ ,  $d(CACGTG)$ , and  $d(m^5CGUAm^5CG)$  as Z-DNA are listed in Table 2.5, along with the  $\Delta G^{\circ}T(B-Z)$  calculated from their  $\Delta\Delta G^{\circ}H(Z-B)$  values. The hexamer sequence  $d(m^5CGUAm^5CG)$  was predicted to be more stable as Z-DNA than the analogous

d(m<sup>5</sup>CGTAm<sup>5</sup>CG) sequence and less stable than d(m<sup>5</sup>CGm<sup>5</sup>CGm<sup>5</sup>CG). We found that indeed the cationic strength required to form diffraction-quality single crystals of Z-DNA of the uridine-containing hexamer was approximately 2-fold lower than that reported for the sequence d(m<sup>5</sup>CGTAm<sup>5</sup>CG). When these were compared to the crystallization conditions of the sequence d(m<sup>5</sup>CG)<sub>3</sub>, we observed that methylated d(CG) dinucleotides require the lowest cation strengths to obtain Z-DNA crystals. Thus, the dependence of Z-DNA crystal formation on these salt conditions mirrors previous findings that methylation of cytosine at the C-5 position stabilizes the Z conformation and d(TA) dinucleotides are the least stable as Z-DNA. The d(UA) dinucleotide, as predicted, is intermediate in its ability to crystallize as Z-DNA.

An overall comparison of the log of the cationic strength for crystallization of each hexamer and their associated B-Z transition energies is shown in Figure 2.5. This relationship is essentially linear with a slope calculated to be 0.585 and the y-intercept of -0.306 M from a linear least-squares fit with  $R = 0.96$ . The hexamer sequences that did not contain methylated cytosines were found to be overall less stable as Z-DNA than sequences containing methylated cytosines, regardless of the number of d(TA) base pairs in the structure, according to both the hydration energy calculations and the cationic strength values. This can be attributed to the very strong Z-DNA stabilizing ability of the flanking d(m<sup>5</sup>CG) dinucleotides in the three methylated hexamers. Thus, the hydration energy calculations were found to be useful not only for predicting *whether* alternating pyrimidine and purine dinucleotides form Z-DNA but also for predicting the *extent* to which they will adopt the left-hand conformation.

Table 2.5: Calculated B-Z transition free energies ( $\Delta G^\circ T(B-Z)$ ) and crystallization conditions for synthetic hexamers as Z-DNA.

hexamer sequence	B-Z equilibrium / $\Delta G^\circ T(B-Z)$	crystallization conditions (M)				ref
		[Na+]	[Mg++]	CS <sup>a</sup>	log (CS)	
d(m <sup>5</sup> CG) <sub>3</sub>	-0.706	0.150	0.020	0.230	-0.638	Fujii et al. (1982)
d(m <sup>5</sup> CGUAm <sup>5</sup> CG)	-0.031	0.180	0.045	0.360	-0.444	Zhou & Ho (1990)
d(m <sup>5</sup> CGTAm <sup>5</sup> CG) <sup>b</sup>	0.238	0.188	0.094	0.563	-0.249	Wang et al. (1984)
d(CG) <sub>3</sub>	0.765	0.400	0.500	2.200	0.342	Gessner et al. (1985)
d(CACGTG)	1.482	0.460	0.700	3.26	0.513	Coll et al. (1988)

<sup>a</sup> CS is the cationic strength of the crystallization conditions at equilibrium as calculated by  $CS = Z^2_{Na^+} [Na^+] + Z^2_{Mg^{2+}} [Mg^{2+}]$ , where Z is the cation charge. <sup>b</sup> The polycation spermine was reported for the crystallization conditions of d(m<sup>5</sup>CGTAm<sup>5</sup>CG) but was not found in the crystal structure. For this reason, we excluded this polycation from the CS calculations. We have recently crystallized this sequence in the absence of spermine under a cationic strength of 0.62M (Zhou & Ho, unpublished results), confirming that the presence of this polycation is not essential to obtaining Z-DNA crystals of this hexamer.

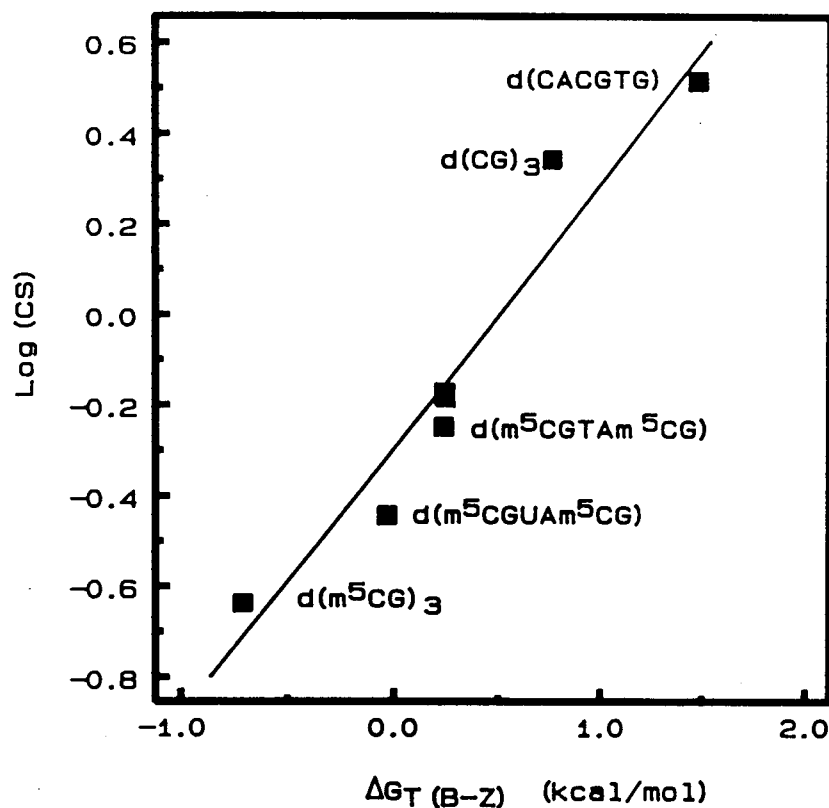


Figure 2.5. Plot of the free energies of transition calculated from the hydration free energies versus logarithm of cationic strengths for crystallizing synthetic hexanucleotides as Z-DNA.  $\Delta G_T(B-Z)$  in the figure is equivalent to  $\Delta G^*T(B-Z)$  in the text. Average hydration energies were calculated for the dinucleotides within each hexamer by summing the hydration free energies for each type of dinucleotide present and dividing by 3. Cationic strengths were calculated from the equilibrium sodium cacodylate and  $MgCl_2$  concentrations present under the conditions found to give diffraction-quality crystals of each hexamer sequence as Z-DNA (Table 2.5). A linear least squares fit of the data yielded a line described by  $\log(CS) = 0.586\Delta G^*T(B-Z) - 0.306$ .



## Discussion

In this study, we have examined the quantitative relationship between calculated solvent-accessible surfaces of dinucleotide sequences in B- and Z-DNA and the experimental transition energies for converting B- to Z-DNA for these dinucleotides. We found that a linear relationship can be drawn between the calculated energies and the experimentally determined transition energies. In addition, the magnitude of the differences in stabilities calculated from the hydration energies of the dinucleotides as Z- and B-DNA is similar to the B- to Z-DNA transition energies. This comparison showed that a simple model in which the stability of Z-DNA is attributed to differences in hydration of the DNA dinucleotides can account for the experimentally observed hierarchy of  $d(m^5CG) > d(CG) > d(CA) \cdot d(TG) > d(TA)$  dinucleotides in stability as Z-DNA.

The hydration energies calculated assume that all solvent interactions are averaged over the DNA molecule. To accommodate specific cationic interactions in these calculations, we would need to locate the positions of the cations from high-resolution crystal structures and treat these as essentially integral parts of the DNA structures. The cations in even the best-studied crystal structures of B- and Z-DNA sequences, however, have not all been located. It is also very unlikely that cations would be fixed in solution at all of these crystallographically determined positions. Soumpasis has quantitatively accounted for the high-salt transition of B- to Z-DNA in poly(dC-dG) by treating the main Coulombic interactions as an ion-ion potential in a homogeneous electrolyte (Soumpasis, 1984). Given our inability to accurately include specific counterions into the hydration energy calculation, the cation interactions were treated as one parameter in the solvent interactions that is essentially constant during the transition. This simplification appears to be valid primarily because the  $\Delta G^{\circ}T(B-Z)$  used as the measure of Z-DNA stability in these studies are restricted to low ionic strength conditions and are primarily supercoil induced.

By including the interactions of the cations into a general solvent interaction term, we have ignored the electrostatic contribution of the aquo-metal complexes to the stabilization of Z-DNA. The solvent interactions are thus underestimated by the hydration energy calculations. Since the strongest electrostatic stabilizations from the cations would be expected to be at the phosphate backbone of both the B and the Z conformations of DNA (Soumpasis, 1984), the electrostatic contribution to the hydration energies of the bases should be minimal. The electrostatic terms should therefore contribute a systematic error to the hydration energy calculations. Despite this, a relationship between the calculated differences in hydration and transition energies for B- and Z-DNA was derived that appears to be useful in predicting the relative stabilities of left- versus right-handed DNA.

An analysis of how the hydrophilic and hydrophobic groups contribute to the overall hydration energies of each dinucleotide allows us to understand the differences in stability of d(m<sup>5</sup>CG), d(CG), d(CA)•d(TG), and, in particular, d(TA) dinucleotides as Z-DNA at the atomic level. We find that the decreased overlap between the adenine and adjacent ribose resulting from the stacking of TA and UA base pairs in Z-DNA is responsible for the greater exposure of the ribose carbon surfaces when compared to that of d(CG) dinucleotides. The absence of the N2 amino group in adenine bases also results in a greater exposure of hydrophobic surfaces in the minor groove of d(TA) dinucleotides versus d(CG) dinucleotides. This suggests that the N2 amino group of purine bases would have a stabilizing effect on the Z-conformation.

The methyl substituent of thymine in d(TA) base pairs destabilizes Z-DNA contrary to the effect of C-5 methylation of cytosines by effectively decreasing the accessibility of the hydrophilic base and ribose surfaces to solvent. Thus, removal of this methyl group should help to stabilize d(UA) dinucleotides as Z-DNA. By comparing d(TA) to the d(UA) dinucleotides in their Z- and B-DNA conformations, we can assign a 0.81 kcal mol<sup>-1</sup> contribution of the C-5 methyl group to the instability of d(TA) dinucleotides as Z-

DNA. The effect of the C-5 methyl group on the stability of DNA conformations is not limited to the B and Z conformations. The thymine methyl group in poly(dA-dT) has, in contrast, been reported to help stabilize an alternating B and an X form of DNA (Vorlickova & Jaroslav, 1984). Since in these studies, we dealt only with two states for the conformation of each dinucleotide, the contribution of other forms of DNA to the overall equilibrium state of these sequences has been ignored. A significant contribution of other forms of DNA to the equilibrium, for example, the alternating B and the X forms of d(TA), would tend to effectively destabilize the Z conformation of that sequence. This two-state approximation may thus contribute to underestimation of the relative instability of Z-DNA for the dinucleotides d(CA)•d(TG) and d(TA) by the hydration energy calculations.

A comparison of the crystallization conditions required to obtain single crystals of the hexamer sequence d(m<sup>5</sup>CGUAm<sup>5</sup>CG) in the Z conformation demonstrated that indeed the methyl group of thymine contributes to the inability of d(TA) dinucleotides to form Z-DNA, as predicted from the calculations. The concentration of cations present at equilibrium in the conditions required to crystallize d(m<sup>5</sup>CGUAm<sup>5</sup>CG) as Z-DNA is approximately 2-fold lower than that of the sequence d(m<sup>5</sup>CGTAm<sup>5</sup>CG). In fact, a linear relationship can be obtained by comparing the cationic strength required to crystallize the sequences d(m<sup>5</sup>CGm<sup>5</sup>CGm<sup>5</sup>CG), d(m<sup>5</sup>CGUAm<sup>5</sup>CG), d(m<sup>5</sup>CGTAm<sup>5</sup>CG), and d(CGCGCG) with the Z- to B-DNA equilibria calculated from the solvent-accessible surfaces. This may ultimately be a useful relationship for helping to define the optimum conditions for crystallizing synthetic hexamers as Z-DNA.

## Chapter 3

### **Effects of Base Substituents on the Hydration of B- and Z-DNA: Correlations to the B- to Z-DNA Transition**

**Todd F. Kagawa, Meredith L. Howell, Kaihan Tseng, and P. Shing Ho**

### Abstract

We present a study of how substituent groups of naturally occurring and modified nucleotide bases affect the degree of hydration of right-handed B-DNA and left-handed Z-DNA. A comparison of poly(dG-dC) and poly(dG-dm<sup>5</sup>C) titrations with the lipotropic salts of the Hofmeister series infers that the methyl stabilization of cytosines as Z-DNA is primarily a hydrophobic effect. The hydration free energies of various alternating pyrimidine-purine sequences in the two DNA conformations were calculated as solvent free energies from solvent accessible surfaces. Our analysis focused on the N2 amino group of purine bases that sits in the minor groove of the double helix. Removing this amino group from guanine to form inosine (I) destabilizes Z-DNA, while adding this amino group to adenines to form 2-aminoadenine (A') stabilizes Z-DNA. These predictions were tested by comparing the salt concentrations required to crystallize hexanucleotide sequences that incorporate d(CG), d(CI), d(TA) and d(TA') base pairs as Z-DNA. Combining the current results with our previous analysis of major groove substituents, we derived a thermodynamic cycle that relates the systematic addition, deletion, or substitution of each base substituent to the B- to Z-DNA transition free energy.

## Introduction

The sequence-dependent behavior of DNA structure has been suggested to play a role in a number of transcriptional and replicative processes (reviewed in Sinden & Wells, 1992). An understanding of how DNA conformations are affected by the various substituent groups of the nucleotide bases, therefore, helps to extend our understanding of the various mechanisms available to control cellular functions. Aside from the canonical right-handed B-form, the left-handed Z-conformation is perhaps the best studied structural form of DNA (reviewed in Jovin et al, 1987). The structural and thermodynamic differences between B- and Z-DNA are fairly well understood in empirical terms. Studies show that, for the basic pyrimidine-purine dinucleotide (dn) repeat of Z-DNA, substitution of the cytosines with thymines, and guanines with adenines reduces the ability of sequences to adopt the left-handed conformation in essentially an additive manner (reviewed in Rich et al., 1984). The d(TA) dinucleotide is therefore less stable as Z-DNA than d(CA)•d(TG), and both are less stable than the prototypical d(CG) Z-DNA dinucleotide. The relative abilities of these dinucleotides to adopt the Z-conformation have been measured as the free energy of transition for the negative supercoiled induced B- to Z-DNA transition ( $\Delta G^{\circ}T(B-Z)$ ). The values for  $\Delta G^{\circ}T(B-Z)$  have been determined experimentally to be 0.66 kcal mol<sup>-1</sup> for d(CG) (Peck & Wang, 1983), 1.2 kcal mol<sup>-1</sup> for the d(CA)•d(TG) (Vologodskii & Frank-Kamenetskii, 1984), and 2.4 kcal mol<sup>-1</sup> for d(TA) dinucleotides (Ellison et al., 1986). The B- to Z-DNA transition, therefore, provides a unique system for studying the effects of base substituent groups on the thermodynamic differences between two interconvertible conformations of DNA, and allows one to test models and methods that predict the stability of macromolecular structures.

A number of theoretical methods have been employed, with varying success, to study the effects of substituent groups on the stability of the Z-conformation. Molecular mechanics calculations suggested that the Z-DNA stabilizing effect of methylating cytosines at the C5 position could be understood in enthalpic terms (Kollman et al., 1982). The inability of d(TA) dinucleotides to form Z-DNA, however, could not be readily explained by these methods, primarily because solvent interactions were not included in the calculations. Free energy perturbation calculations in which d(CG) dinucleotides were gradually "mutated" into d(TA) in an aqueous environment showed that mutation of the first base pair was detrimental to the stability of Z-DNA, but that a transformation of the second d(CG) base pair to d(TA) should potentiate formation of Z-DNA (Dang et al., 1990). These studies properly predicted that d(CA)•d(TG) would be less stable as Z-DNA compared to d(CG) dinucleotides; however, they also predicted that d(TA) dinucleotides would adopt the Z-conformation more readily than d(CA)•d(TG). This discrepancy between the calculated and the experimental results could not be readily explained.

In our own studies on B- and Z-DNA stabilities, we analyze solvent accessible surfaces (SASs) to determine the difference in solvent free energies (SFEs) for sequences as B- and Z-DNA (Kagawa et al., 1989). The differences in free energies between the two conformations for d(CG), d(CA)•d(TG) and d(TA) were of the same order of magnitude as those determined experimentally, and followed the trend that  $d(CG) < d(CG)•d(TG) < d(TA)$  in terms of hydrophobicity as Z-DNA. The general order of Z-DNA stability for these dinucleotides could thus be explained in terms of the differences in SFE associated with the base substituents in B- versus Z-DNA.

Using this same method, we have also shown that the stabilizing effect of methylating cytosines at the C5 position results from the methyl group actually filling in a hydrophobic pocket at the major groove of Z-DNA (Ho et al., 1988). The increase in hydrophobic surface that would be expected from adding a methyl group to the major

groove surface of Z-DNA is thus more than compensated for by the decrease in the exposed hydrophobic surface of the C5 pocket. An analysis of the C5 methyl of thymine, however, led to an entirely different conclusion (Kagawa et al., 1989). For d(TA) dinucleotides in Z-DNA, the C5 pocket was found to be overall more hydrophilic than that of d(CG). The C5 methyl group of thymine actually buries both hydrophilic as well as hydrophobic surfaces. We predicted, therefore, that demethylating the thymine base would actually facilitate the formation of Z-DNA. Thus, the effect of methylation at the C5 carbon of pyrimidine bases on the relative stability of Z- versus B-DNA is dependent on the specific base being modified.

Crystallographic studies on the self-complementary sequence d(m<sup>5</sup>CGUAm<sup>5</sup>CG) supported the prediction that the decreased hydrophobicity of the Z-DNA major groove surface helped to stabilize d(UA) dinucleotides in the left-handed conformation (Zhou & Ho, 1990). We had previously shown that the ion concentrations in solutions that yield Z-DNA crystals of hexanucleotide duplexes are related to the ability of these sequence to adopt the Z-conformation in solution. This relationship thus allows us to systematically test predictions for how sequence modifications affect the ability of hexanucleotides to form Z-DNA, and, in fact, has been useful in this laboratory to predict how to crystallize a particular sequence as Z-DNA (Ho et al., 1991). The self-complementary hexamer sequence d(m<sup>5</sup>CGUAm<sup>5</sup>CG) was crystallized as Z-DNA under significantly lower cation concentrations than the analogous d(TA) containing sequence.

The inability of d(TA) dinucleotides to adopt the Z-conformation was, from our analysis, predicted to be only partially related to the destabilizing effect of the thymine methyl group. The other major factor was observed to be the lack of an amino group at the C2 position of the adenine base, thus rendering the minor groove crevice of Z-DNA less hydrophilic than that of d(CG) dinucleotides. In the present study, we first demonstrate experimentally that the methylation of cytosine stabilizes Z- versus B-DNA by affecting the hydrophobicity of the two conformations. We then focus on the



contribution of other substituent groups, particularly the N2 amino group in the minor groove of the purine bases, on the thermodynamic stability of Z-DNA. Using the results of these studies, we have derived a thermodynamic cycle which describes the stability of Z-DNA as the base substituents of alternating pyrimidine-purine (APP) dinucleotides are systematically substituted to evolve d(m<sup>5</sup>CG) to d(TA) and back again.

## Experimental Methods

Salt titrations of poly(dG-dC) and poly(dG-dm<sup>5</sup>C). The polynucleotides poly(dG-dC) and poly(dG-dm<sup>5</sup>C) were obtained from Pharmacia. Titrations were performed by adding polynucleotide to various concentrations of MgCl<sub>2</sub>, LiCl, NaCl, and KCl solutions to give approximately 1 O.D. at 260 nm. The solutions were heated to 60° C for 10 minutes to facilitate formation of Z-DNA (Behe & Felsenfeld, 1981). Spectra were recorded at room temperature on an HP8452 diode array spectrophotometer. The formation of Z-DNA at each salt concentration was determined by monitoring the ratio of absorbance at 294 nm versus 260 nm. B-DNA has an absorbance ratio of 0.15 to 0.2 for poly(dG-dC) and 0.25 to 0.4 for poly(dG-dm<sup>5</sup>C), while Z-DNA has a ratio of 0.35 to 0.5 for poly(dG-dC) and 0.35 to 0.7 for poly(dG-dm<sup>5</sup>C), depending on the salt.

Solvent-accessible surface and solvent free energy calculations. The general methods for calculating the surfaces of DNA structures that are exposed to solvent (solvent accessible surfaces, or SASS) and the free energies for solvating these surfaces (solvent free energies, or SFEs) were previously described (Kagawa et al., 1989). The method involves first building models for hexanucleotide sequences in either the B- or the Z-conformations. The atomic coordinates of sequences as B-DNA were generated using standard helical parameters for B-DNA. The atomic coordinates of these same sequences as Z-DNA were generated from the crystal structures of previously crystallized sequences. Models of d(TA') containing sequences, where A' is an aminated adenine at the C2 position, were constructed by adding an sp<sup>2</sup> amino group at the C2 carbon of an adenine base using standard distances and geometries. Sequences containing d(CI) base pairs, where I is an inosine base, were constructed by removing this same amino group from the guanine of a d(CG) base pair. The d(UA) and d(UA') containing sequences were constructed by demethylating d(TA) and d(TA') base pairs, respectively. In specific cases where these bases have been crystallized, we have incorporated the

conformations from single crystal structures into our models. Thus, for the d(TA') and d(UA') containing sequences, we can compare the simplest model for these base pairs in Z-DNA, as generated by adding or removing substituents from the naturally occurring bases, to the actual conformations of these base pairs in the crystal structures to assess their effects on Z-DNA stability.

The calculation of the solvent free energies (SFEs) of each DNA structure requires first a calculation of the solvent accessible surfaces (SASs) for each sequence in both the B- and the Z-conformations. The SASs of the internal four base pairs of each hexanucleotide sequence were calculated using the Connolly rolling ball method (Connolly, 1983). The SFEs of these models were calculated as previously described. In short, each surface type was converted to a free energy for hydration by applying an atomic solvation parameter (ASP) that describes the energy required to transfer that surface type from an organic phase to an aqueous solvent phase (see Table 3.1). The total free energy of hydration ( $\Delta G^{\circ}_H$ ) for each hexamer model was calculated using Equation 1 (Eisenberg & MacLachlan, 1986), where  $SAS_i$  is the solvent accessible surface for each surface type  $i$ , and  $ASP_i$  is the atomic solvation parameter for that surface.

$$\Delta G^{\circ}_H = \sum_i (SAS_i \times ASP_i) \quad \text{Eq. 1.}$$

The two base pairs at either end of the hexamers were not included in the calculations to eliminate possible artifacts due to end effects. The total  $\Delta G^{\circ}_H$  was then divided by 2 to obtain the average SFE for the dinucleotide.

The differences in the hydration free energy between Z- and B-DNA ( $\Delta\Delta G^{\circ}H(Z-B)$  or solvent free energy (SFE)) for the previously studied d(m<sup>5</sup>CG), d(CG), d(CA)•d(TG), and d(TA) dinucleotides are linearly related to the B- to Z-DNA transition free energies ( $\Delta G^{\circ}T(B-Z)$ ) by Equation 2 (Kagawa et al., 1989).

$$\Delta\Delta G^{\circ}H(Z-B) = 0.71\Delta G^{\circ}T(B-Z) - 0.464 \quad \text{Eq. 2}$$

Crystallization solutions for Z-DNA hexanucleotides. In general, the most variable component in the crystallization of Z-DNA are the types and concentrations of the cations in the solutions. To normalize all the different types of solutions, we converted the various crystallization conditions for Z-DNA to a measure of the effective cation concentration (the cation strength) in each solution. The cation strengths ( $CS$ ) were estimated as the sum of the concentrations of cation added to the crystallization solutions ( $[cation]$ ), upon equilibration against the precipitant in the reservoir, corrected for the effective charge of each cation species ( $Z^2$ ) (*i.e.*,  $CS = \sum Z^2[cation]$ , as previously described (Kagawa et al., 1989).

To determine whether the calculated transition free energies were related to the actual driving force required for inducing a B- to Z-DNA transition, we compared  $\Delta G^{\circ}T(B-Z)$  calculated from Eq. 2 to the cation strengths required to crystallize hexanucleotides as Z-DNA. The  $\log_{10}$  of the  $CS$  values ( $\log CS$ ) were compared to the  $\Delta G^{\circ}T(B-Z)$  of each sequence. With the exception of the d(CICGCG) sequence, the crystallization conditions for the various hexanucleotides used in this study were those published in the original papers (see Table 3.4). The d(CICGCG) sequence was crystallized from a solution containing 2.0 mM DNA, 33 mM sodium cacodylate buffer at pH 7.0, 0.2 M  $MgCl_2$ , and 10% 2-methyl-2,4-pentanediol (MPD) equilibrated against a reservoir of 50% MPD.

Table 3.1. Atomic solvation parameters (ASP) of hydrophilic, hydrophobic, and charged phosphate surfaces in nucleic acids. The parameters were derived from the partition coefficients and calculated solvent-accessible surfaces of small organic molecules.

group	surface type	ASP (kcal mol <sup>-1</sup> Å <sup>-2</sup> )
base	hydrophobic (C)	0.034
	methyl (C)	0.043
	hydrophilic (O/N)	-0.068*
ribose	hydrophobic (C)	0.043
	hydrophilic (O)	-0.038
phosphate	charged (O/P)	-0.100

\* This value differs slightly from that reported by Kagawa, et al.(1989).

## Results

Our previous analysis of the hydrated surfaces of DNA duplexes suggested that the effect of base substituent groups on the relative stability of left-handed Z-DNA versus right-handed B-DNA is due to differences in the hydrophobicity of the solvent accessible surfaces (SAS) of the DNA structures. In this analysis, these base substituent effects are quantitated by converting the SAS values to solvent free energies (SFE). Here, we present experimental evidence that stabilization of Z-DNA by methylating cytosine bases is associated with differences the hydrophobicity of the DNA conformations. In addition, we have extended the surface and hydration analyses to the minor groove substituents of the DNA bases to study the role of the N2 amino substituent group of purine bases in stabilizing the minor groove crevice of Z-DNA. Our previous studies (Kagawa et al., 1989; Zhou & Ho, 1990) suggested that the inability of d(TA) dinucleotides to adopt the Z-conformation was partially related to the increased hydrophobicity of the major groove surface of Z-DNA associated with the C5 methyl group of the thymine bases. In this study, we focus on the role of the amino group of the purine bases in stabilizing Z-DNA by comparing the hydration free energies of d(TA') and d(CI) base pairs, where A' is adenine aminated at the C2 carbon and I is inosine, in their B- and Z-conformations. For these two sequences, we address the question of how adding an amino group to an adenine base pair affects the ability of d(TA), and how removing the amino group from guanine would affect the ability of d(CG) to adopt the Z-conformation. Using the results from our current and previous studies, we have constructed a thermodynamic scheme that describes the contribution of each nucleotide base substituent group to the stability of Z- versus B-DNA.

The Hofmeister hydrophobicity cation series and the B- to Z-DNA transition. One of the accepted measures of hydrophobicity is to monitor the influence of certain cations and anions on a process or transition. For cations, it has been shown that the

hydrophobic effect follows the Hofmeister series  $\text{Mg}^{+2} > \text{Li}^+ > \text{Na}^+ > \text{K}^+ > \text{NH}_4^+$ . According to this series,  $\text{Mg}^{+2}$  and  $\text{Li}^+$  would show the most pronounced hydrophobic effect, while  $\text{K}^+$  and  $\text{NH}_4^+$  deemphasize the effect (Melander & Horvath, 1977). We can use this series to test our hypothesis that substituent groups of the DNA bases stabilize or destabilize Z-DNA by either decreasing or increasing the hydrophobicity of the Z- and B-conformations.

To determine whether the Z-DNA stabilizing effect of methylating cytosine bases at the C5 position is related to hydrophobicity, we monitored the B- to Z-DNA transition for poly(dG-dC) to poly(dG-dm<sup>5</sup>C) as induced by cations of the Hofmeister series (Figure 3.1). If indeed the methyl group is primarily affecting the hydrophobicity of the DNA structures, there should be a dramatic difference in the amount of  $\text{Mg}^{+2}$  required to induce a B- to Z-DNA transition between the methylated and unmethylated polynucleotides. This difference should decrease as we proceed down the series, with  $\text{NH}_4^+$  showing very little if any difference between the two polynucleotides. The titration curves in Figure 3.1 show the effects of cations on the midpoint of the salt induced transition from B- to Z-DNA. The largest difference was observed for  $\text{Mg}^{+2}$  (> 1000-fold difference for the midpoints of the titrations), followed by  $\text{Li}^+$  (> 10-fold difference). Titrations with  $\text{Na}^+$  and  $\text{K}^+$  showed nearly identical differences (~4-fold differences) in their midpoints. Titrations with  $\text{NH}_4^+$ , which are not included in Figure 3.1, were again nearly identical to  $\text{Na}^+$  and  $\text{K}^+$ , with an approximate 5-fold difference for the unmethylated versus the methylated sequences. The decreasing differences going from  $\text{Mg}^{+2}$  to  $\text{Li}^+$  to  $\text{Na}^+$  suggest that there is a significant hydrophobicity component to the Z-DNA stabilizing effect of cytosine methylation, as we had previously suggested (Ho et al., 1988). The near identical behaviour of  $\text{Na}^+$ ,  $\text{K}^+$  and  $\text{NH}_4^+$  on the B- to Z-DNA transition indicates that there are additional stabilizing effects, such as base stacking or electrostatic interactions, of the cytosine methyl group on Z-DNA stability.

Figure 3.1. Titrations of poly(dG-dC) and poly(dG-dm<sup>5</sup>C) with MgCl<sub>2</sub>, LiCl, NaCl, and KCl. The percent of Z-DNA induced by titration of polynucleotides were measured by monitoring the ratio of absorbance at 294 nm versus 260 nm as a function of the salt concentrations. Interpolated curves are drawn to facilitate analysis of the results, and are not fit using any theoretical model. Titrations of poly(dG-dC) are shown as closed symbols and are traced with dotted lines, while those of its methylated analogue poly(dG-dm<sup>5</sup>C) are shown as open symbols are traced using a solid line. The double-headed arrows indicate the difference between the approximate midpoint of titrations with each salt for methylated and unmethylated polynucleotides. The squares represent titrations with MgCl<sub>2</sub>, diamonds with NaCl, circles with KCl, and triangles with LiCl.



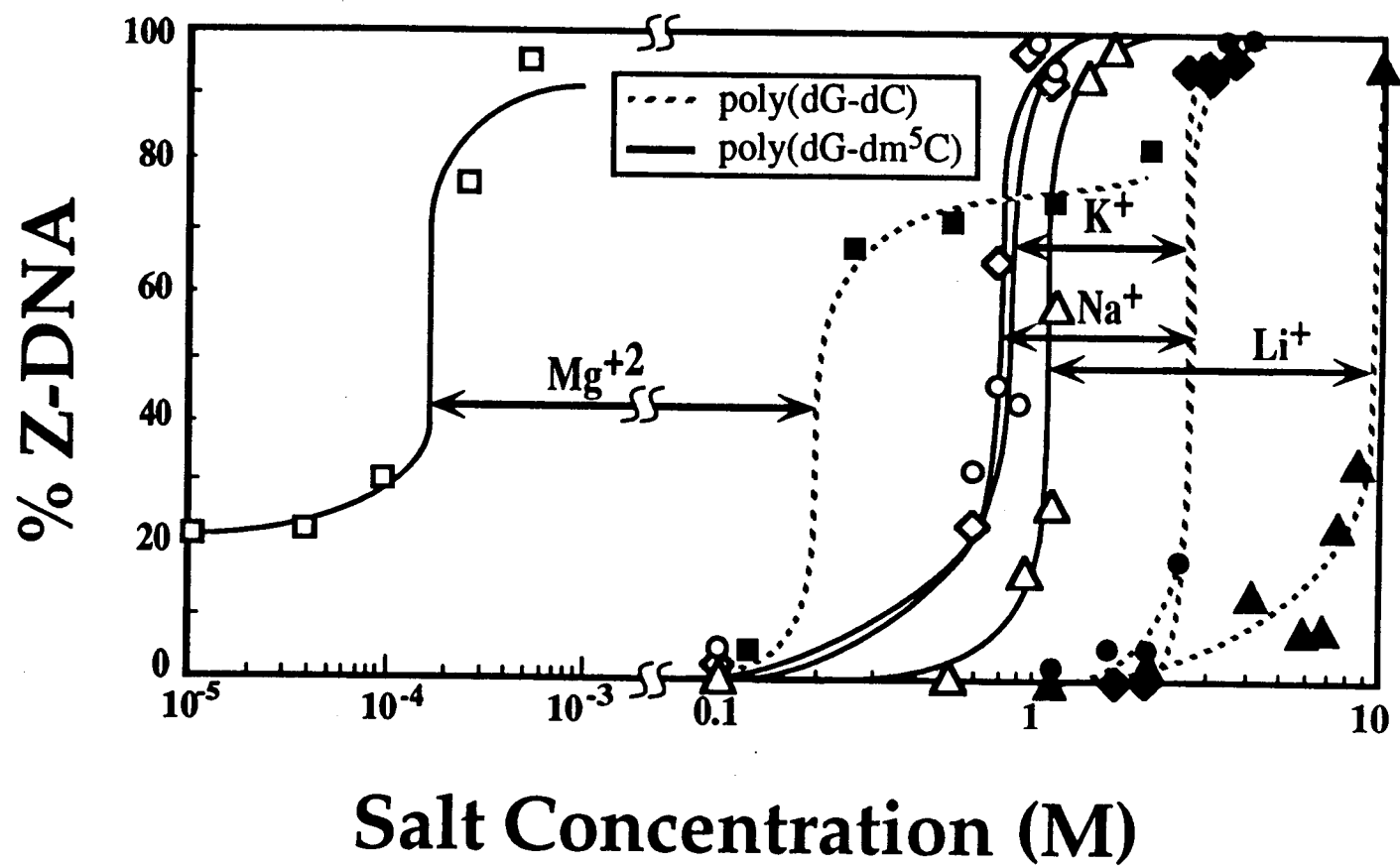


Figure 3.1

Stability of d(TA') dinucleotides in Z-DNA. Adding an N2 amino group to the C2 carbon of the adenine base in a d(TA) base pair generates a d(TA') base pair. The net effect is to render the minor groove of both B- and Z-DNA more hydrophilic. The effect on the relative solvent free energies of the two conformations, however, depends as much on the surface types which are lost as those which are gained by adding this N2 amino group. The changes in the SAS of d(TA) and d(TA') dinucleotides are compared for B- and Z-DNA in Table 3.2. In general, this amino group has a greater effect on the stability of the Z-conformer as compared to the B-form.

In both the B- and the Z-conformations, the C2 carbon becomes entirely inaccessible to solvent with addition of the N2 amino group; the accessible carbon surfaces of the aromatic bases are therefore reduced for both B- and Z-DNA (Table 3.2). There is a greater reduction in the exposure of base carbons in B-DNA ( $>17 \text{ \AA}^2$ ) as compared to Z-DNA ( $\sim 12.6 \text{ \AA}^2$ ) in going from d(TA) to d(TA'). Similarly, the exposure of the added amino group is greater for B-DNA ( $26.0 \text{ \AA}^2$ ) than for Z-DNA ( $21.5 \text{ \AA}^2$ ). These results reflect the greater exposure of the C2 position of the purine base to solvent in the minor groove of the B- versus Z-conformation and, in themselves, would suggest that adding the N2 amino to adenine bases should destabilize Z-DNA. However, if we include in our comparisons the additional surfaces that are buried, we see that neighboring hydrophilic atoms also become less accessible in the d(TA') dinucleotides. For B-DNA, there is a total loss of  $>11 \text{ \AA}^2$  of base oxygen and nitrogen surfaces with the addition of the N2 amino. In the case of Z-DNA, the loss of hydrophilic surface is only  $1.3 \text{ \AA}^2$ . In addition, the added N2 amino of the A' base is stacked directly above the ribose sugar of an adjacent base pair in Z-DNA. This greatly reduces the accessible surfaces, particularly of the C1' and C2' carbons, of the ribose (Figure 3.2). In total, the

Table 3.2. Solvent accessible surface areas ( $\text{\AA}^2$ ) of alternating pyrimidine-purine sequences (Py-Pu Seq) as B and Z-DNA\*.

Conf. (B/Z)	Py-Pu Seq.	Nucleotide Base					Ribose-Phosphate			TOTAL
		C	N	O	N2	C5 Methyl	C	O	P	
B	TA	43.6	55.6	32.6	NA	44.8	182.8	51.4	132.8	543.8
Z	TA	46.0	50.8	27.0	NA	46.2	188.2	42.0	133.6	533.8
B	TA'	28.8	46.8	29.8	26.0	44.8	183.8	43.8	132.8	536.6
Z	TA'	33.3	49.7	27.4	21.5	46.3	170.9	41.8	133.2	524.1
B	CG	49.4	59.1	31.0	23.6	NA	185.5	47.2	132.6	528.4
Z	CG	56.3	48.7	44.2	19.6	NA	184.0	47.4	132.1	532.4
B	CI	64.8	65.2	38.0	NA	NA	197.7	52.3	132.6	538.6
Z	CI	71.4	48.8	47.2	NA	NA	199.4	46.8	133.0	546.2
B	UA	63.2	57.6	39.2	NA	NA	190.0	51.4	133.6	534.6
Z	UA	68.6	57.0	37.8	NA	NA	194.6	42.0	133.8	534.0
B	UA' <sup>1</sup>	54.0	45.8	34.0	28.2	NA	197.4	47.2	124.2	530.6
Z	UA' <sup>1</sup>	60.0	54.4	33.6	20.6	NA	193.0	40.0	130.8	540.2
B	UA' <sup>2</sup>	45.6	45.8	33.4	29.2	NA	192.8	42.2	133.6	522.4
Z	UA' <sup>2</sup>	56.2	57.0	36.4	18.4	NA	182.8	41.8	133.0	525.6

\*Surfaces were calculated for the four internal base pairs of model hexanucleotides of alternating pyrimidine-purine sequence in the table. The total for each surface type are therefore reported for four base pairs (or two dinucleotides). The corresponding values for a single dinucleotide would be half of those in the table.

<sup>1</sup>Calculated from models where the UA' base pairs were constructed by simply removing the methyl from the thymine and adding an amino to the adenine bases of a d(TA) base pair.

<sup>2</sup>Calculated from models constructed using the atomic coordinates of d(UA') dinucleotides of the Z-DNA structure d(m<sup>5</sup>CGUA'm<sup>5</sup>CG) (Schneider et al., 1992).

Figure 3.2. Solvent-accessible surfaces of d(TA) and d(TA') dinucleotides as B- and Z-DNA. The dinucleotides are viewed down the helical axis. The *anti-syn* over *syn-anti* stacking of the base pairs are shown for the dinucleotides in Z-DNA. Hydrogens have been omitted from this figure for clarity, although they were included for the surface calculations. The open dots represent hydrophilic contact points with water, while the filled dots represent hydrophobic contact points. SAS were calculated using a probe radius of 1.45 Å and a dot density of 10 dots Å<sup>-2</sup> (5 dots Å<sup>-2</sup> are shown for clarity).

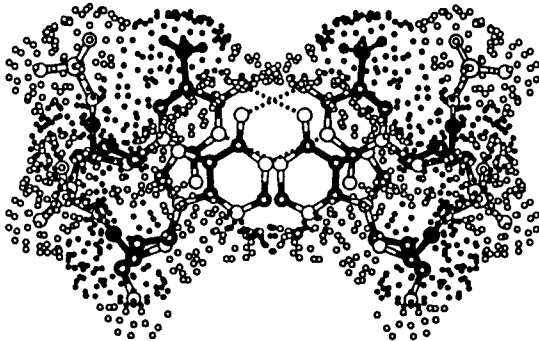
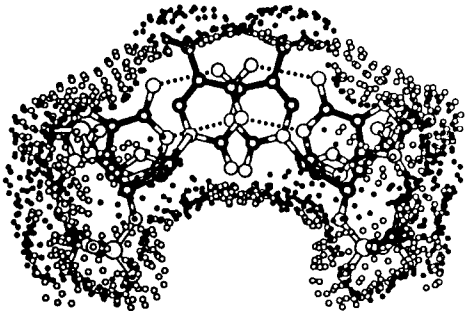
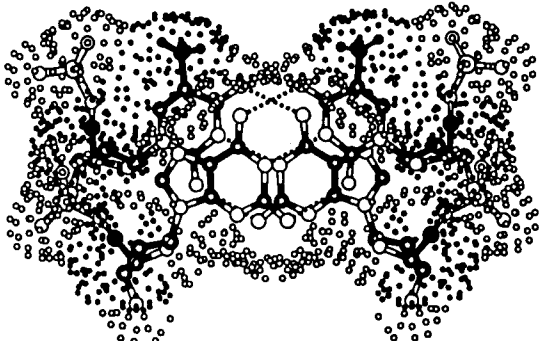
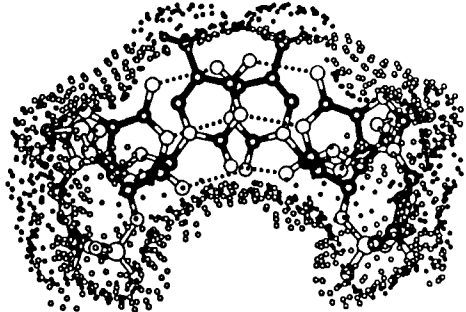
	B-DNA	Z-DNA
$d(TA)$		
$d(TA')$		

Figure 3.2

surface of B-DNA becomes only slightly more hydrophilic (with a net loss of  $17.4 \text{ \AA}^2$  of exposed hydrophobic surface and only a  $6.8 \text{ \AA}^2$  net increase in hydrophilic surface), while for the Z-conformer, the loss of hydrophobic surface is  $30 \text{ \AA}^2$ , with a concomitant gain of  $20.2 \text{ \AA}^2$  of hydrophilic surface.

When these changes in SASs are translated into SFE values (Table 3.3), we see that there is a  $1.35 \text{ kcal mol}^{-1}$  per dinucleotide ( $\text{kcal mol}^{-1} \text{dn}^{-1}$ ) difference in the hydration free energy of Z versus B-DNA ( $\Delta\Delta G^\circ \text{H}(\text{Z-B})$ ) in the case of the d(TA) dinucleotide, while for the d(TA') dinucleotide,  $\Delta\Delta G^\circ \text{H}(\text{Z-B}) \cong 0$ . Using the relationship between  $\Delta\Delta G^\circ \text{H}(\text{Z-B})$  and the B- to Z-DNA transition free energy ( $\Delta G^\circ \text{T}(\text{B-Z})$ ) (Kagawa et al., 1989), this translates to  $1.01 \text{ kcal mol}^{-1} \text{dn}^{-1}$  difference in stability of Z- versus B-DNA for the d(TA') dinucleotides, as compared to the  $2.4 \text{ kcal mol}^{-1} \text{dn}^{-1}$  for d(TA). Thus mutation of d(TA) dinucleotides to d(TA') would have an overall effect of stabilizing the Z-conformation. The extent of this stabilization was greater than expected. The SAS of the d(TA') dinucleotide is overall slightly less hydrophilic compared to the d(CG), but this was true for both the B- and the Z-conformations. We therefore would predict that, compared to the naturally occurring APP dinucleotides of d(CG), d(CA)•d(TG), and d(TA), the d(TA') dinucleotide would have a propensity to adopt the Z-conformation that is comparable to that of the d(CG).

When only a single d(TA) base pair of the dinucleotide is converted to a d(TA'), the effect is even more dramatic. Comparing the areas of d(CPu)•d(PyG) dinucleotides (where Pu is the purine nucleotide G, A, or A', and Py is the pyrimidine base C or T), we calculated  $\Delta G^\circ \text{T}(\text{B-Z}) = 0.08 \text{ kcal mol}^{-1} \text{dn}^{-1}$  for the d(CA')•d(TG) dinucleotide, which is lower than we would expect for the average between the d(CG) and a d(TA') dinucleotides. Thus, the effect of each substituent on the stability of a dinucleotide in Z-DNA cannot be considered as simply the sum of the  $\Delta G^\circ \text{T}(\text{B-Z})$  for the two base pairs that

Table 3.3 Calculated Solvation Free Energies (SFEs) and the corresponding B- to Z-DNA transition free energies ( $\Delta G^\circ T(B-Z)$ ) of dinucleotide sequences.

Dinucleotide Sequence	SFE (kcal mol <sup>-1</sup> dn <sup>-1</sup> )			$\Delta G^\circ T(B-Z)^*$ (kcal mol <sup>-1</sup> dn <sup>-1</sup> )
	$\Delta G^\circ H(B)$	$\Delta G^\circ H(Z)$	$\Delta \Delta G^\circ H(Z-B)$	
d(m <sup>5</sup> CG)	-10.49	-11.36	-0.87	-0.45
d(CG)	-13.05	-12.76	0.29	1.07
d(CA)•d(TG)	-11.67	-11.34	0.33	1.12
d(TA)	-9.90	-8.53	1.35	2.45
d(UA)	-11.50	-10.64	0.86	1.81
d(CI)	-12.01	-10.52	1.50	2.64
d(TA')	-10.80	-10.56	0.25	1.01
d(CI)•d(CG)	-12.51	-11.69	0.82	1.76
d(CA')•d(TG)	-11.87	-12.34	-0.47	0.08
d(UA')	-11.15	-11.87	-0.72	-0.25

\*Calculated using the relationship  $\Delta G^\circ T(B-Z) = 1.30 \times \Delta \Delta G^\circ H(Z-B) + 0.69$  kcal mol<sup>-1</sup> dn<sup>-1</sup>, derived from the relationship analogous to that in Kagawa, et al (1989).

form the dinucleotide, as has been suggested (Anshelevich et al., 1988). A comparison of the predicted  $\Delta G^\circ T(B-Z)$  values in Table 3.2 suggests that the dinucleotide must be considered as the minimum unique repeat unit for Z-DNA in terms of the thermodynamic stability.

We have previously observed a correlation between the cation strengths (CS) required to crystallize a series of isomorphous hexamer duplex structures as Z-DNA and  $\Delta G^\circ T(B-Z)$  calculated for these sequences (Ho et al., 1991). Our spectroscopic studies on Z-DNA formation under solutions for crystallization showed that the conditions for crystallization of hexanucleotides as Z-DNA were also conditions that stabilize the left-handed conformation in solution (Ho et al., 1991). Thus, the Z-DNA crystallization conditions for hexamer sequences mirror, and are indicative of the ability of these sequences to

adopt the Z-conformation. The incremental increase in the  $\Delta G^{\circ}T(B-Z)$  expected for introducing a d(TA') dinucleotide into a Z-DNA forming sequence was predicted from the SAS analysis to be  $\sim 1.01 \text{ kcal mol}^{-1}\text{-dn}^{-1}$ . This suggests that replacing a single d(CG) dinucleotide by d(TA') in a hexanucleotide sequence would not significantly affect the cation strength required to crystallize this sequence as Z-DNA. Alternatively, replacing a d(TA) dinucleotide with d(TA') would lower the salt requirement for crystallization. Furthermore, the even lower  $\Delta G^{\circ}T(B-Z)$  of  $0.08 \text{ kcal mol}^{-1}\text{-dn}^{-1}$  predicted for d(CA')•d(TG) would suggest that a sequence that contains this dinucleotide would require less salt to crystallize than an analogous d(CG) containing sequence.

Coll et al. (1986) obtained single crystals of the self-complimentary sequence d(CGTA'CG), which was disordered but presumed to be in the Z-conformation, and of d(CA'CGTG), which was indeed in the Z-conformation and isomorphous to other Z-DNA hexanucleotide crystals. Table 3.4 compares the crystallization conditions published for these two hexanucleotides, and also compares the Z-DNA crystallization conditions and the calculated  $\Delta G^{\circ}T(B-Z)$  for other well studied hexamer sequences. The cation strength required to crystallize d(CA'CGTG) was indeed similar to that for d(CG)<sub>3</sub>



Table 3.4. Comparison of the calculated B- to Z-DNA transition free energies ( $\Delta G^{\circ}T(B-Z)$ ) to cation strength ( $CS$ ) in solutions that yield crystals of hexanucleotide sequences as Z-DNA.

Hexanucleotide Sequence	$\Delta G^{\circ}T(B-Z)^*$ (kcal mol <sup>-1</sup> dn <sup>-1</sup> )	$CS^{**}$ (M)	$\log(CS)$	Reference
d(m <sup>5</sup> CG) <sub>3</sub>	-0.45	0.23	-0.64	Fujii et al., 1982
d(m <sup>5</sup> CGUAm <sup>5</sup> CG)	0.31	0.36	-0.44	Zhou & Ho, 1990
d(m <sup>5</sup> CGTAm <sup>5</sup> CG)	0.55	0.56	-0.25	Wang et al., 1984
d(CA'CGTG)	0.41	0.66	-0.18	Coll et al., 1986
d(CGUA'CG)	0.63	0.38	-0.42	Schneider et al., 1992
d(CGTA'CG) <sup>***</sup>	1.05	0.80	-0.10	Coll et al., 1986
d(CG) <sub>3</sub>	1.07	2.19	0.34	Wang et al., 1979
d(CACGTG)	1.09	3.24	0.51	Coll et al., 1979
d(CICGCG)	1.52	4.17	0.62	This work

\*  $\Delta G^{\circ}T(B-Z)$  calculated as the average across the three dinucleotides of the hexamer sequence (units are in kcal/mol per dinucleotide = kcal mol<sup>-1</sup> dn<sup>-1</sup>).

\*\*  $CS$  estimated at equilibrium from crystallization solutions by the equation  $CS = \sum Z^2[\text{cation}]$  (where  $Z$  is the charge of the cation).

\*\*\* This structure was crystallized in a disordered crystal lattice, but presumed to be in the Z-conformation.

Figure 3.3. Comparison of the log of the cation strength ( $\log CS$ ) for crystallization of Z-DNA hexanucleotides to the calculated B- to Z-DNA transition free energies ( $\Delta G^{\circ}T(B-Z)$ ). Open squares are for the hexanucleotide sequences  $d(m^5CG)_3$ ,  $d(m^5CGTAm^5CG)$ ,  $d(m^5CGUAm^5CG)$ ,  $d(CG)_3$ , and  $d(CACGTG)$  from earlier studies (see references in Table 3.4). The diamonds represent sequences for this study. Closed diamonds are for the sequences  $d(CA'CGTG)$ ,  $d(CGTA'CG)$ , and  $d(CICGCG)$ . The open diamond is for the sequence  $d(CGUA'CG)$ . The cation strength ( $CS$ ) is estimated from the crystallization conditions by the relationship  $CS = \Sigma Z^2[\text{cation}]$ , and where  $Z$  is the charge of the cation. The line represents a linear least square fit to the data, giving a relationship:  $\log CS = 0.74\Delta G^{\circ}T(B-Z) - 0.47$  ( $R = 0.93$ ).

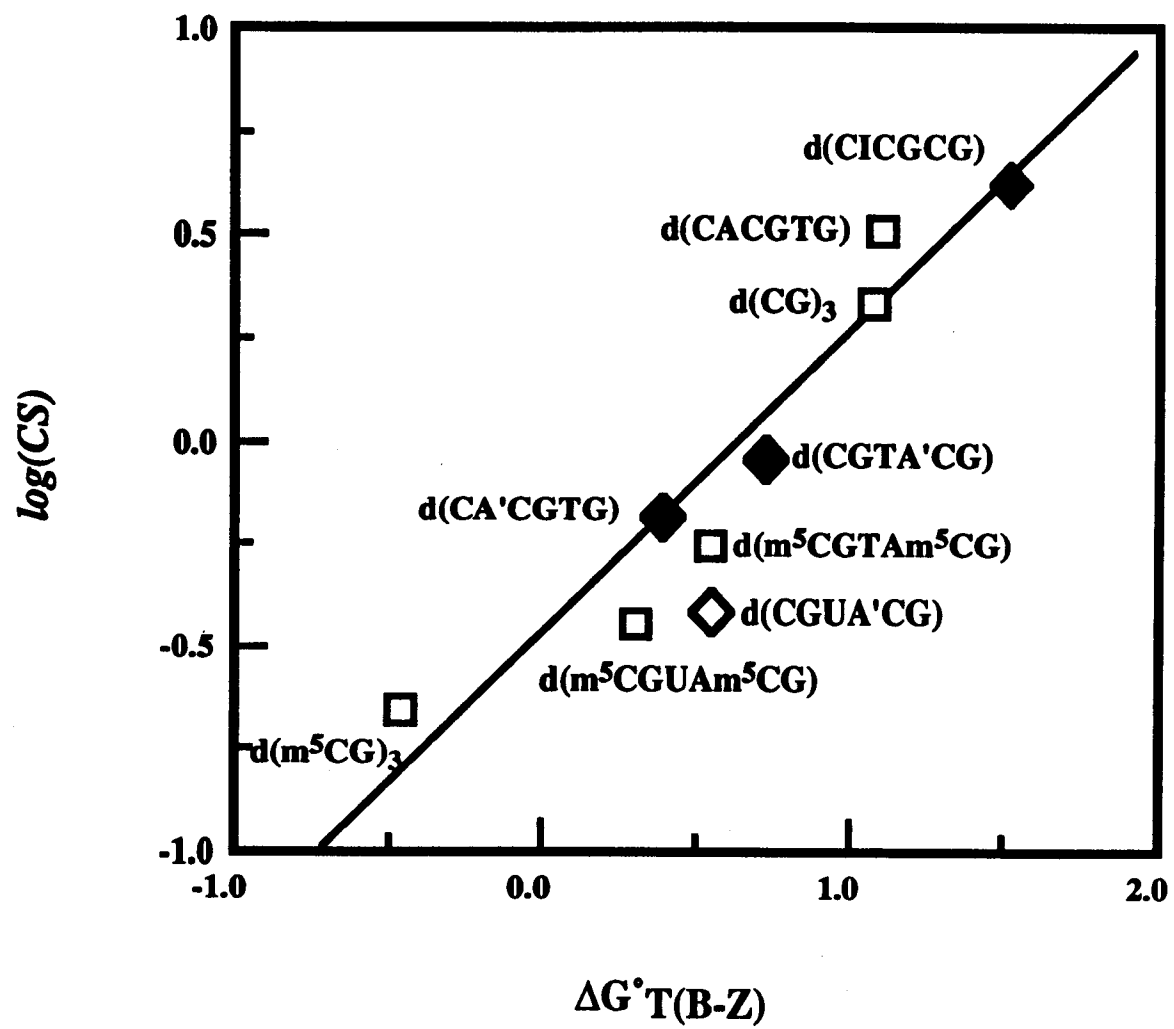


Figure 3.3

and lower than that for d(CACGTG), as was predicted from the SFE calculations (Figure 3.3). The sequence d(CGTA'CG) also behaved as predicted, requiring less salt for crystallization than even d(CG)<sub>3</sub>.

The structure of the d(CA'CGTG) sequence provides a molecular test for some aspects of this hydration model (Coll et al., 1986). Of particular interest was the observation that a continuous spine of water molecules was located in the minor groove crevice of the d(TA') containing hexanucleotide. Similar sets of waters were observed in the crystal structures of d(CGCGCG) and d(m<sup>5</sup>CGm<sup>5</sup>CGm<sup>5</sup>CG), suggesting that these help to stabilize Z-DNA conformation. The d(TA) base pairs of the d(m<sup>5</sup>CGTAm<sup>5</sup>CG) and d(CACGTG) sequences, however, disrupt this spine of water. Our results are consistent with these observations, and would have predicted this difference in the interaction of solvent molecules in the minor groove crevice of Z-DNA.

Stability of d(CI) dinucleotides in Z-DNA. The N2 amino groups of guanine bases render the minor grooves of both B- and Z-DNA very hydrophilic (Figure 3.4). As we would suspect from extrapolation of the results from the analysis of d(TA) and d(TA'), removing this amino group from the minor groove of a d(CG) base pair would greatly reduce the stability of Z-DNA. The SAS calculations show that there is a slight increase of the hydrophobic surfaces of the bases for both B- and Z-DNA due to the exposure of the C2 carbon of the purine base, but, again, it is the now increased exposure of other neighboring hydrophilic groups of the B-DNA bases that helps to destabilize the Z-form (Figure 3.4 and Table 3.2).

These calculations show that the N2 amino group, when removed, would destabilize the Z-form much more than the B-conformer. The overall loss in stability is observed as a  $\Delta\Delta G^\circ_{\text{H}}(\text{Z-B}) = 1.5 \text{ kcal mol}^{-1}\text{-dn}^{-1}$ , or a  $2.64 \text{ kcal mol}^{-1}\text{-dn}^{-1}$  for  $\Delta G^\circ_{\text{T}}(\text{B-Z})$  (Table 3.3). A comparison of the actual Z-DNA crystallization conditions for the sequence d(CICGCG) to those of other Z-DNA hexanucleotides and to the predicted cation strength show that indeed the d(CI) dinucleotide requires a dramatically higher salt

for crystallization as compared to other APP hexanucleotide sequences. Thus d(CI) greatly diminishes the stability of the left-handed conformation (Table 3.4). The degree of destabilization resulting from deamination of guanines is comparable to the stabilizing effect of aminating adenine bases. Consequently, the d(CI) dinucleotide is predicted to have the lowest propensity to form Z-DNA of any APP dinucleotide studied so far.

Figure 3.4. Solvent-accessible surfaces of d(CG) and d(CI) dinucleotides as B- and Z-DNA. The dinucleotides are viewed down the helical axis. The *anti-syn* over *syn-anti* stacking of the base pairs are shown for the dinucleotides in Z-DNA. Hydrogens have been omitted from this figure for clarity, although they were included for the surface calculations. The open dots represent hydrophilic contact points with water, while the filled dots represent hydrophobic contact points. SAS were calculated using a probe radius of 1.45 Å and a dot density of 10 dots Å<sup>-2</sup> (5 dots Å<sup>-2</sup> are shown for clarity).

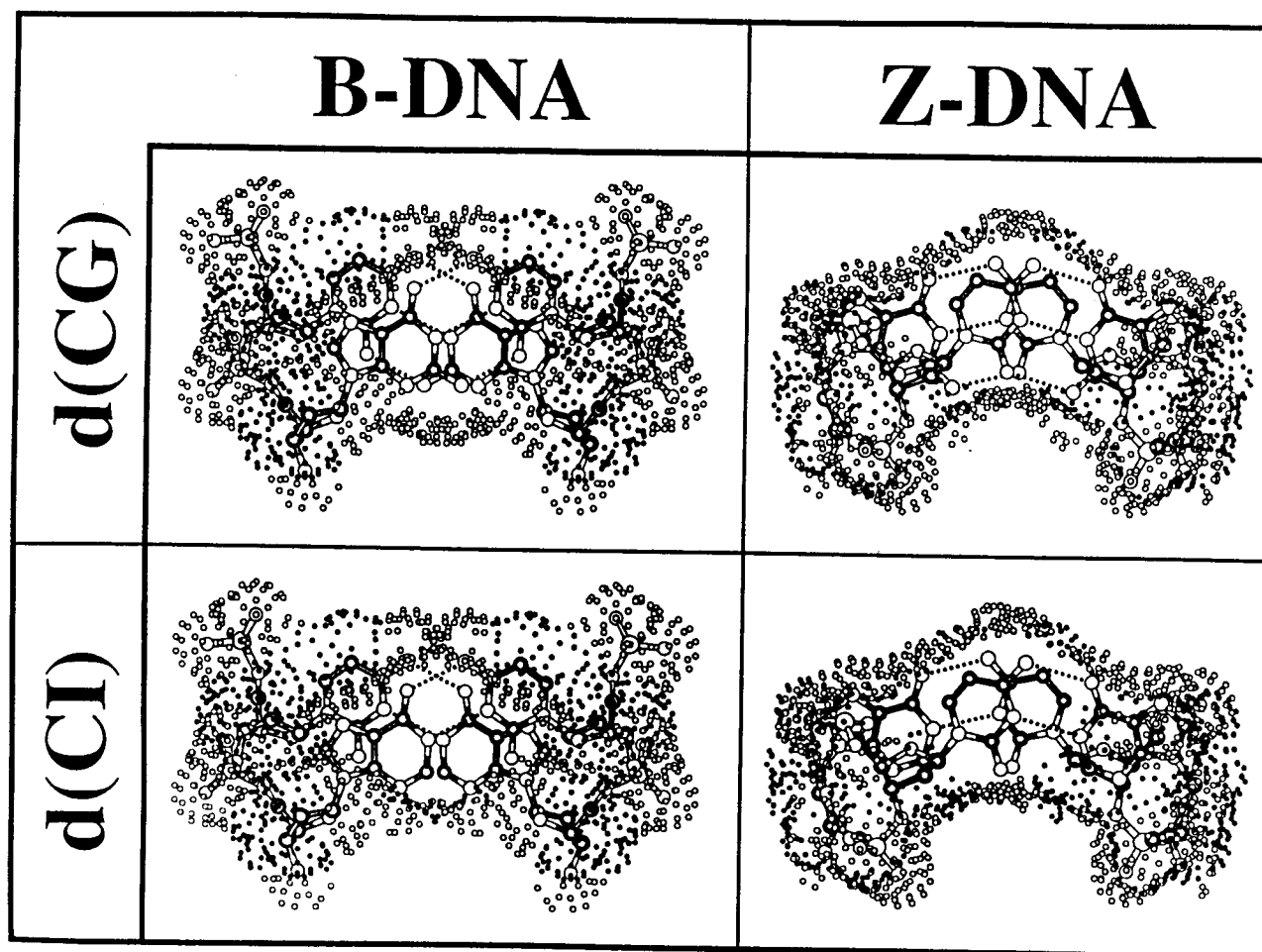


Figure 3.4

## Discussion

The energetics of DNA folding, as defined by a purely thermodynamic approach, have yet to be fully described in a satisfactory manner. Our work has focused on the contribution of the hydrophobic effect on the ability of various APP sequences to adopt the left-handed structure of Z-DNA. In the present studies, we bring together spectroscopic and crystal growth experiments, along with SFE calculations, to show that indeed there is a strong hydrophobic effect on the stability of Z-DNA and that SFE calculations can be used in a predictive manner to estimate the  $\Delta G^\circ T(B-Z)$  of APP sequences.

Our results from the titration of poly(dG-dC) and poly(dG-dm<sup>5</sup>C) with the cations of the lipotropic Hofmeister series show that hydrophobicity contributes to the Z-DNA stabilizing effect of cytosine methylation, as had been previously suggested (Ho et al., 1988). The results from these experiments, however, also demonstrate that there is an intrinsic ability of the methyl group to stabilize Z-DNA. This additional stabilization may arise from perturbations to the electrostatic properties of the nucleotides as suggested by Soumpasis, et al (1986), or by affecting the base stacking, as suggested from the original crystal structure of d(m<sup>5</sup>CG)<sub>3</sub> (Fujii et al., 1982). This is consistent with the relationship that we have derived between the calculated SFEs and  $\Delta G^\circ T(B-Z)$ , which showed that solvent interactions account for approximately 70% of the sequence dependence for Z-DNA formation.

We have extended this SFE analysis to other substituent groups of the DNA bases, and predicted that the N2 amino group is critically important in defining the stability of Z- versus B-DNA. The resulting estimates for  $\Delta G^\circ T(B-Z)$  faithfully predict the crystallization conditions, and thus the salt required to induced the B- to Z-DNA transition of d(CI) and d(TA') containing hexanucleotide sequences. The results from the current studies on the minor groove substituents and our previous work on methylation at



the major groove surface (Kagawa et al., 1989; Zhou & Ho, 1990) demonstrate that the approach of estimating SFEs from the solvent exposed surfaces of a DNA sequence is useful for predicting the relative abilities of various sequences to adopt the Z-DNA conformation. Thus, these studies support the significance of solvent interactions in DNA folding.

We have used the results of these analyses to construct a thermodynamic cycle that relates the effects of systematically adding, removing, or inverting the positions of the various substituent groups on the stability of APP dinucleotides as Z-DNA (Figure 3.5). Starting with the d(m<sup>5</sup>CG) dinucleotide, demethylation destabilizes Z-DNA by 1.5 kcal mol<sup>-1</sup>-dn<sup>-1</sup>. Removing the amine in the minor groove crevice of the guanine base to form the d(CI) dinucleotide has a more dramatic effect of destabilizing Z-DNA by 1.6 kcal mol<sup>-1</sup>-dn<sup>-1</sup>. If we now reverse the positions of the O6 keto oxygen of the guanine base and the N4 amino of the cytosine base to form a d(UA) dinucleotide, Z-DNA is stabilized by -0.83 kcal mol<sup>-1</sup>-dn<sup>-1</sup>. Methylation of the uridine bases generates the d(TA) dinucleotide which is now 0.64 kcal mol<sup>-1</sup>-dn<sup>-1</sup> less stable as Z-DNA than d(UA). Adding an amino group to the adenine base to give the d(TA') dinucleotide stabilizes Z-DNA by -1.4 kcal mol<sup>-1</sup>. Finally, we can again reverse the positions of the keto oxygen of the pyrimidine and the amine of the purine at the major groove surface to now complete the thermodynamic cycle, bringing us back to the original d(m<sup>5</sup>CG) dinucleotide, which is -1.5 kcal mol<sup>-1</sup>-dn<sup>-1</sup> more stable than the d(TA') dinucleotide.

Of the base pairs in this cycle, each has been incorporated in at least one sequence that has been crystallized as Z-DNA. The crystallization conditions can in fact be predicted for these hexanucleotides of APP sequences using the simple rules in this thermodynamic cycle. In addition, the free energies for the negative supercoil induced B- to Z-DNA transition have been determined for three of the dinucleotides in this cycle, and follow the rules d(m<sup>5</sup>CG) > d(CG) > d(TA) as predicted by this thermodynamic scheme.

One interesting observation from the scheme in Figure 3.5 is that the positions of the keto- and amino groups at the major groove surface have a greater effect on the stability of Z-DNA than expected. The positions of these two substituents specify whether C5 methylation of the pyrimidines stabilizes or destabilizes Z-DNA. When the keto oxygen is placed on the pyrimidine (U or T), the C5 methyl has a destabilizing effect, while an amino at the C-4 position gives the methyl group a stabilizing influence on Z-DNA. The hydration model suggests that the placement of these two substituents defines the pocket at the C5 position of the pyrimidine in Z-DNA as either overall hydrophilic or hydrophobic. In the case of the d(CG) dinucleotide, C5 methylation buries a hydrophobic pocket and thus stabilizes the Z-form. For the d(TA) dinucleotides, demethylation of the thymine exposes a more hydrophilic pocket and thus stabilizes the Z-form.

There are a number of possible pyrimidine-purine dinucleotides that are not represented in this thermodynamic cycle. The most obvious is d(CA)•d(TG). Our previous studies show that this dinucleotide would be as stable in the Z-conformation as would d(UA). The most interesting dinucleotide that is not in this figure is d(UA'). Considering the major groove and minor groove substituent effects that we observe here, this pair of unusual bases was predicted to be one of the most stable dinucleotides as Z-DNA. In a two state system in which the DNA duplex can adopt only the B- or the Z-conformations, alternating d(UA') would be expected to be even more stable in the Z-form than even d(CG), and may be comparable to d(m<sup>5</sup>CG) (Table 3.3). This is consistent with the low concentration of salt reported in the crystallization of the sequence d(CGUA'CG) (Schneider et al., 1992) (Table 3.4).

Obviously, there may be alternative structures of DNA, other than the B- or the Z-forms, that would affect the ability of sequences to adopt the Z-conformation. For example, polymeric d(UA), which should be as stable in the Z-form as a d(CA)<sub>n</sub>•d(TG)<sub>n</sub> polymer, has not been observed to form Z-DNA in solution. There appears to be an

alternative conformation (the yet undefined X-form) which d(UA) prefers under conditions that would normally induce other APP polymers to form Z-DNA (Vorlickova & Jaroslav, 1984). Whether there are any alternative structures for d(UA') that could effectively compete with Z-DNA is yet to be determined.

Figure 3.5. Thermodynamic cycle comparing the substituent effects at the major and minor grooves of the double helix on the stability of Z- versus B-DNA for alternating pyrimidine-purine dinucleotides. The effects of systematic additions or mutations of substituent groups, going from the most stable d(m<sup>5</sup>CG) dinucleotide to d(CG) to d(CI) to d(UA) to d(TA) to d(TA') and back to d(m<sup>5</sup>CG), on the relative stability of Z-DNA are shown. Atoms or groups in shaded spheres indicate positions that are added, removed, or mutated at each step in the cycle. The direction of the cycle was arbitrarily chosen, although each step should be treated as an equilibrium. Only one base pair of each dinucleotide is shown for clarity.

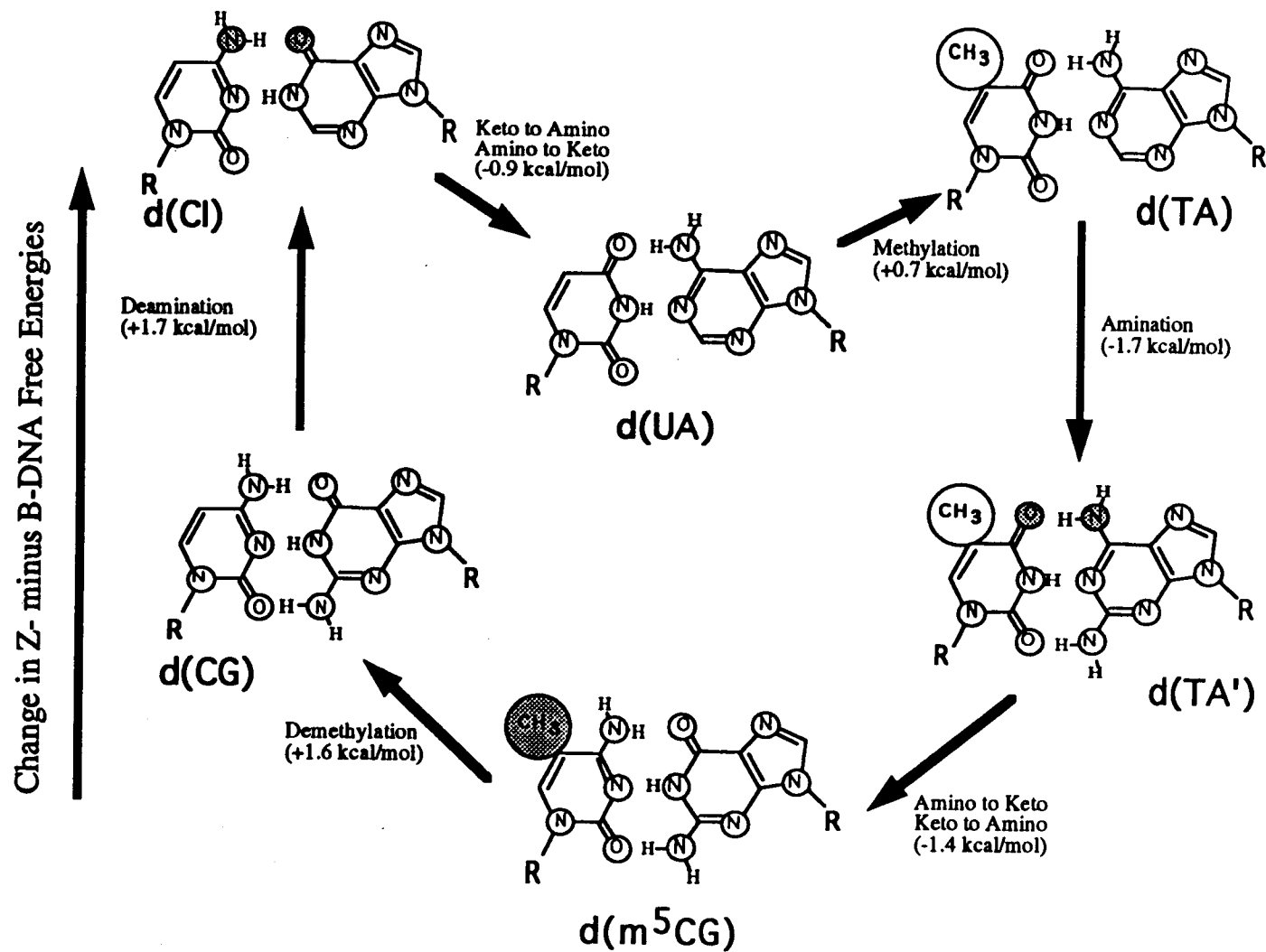


Figure 3.5

## Chapter 4

### The Contribution of the Hydrophobic Effect to Z-DNA Helix Packing: A Crystallographic Study

#### Introduction

X-ray diffraction studies on single crystals of macromolecules have provided detailed structural information which have contributed to the understanding of biological processes at the molecular level. With the advances in X-ray diffraction data collection technology, improved computational capabilities, and increased availability of pure material, X-ray crystallography has become a more routine tool for studying macromolecular structure. The availability of new structural information from single crystal X-ray diffraction studies is currently limited by the ability to grow diffraction quality crystals of macromolecules. The ability to produce crystals would be aided by a better understanding of the mechanisms involved in crystallizing biological macromolecules.

The complexities associated with macromolecular crystallization, as opposed to the crystallization of small organic molecules, are associated with the sensitivity of conformational stability and solubility to environmental conditions. Therefore, the successful crystallization of macromolecules requires the control of a large number of parameters. The crystallization of macromolecules are significantly influenced by physico-chemical parameters which define solution characteristics such as viscosity, pH, ionic environment, and temperature. These factors affect the characteristics of solution dynamics, as well as the conformational homogeneity of the sample. Recent advances in

the ability to obtain diffraction quality crystals have stemmed from the improved methodologies which allow better control of these parameters (reviewed in Geige et al., 1994).

In addition to improvements in crystal growth methods, studies on the interactions within the crystal lattice have contributed to the understanding of the crystallization of macromolecules. The internal order and diffracting capabilities of crystals of macromolecules are determined by the intermolecular contacts within the crystal lattice. These crystal packing interactions involve specific interfaces between molecules in the crystal lattice. Therefore, it should be possible to improve the crystallizability of a particular protein with strategic manipulation of the surface residues at these critical interfaces. An analysis of cytochrome c' and lysozyme crystal packings (Salemme et al., 1988) showed that stabilizing lattice interactions involve van der Waals, hydrogen bonding, and ionic bonding which occur at a few protein - protein contacts. These contacts within the crystal lattice were reported to occur frequently at surface loops and flexible side chains. The flexibility of these surface groups was thought to facilitate the formation of hydrogen bonding networks of ordered solvent which help to stabilize the lattice (Salemme et al., 1988). As with the protein crystals, the stabilization of the tRNA<sup>Asp</sup> crystal lattice involves both van der Waals, and specific or non-specific hydrogen bonding interactions. However, the primary contacts in this lattice have been characterized as hydrophobic in nature. In addition, these contacts in the tRNA<sup>Asp</sup> crystals occur between regions which have the greatest mobility, suggesting that structural adaptability was a favorable characteristic for these molecular interfaces (Moras & Bergdoll, 1988). The pliability of the structure at these interfaces for both protein and RNA crystals is presumed to allow the intermolecular interactions to be optimized in the crystal lattice (Salemme et al., 1988; Moras & Bergdoll, 1988).

Advances resulting from the manipulation of surface and internal residues of proteins have been reported for the crystallization of thymidylate synthase (McElroy et al., 1992), the H-chain of human ferritin (Lawson et al., 1991), and glutathione reductase (Mittl et al., 1994). The surface residue modifications reported in these studies have either facilitated crystal formation or improved the quality of the crystal. These modifications have affected the crystallization process by creating intermolecular metal complex sites (Lawson et al., 1991), introducing intermolecular hydrogen bonding (Mittl et al., 1994), or by modifying the solubility of the protein (McElroy et al., 1992). These studies indicate that strategic manipulation of the crystal packing interactions can aid in the crystallization of macromolecules.

Modifications of the crystal packing interactions are likely to influence the crystallization process with adjustments to the intrinsic solubility, structural stability, and the potential lattice interactions. The development of a rational approach for the improved crystallization of macromolecules would be aided with information concerning the structural and energetic consequences of these modifications on the process of crystallization. Toward this goal, the studies in this chapter were designed to examine the relative significance of van der Waals interactions versus solvent interactions in the process of macromolecular crystallization.



### Statement of purpose

The final chapter of the thesis describes experiments designed to study the relative significance of two competing crystal packing forces involved in determining the orientation of Z-DNA hexamer duplexes in the crystal lattice of the  $P2_12_12_1$  space group. The significance of these energetic terms was studied by monitoring the change in distribution of the duplex between the two distinct packing orientations in the lattice of a non-selfcomplementary sequence  $d(m^5CGGGm^5CG) \cdot d(m^5CGCCm^5CG)$ . This Z-DNA parent structure has previously been solved by Schroth et al. to 1.3 Å resolution (Schroth et al., 1993). The crystal packing interactions were manipulated with the methylation of a single cytosine residue and the demethylation of four specific C5 methylated cytosine residues of this asymmetric hexamer duplex. A total of five modifications were introduced in the structure of the asymmetric hexamer. The modified duplexes have the general sequence  $d(xCGGGxCG) \cdot d(xCGCYyCxCG)$  in which  $x$  denotes the sites of demethylation of methylated cytosine residues, and  $y$  denotes the site of the methylation of a cytosine residue. The distribution of the hexamer orientation in the crystals results from differences in the potential interactions in the crystal lattice, and reflects the relative thermodynamic stability of the two crystal packing orientations. The observed free energy differences between the two packing interactions in these five modified structures and the parent structure were compared to calculated crystal packing van der Waals and solvation free energies.

## Introduction to the Crystal Structure of the Asymmetric Z-DNA Hexamer

### The structure of the $d(m^5CGGGm^5CG) \cdot d(m^5CGCCm^5CG)$ asymmetric duplex.

The crystal structure of the  $d(m^5CGGGm^5CG) \cdot d(m^5CGCCm^5CG)$  parent duplex has been solved to 1.3 Å resolution (Schroth et. al., 1993). The duplex crystallized in the  $P2_12_12_1$  space group and maintained the general structural characteristics of the Z-DNA helix. The structure of the asymmetric duplex was originally designed to study the effect of placing a cytosine residue in the sterically unfavorable *syn* conformation. The structure exhibited other key features which could be exploited for studying the thermodynamics of crystal packing. From this crystal structure, it was determined that the asymmetric hexamer duplex assumes two distinct orientations in the crystal lattice of the  $P2_12_12_1$  space group. In addition, the distribution between the two packing orientations favored one orientation by approximately a 2 to 1 ratio, reflecting a difference in crystal packing interaction energies of  $0.46 \text{ kcal mol}^{-1} \text{ hexamer duplex}^{-1}$  between the two packing orientations. The ability to distinguish between the two potential packing orientations and quantitate the distribution of the duplex between the two orientations in the crystal, provides a novel system for studying the thermodynamics of the packing of the hexamers of Z-DNA in the crystal lattice. The effect of simple modifications to the parent structure on the distribution of the duplex in the crystal lattice is expected to reflect the balance of thermodynamic forces that determine the orientation of the duplex in the crystal lattice. In addition, the high resolution structures available from these crystals allow a detailed analysis of the changes in both the crystal packing interactions.

General structural features of the asymmetric hexamer duplex. The two key features of the parent structure that are pertinent to the studies presented in this chapter are the asymmetry of the duplex, which allows the identification of the two packing orientation in the crystal structure, and the sites of methylation which occur at each of the

cytosine residues. These general features of the sequences are illustrated with the Z-DNA structure of the parent asymmetric hexamer duplex in figure 4.1. The asymmetry of the hexamer duplex is highlighted in the figure with the different coloring of the two unique hexanucleotides. The sequence of the blue hexanucleotide is d(m<sup>5</sup>CGGGm<sup>5</sup>CG). This sequence has as its distinguishing feature two central guanine residues, and is referred to as the 'GG' strand. The complementary red hexanucleotide, referred to as the 'CC' strand, has the sequence d(m<sup>5</sup>CGCCm<sup>5</sup>CG) with two central cytosine residues. The residues of each hexanucleotide in the figure are numbered from 1 through 6 in the 5' to 3' direction. The duplex generated by the two non-selfcomplementary strands is asymmetric in that the two strands are not interchangeable as is the case for self-complementary duplexes. Since the two packing orientations differ in the position of the strands in the duplex, this asymmetry is a key feature in distinguishing between two packing orientations in the crystal lattice.

The second feature shown in this figure are the methylation sites that were modified in these studies. These sites are highlighted in the figure with the coloring of the space filling models of the methyl hydrogen atoms. A total of 5 modifications to the parent structure were studied involving a combination of changes at these potential methylation sites. These combinations include a single methylation, and the combined effects of the methylation and 4 individual demethylations. The methylation site at the fourth cytosine residue (CYT 4) of the CC strand is highlighted in the figure with the hydrogens of the methyl group colored green. This methylation occurs at either residue 4 or residue 10 in the crystallographic asymmetric unit depending on the orientation of the duplex in the lattice. Therefore, this methyl group is referred to as the "METH4-10" methyl group. The two crystal packing orientations are described in a later sections. The four demethylations of the asymmetric duplex were at the C5 methylated cytosine residues

Figure 4.1 Space filling model of the asymmetric Z-DNA hexamer duplex showing the modification sites. The asymmetry of the hexamer duplex is highlighted in the figure with the different coloring of the two unique hexanucleotides. The sequence of the blue hexanucleotide is d(m<sup>5</sup>CGGGm<sup>5</sup>CG) and has as its distinguishing feature two central guanine residues. This strand is referred to as the 'GG' strand. The complementary red hexanucleotide has the sequence d(m<sup>5</sup>CGCCm<sup>5</sup>CG) and has two central cytosine residues, and is referred to as the 'CC' strand. The residues of each hexanucleotide in the figure are numbered from 1 through 6 in the 5' to 3' direction.

The methylation sites that were modified in these studies are highlighted in the figure with the coloring of the hydrogen atoms of the methyl groups. A total of 5 modifications to the parent structure were studied involving a combination of changes at these potential methylation sites. The site of methylation is at cytosine (CYT) residue 4 of the CC strand. This methyl group is referred to as the METH4-10 methyl group, and is highlighted in the figure with the hydrogens of the introduced methyl group colored green. The four demethylation sites on the asymmetric duplex are at cytosine (CYT) residues 1 and 5 of the GG strand, and cytosine residue 1, and 5 of the CC strand. The sites of demethylation are highlighted in the figure with white hydrogen atoms for these methyl groups.

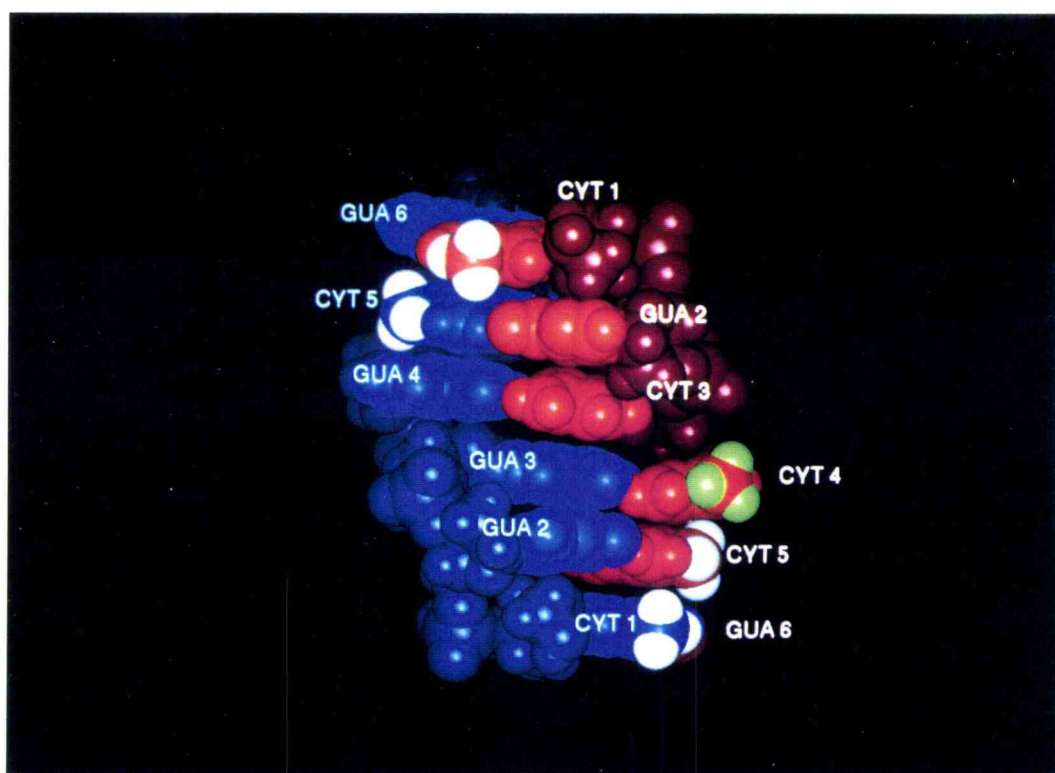


Figure 4.1

(MCY) 1 and 5 of the GG strand, and the MCY residues 1, and 5 of the CC strand. The sites of demethylation are highlighted in the figure with white hydrogen atoms for these methyl groups.

The nomenclature for the parent and modified crystal structures are given on table 4.1. The structures are named according to the modifications to the parent GG and CC hexanucleotides. The first structure listed on the table differs from the parent sequence by the addition of a single methylation at the *fourth* CYT residue (METH4-10 group) of the CC hexanucleotide and is therefore referred to as the 4mCC structure. The remaining modifications are systematic demethylations of the external dinucleotides which generate hemi-methylated d(CG) dinucleotides. These sequences were designed to study the combined effects of this METH4-10 methylation with the systematic demethylation of the MCY residues of the four external base pairs. The sites of demethylation are specified in the names of the structures which designate the strand and residue modified within the duplex. For example, the GG-MC1 structure refers to the crystal structure of the asymmetric duplex in which the methyl group of the first MCY residue of the GG strand has been removed. The 4mCC structure is related to the parent and the remaining four modified structures by the removal of a single specific methyl groups and is therefore defined as the 'reference' structure.

Table 4.1. Nomenclature for the crystal structures of the six asymmetric hexamer duplexes. The positions of the cytosine residues which were modified are highlighted with bold characters. Each strand of the duplex is crystallographically unique, therefore, structures are named according to their component DNA hexanucleotides. For example, the parent hexamer duplex is named the GG/CC structure according to its component CC and GG hexanucleotides. The structure of the modified duplexes are named according to the modifications made to the parent GG and CC sequences. Therefore, the 4mCC structure refers to the structure in which the parent structure is modified by a methylation of fourth cytosine residue of the CC strand. This methylation occurs in all modified structures and is referred to as the METH4-10 methyl group. The 4mCC is related to the parent, and the remaining four structures by demethylations of specific MCY residues, and is therefore designated as the reference structure. The demethylated structures are named according to the strand and residue which is demethylated in the hexamer duplex.

Table 4.1.

Crystal Structure	Hexanucleotide Sequences	Modification
GG / CC	$d(m^5CGGGm^5CG) \cdot d(m^5CGCCm^5CG)$	Parent structure. Duplex of GG and CC strands.
4mCC	$d(m^5CGGGm^5CG) \cdot d(m^5CGCm^5Cm^5CG)$	Reference structure. Methylation of CC hexanucleotide at residue 4 (METH4-10)
GG-MC1	$d(CGGGm^5CG) \cdot d(m^5CGCm^5Cm^5CG)$	Demethylation of residue 1 of GG hexanucleotide.
GG-MC5	$d(m^5CGGGCG) \cdot d(m^5CGCm^5Cm^5CG)$	Demethylation of residue 5 of GG hexanucleotide.
4mCC-MC1	$d(m^5CGGGm^5CG) \cdot d(CGCm^5Cm^5CG)$	Demethylation of residue 1 of 4mCC hexanucleotide.
4mCC-MC5	$d(m^5CGGGm^5CG) \cdot d(m^5CGCm^5CCG)$	Demethylation of residue 5 of 4mCC hexanucleotide.



### Introduction to the Crystal Packing of Z-DNA Hexamers in the $P2_12_12_1$ Lattice.

Modifications and their environments in the Z-DNA  $P2_12_12_1$  crystal lattice. The potential effects of each of these modifications on the lattice interactions of the Z-DNA duplex can be better understood if they are examined in the context of the  $P2_12_12_1$  crystal lattice. Figure 4.2 shows a diagram of the crystal packing arrangement of Z-DNA hexamers in the  $P2_12_12_1$  crystal lattice. The asymmetric unit in this space group is a single Z-DNA hexamer duplex (colored red in figure 4.2). The figure shows three views of the lattice looking down the three principle crystallographic axes. In the crystal lattice, the duplexes stack with their helical axes aligned parallel to the crystallographic *c*-axis generating essentially continuous helices of Z-DNA (panels B and C). In addition, these stacks of Z-DNA are associated laterally along the *a*- and *b* crystallographic axes generating a tight hexagonal arrangement of columns of Z-DNA (panel A).

The crystal structure of the parent GG/CC duplex showed that the hexamer duplex is able to assume two distinct and discrete packing orientations in the crystal lattice. The features of the two packing orientations available to the duplex are illustrated in figure 4.3. In this figure, the GG hexanucleotides are colored blue, and the CC hexanucleotides are colored red. Each asymmetric unit in the figure (delineated with a box) is comprised of a GG hexanucleotide and a CC hexanucleotide. The nucleotide residues in the crystallographic asymmetric unit are numbered from 1 through 6 in the 5' to 3' direction of one strand, and 7 through 12 in the 5' to 3' direction of the complementary strand. In this nomenclature, residue 1 is base paired with residue 12, residue 2 is base paired with residue 11, and so forth through the duplex. The general difference between the two packing orientations is apparent when comparing the position of each strand within the asymmetric unit in each packing orientation. The packing orientation illustrated in Panel A shows the duplex in the "OG" packing orientation in which the GG strand is assigned residue numbers 1 through 6. Panel B shows the duplex in the alternative "OC" packing

Figure 4.2 Crystal packing diagrams of Z-DNA hexamers in the orthorhombic  $P2_12_12_1$  space group. The unit cell in these crystals have approximate dimensions of  $a=18\text{\AA}$ ,  $b=31\text{\AA}$ , and  $c=45\text{\AA}$ . The unit cell edges are highlighted in each panel as a white box. The asymmetric unit for crystals of Z-DNA hexamers in the  $P2_12_12_1$  space group is a single hexamer duplex and is highlighted in all panels as the red hexamer duplex. Panel A shows a view down the crystallographic  $c$ -axis, while panels B and C illustrate views down the  $b$ - and  $a$ - crystallographic axes respectively. Panel B shows the stacked arrangement of hexamers with a view perpendicular to the crystallographic  $b$ - and  $c$ -axes. Additional interactions are shown in this panel between adjacent stacked helices along the crystallographic  $a$ -axis. The hexamers in each stack in this view are related by a translation of one unit cell length along the  $a$ -axis. Panel A illustrates the interactions between helices along the  $b$ -axis with a view of the unit cell perpendicular to the crystallographic  $a$ - and  $c$ -axes. The interactions along this axis are between helices that are oriented in the opposite directions in the crystal lattice and are displaced by a quarter of a turn of Z-DNA along the  $c$ -axis.

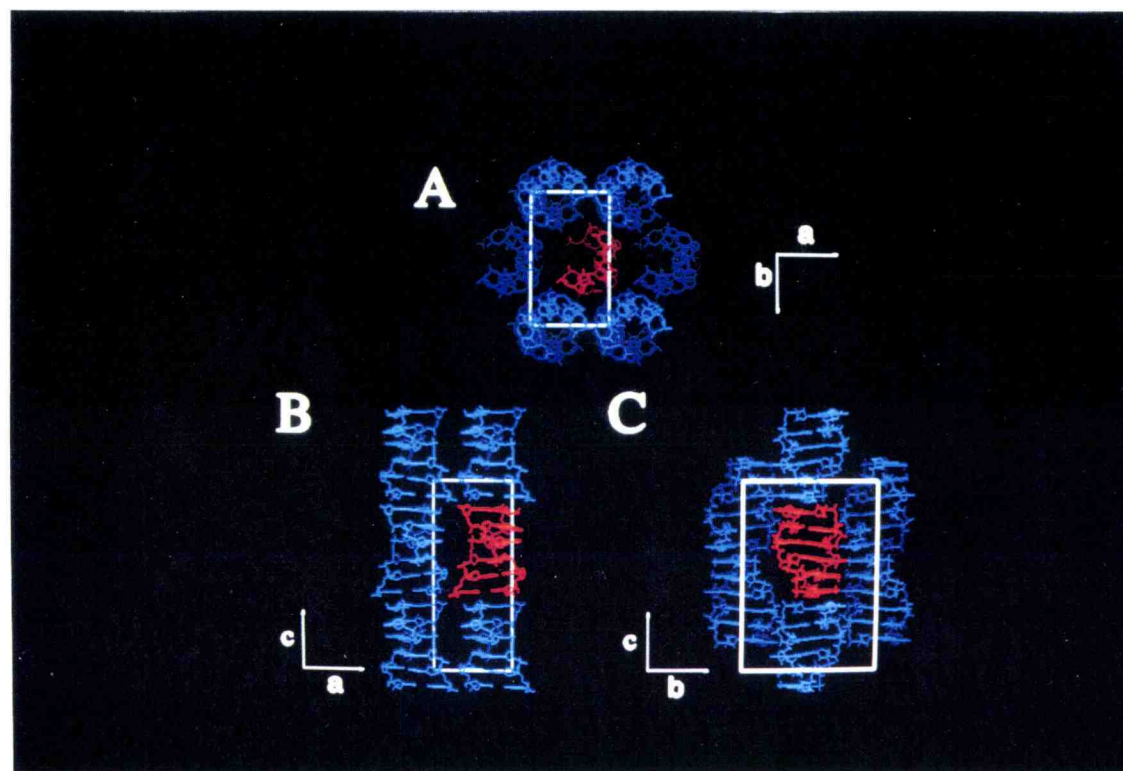


Figure 4.2

Figure 4.3. Crystal packing diagrams of the two orientations available to the hexamer duplex in the  $P2_12_12_1$  space group. The asymmetric unit in crystals of hexamers of Z-DNA in this space group is a single hexamer duplex. Each hexamer duplex in the figure is highlighted with a box, and defines one half-turn of Z-DNA. The nucleotide residues in the hexamer duplex are numbered C1 through G6 in the 5' to 3' direction on one strand, and C7 through G12 in the 5' to 3' direction of the complementary strand. In the crystal lattice, the duplexes stack with their helical axes aligned parallel to the crystallographic *c*-axis generating essentially continuous helices of Z-DNA. The hexamers of adjacent stacks are related by a second two-fold screw transformation in the plane of C5-G8 base pair, parallel to the crystallographic *b*-axis. In both packing diagrams, the d(m<sup>5</sup>CGGGm<sup>5</sup>CG) hexanucleotides are colored blue, and the d(m<sup>5</sup>CGCCm<sup>5</sup>CG) hexanucleotides are colored red. Figure A and figure B are views down the crystallographic *a*-axis, and show the packing diagrams for the "OG" packing orientation and "OC" packing orientation respectively. The models of the two packing orientations differ in their residue assignments for each strand in the asymmetric unit. In the "OG" packing orientation shown in figure A, the d(m<sup>5</sup>CGGGm<sup>5</sup>CG) hexanucleotide (blue strand) is assigned residue numbers 1 through 6, and the d(m<sup>5</sup>CGCCm<sup>5</sup>CG) hexanucleotide (red strand) is assigned residue numbers 7-12. In the "OC" packing orientation shown in figure B, the d(m<sup>5</sup>CGCCm<sup>5</sup>CG) hexanucleotide is assigned residue numbers 1 through 6, and the d(m<sup>5</sup>CGGGm<sup>5</sup>CG) hexanucleotide is assigned as residue numbers 7 through 12. Therefore, the two packing orientations are related by the interchanging of the residue assignments for each of the component hexanucleotides and have different crystal packing environments. This difference can be seen by comparing the environment of the 3-10 base pair in the OG packing versus the OC packing.

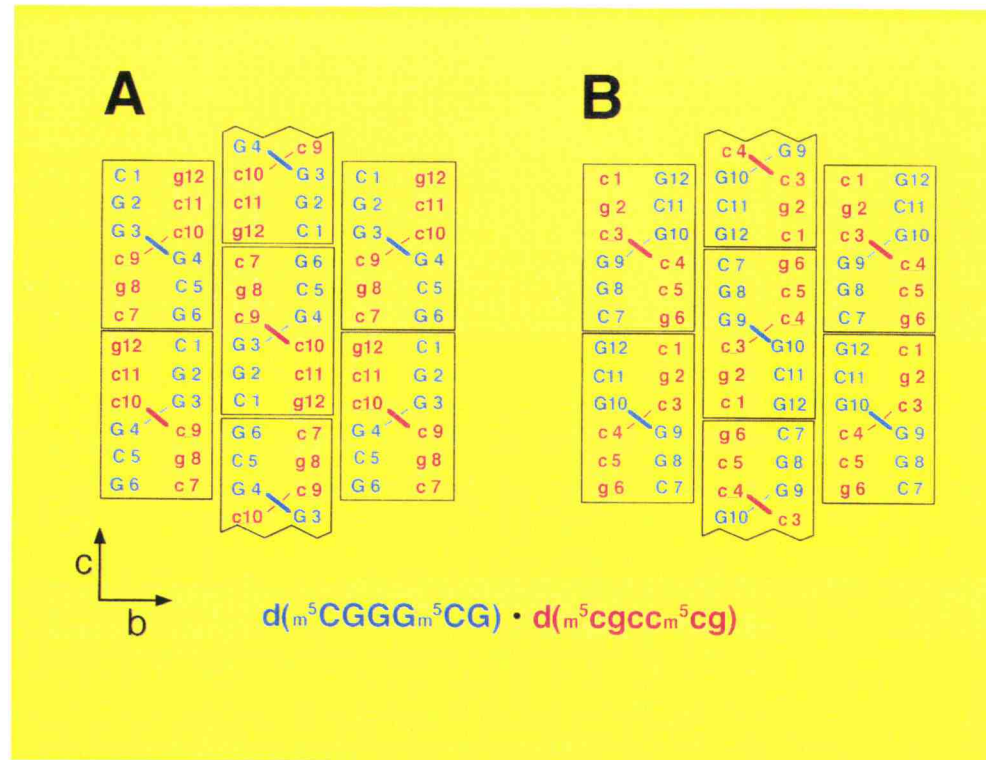


Figure 4.3

orientation in which the CC strand is assigned residue numbers 1 through 6. The most obvious difference between the two packing orientations is the interchange of the residue assignments for the GG and CC hexanucleotides in the asymmetric unit.

A consequence of the availability of two packing orientations in the lattice (related by interchanging the two hexanucleotides in the duplex) is that each residue in the duplex has two potential nucleotide positions in the crystal structure. A modification at the first residue of the CC strand would be apparent at either residue 1 or residue 7 depending on strand assignments in the asymmetric unit. The characteristics of the crystal lattice environment of these residues are determined by the packing of the Z-DNA hexamer in the  $P2_12_12_1$  lattice. This is best illustrated in a comparison of the environment of the nucleotide residues of the central dinucleotide in the OG versus OC orientations. In the OG orientation, the central guanine residues of the GG strand are located in an open channel in the lattice while the central cytosine residues of the CC strand are tightly packed against a symmetry related duplex. In the alternate OC orientation, it is the cytosine residues of the CC strand which sit in the open solvent channel and the guanine residues of the GG strand are tightly packed against the symmetry related duplex. Thus, the relative stability of the two orientations in the crystal lattice is dependent on the differences in the interactions between the two crystal packing environments.

The general characteristics of the methylation sites modified in these studies are given on table 4.2 which lists contacts within 4 Å of the methyl carbons at each site and the non-hydrogen atoms of adjacent duplexes in the lattice. These distances were calculated with the OG and OC models of the parent GG/CC structure. The number of contacts, the residues involved, and the type of interaction are indicated on the table. The types of interactions are classified as either "stacking" interactions between adjacent duplexes in the lattice, or as "side-on" interactions which occur between backbone and

Table 4.2 Crystal packing environments for the methyl groups of the asymmetric hexamer duplex in the  $P2_12_12_1$  space group. The table reports the non-hydrogen contacts with the methyl carbons of the cytosine residues of the parent GG/CC structure. Only non-bonded atomic distances less than 4.0 Å were considered in the calculation. The types of contacts are characterized as being with adjacently stacked residues (stacking) or with the base and backbone atoms of adjacently aligned hexamers (side-on). The residue which the methyl group is in contact with is given as the "Contact Residue". The alternative residue is the residue assignment of the methyl group in the alternative orientation. This allows the comparison of the two environments which the methyl group is allowed to sample in the crystal packing.

Methyl Residue	Number of Contacts (<4Å)	Closest Approach(Å)	Type	Contact Residue	Alternative Residue
1	2	3.67	Ribose (side-on)	GUA 6	7
	2	3.82	Base (stacking)	GUA 6	
4	1	3.88	Ribose (side-on)	GUA 8	10
5	3	3.69	Ribose (side-on)	GUA 8	11
7	4	3.64	Base (stacking)	GUA12	1
	1	3.91	Ribose (stacking)	GUA12	
10	3	2.12	Base (side-on)	GUA12	4
11	0	>4.00	none	none	5

groove surfaces of adjacent duplexes in the crystal packing. The nucleotide residue of each methyl group in the alternative packing orientation (ie. related by interchanging the strand assignments) are listed as the "alternative residue" methyl position. For example, residue positions 1 and 7 are related by strand interchange in the hexamer duplex, and therefore are considered crystal packing alternatives. Table 4.2 shows that the methyl environments differ in both the type of interactions as well as in the distance of closest approach. The lattice environment of each methylation site can be characterized as either relatively solvent exposed (ie. absence of side-on interactions) or located at the interface between helices within the crystal lattice (ie. with side-on interactions). Furthermore, the sites which are characterized as solvent exposed have crystal packing alternatives which are characterized as being involved in interhelix crystal lattice interactions. For example, the residue 7 environment is characterized as solvent exposed in the absence of side-on interactions, while its packing alternative (residue 1) is involved in a side-on interaction with residue 6 of an adjacent duplex. Thus, the duplex may assume an orientation in the lattice with the modification site solvent exposed or involved in crystal packing interaction.

The effect of the methyl group introduced at the fourth CYT residue of the CC strand (METH4-10) is studied in the 4mCC structure. In the OC orientation, the methylation occurs at residue 4 in the asymmetric duplex. In the OG orientation, this same methyl group occurs at the "alternative residue" 10 in the asymmetric duplex. The methyl group environment of the fourth residue is characterized as an open solvent channel. Packing the methyl group in this environment of greater solvent exposure is therefore expected to be penalized by a less favorable solvation free energy. Alternatively, packing the methyl group at residue 10 is associated with a steric collision (2.12 Å) with the major groove surface of residue GUA12 of an adjacent duplex. This close contact occurs only with the duplex packed in the OG orientation. The packing of the duplex with a residue at this position is expected to be unfavorable due to the high



steric repulsion energy for this interaction. However, burying the methyl groups at this positions would be favorable when considering solvation free energies. Therefore, this METH4-10 methyl group examines the balance of steric versus solvent interactions in determining the orientation of the duplex in the crystal lattice.

The remaining structures study the combined effects of the METH4-10 methyl group, and four systematic single demethylations at the 1 and 5 MCY residues of GG and CC strands. The four demethylations generate hemi-methylated external d(CG) dinucleotides, and examine the effects of modifications of the 1, 5, 7 and 11 MCY residues of the hexamer duplex crystal structure. Close contacts occur for methyl groups residues 1 and 5 with interactions between grooves of adjacent hexamers in the lattice. The packing of methyl groups in these environments are expected to be favored by van der Waals dispersion energies, and decreased solvent exposure of the methyl groups. The alternatives residues for these close packing methyl sites are residues 7 and 11. The methyl environments of these residues are relatively solvent exposed. The packing of the methyl groups at these sites are expected to be associated with unfavorable solvation free energies due to the greater solvent exposure of the methyl group. Thus, these single demethylations of the duplex are designed to study the significance of differences in solvent interactions in combination with the balance of steric versus solvent interactions introduced by the METH4-10 methyl group in determining the orientation of the duplex in the crystal lattice.

Comparisons of the observed ( $\Delta G^{\circ}_{occ}$ ) and calculated differences in lattice interaction energies ( $\Delta \Delta SFE^{\circ}_{ref}$  and  $\Delta \Delta pvdW^{\circ}_{ref}$ ). The crystal structures of the asymmetric duplexes provide a measure of the distribution of the duplex between the two orientations in the lattice. From this distribution, the differences in lattice interaction energies ( $\Delta G^{\circ}_{occ}$ ) are calculated by assuming a Boltzmann relationship between two states or orientations. Although the calculation of  $\Delta G^{\circ}_{occ}$  is obtained from the distribution of the duplexes within the crystal lattice, the orientation of the duplex is

considered to be determined during the equilibrium between the Z-DNA hexamers in solution and the growing crystal surface. The assumption that the predominant conformation of the DNA during crystal growth is that of Z-DNA was supported by solution studies on dilute solutions of hexanucleotides in the crystallization solvents (Ho et al., 1991). Thus, the potential interactions which determine the orientation of the duplex and therefore the distribution in the crystal occur at the surface of the growing crystal. The observed *differences* in the distribution between the two orientations for each of the modified structures *relative* to the reference 4mCC structure would not be expected to be dependent on all possible interactions within the lattice. The differences in distribution are likely to be dependent on the interactions which are affected only by the specific modifications of the 4mCC structure. Thus, the values for  $\Delta G^{\circ}_{occ}$  represent the differences in key interaction energies between the two orientations.

The difference in observed lattice interaction energies ( $\Delta G^{\circ}_{occ}$ ) are compared to calculated values of differences in solvation free energies ( $\Delta\Delta SFE^{\circ}_{ref}$ ) and van der Waals interaction energies ( $\Delta\Delta pvdW^{\circ}_{ref}$ ). Both  $\Delta\Delta SFE^{\circ}_{ref}$  and  $\Delta\Delta pvdW^{\circ}_{ref}$  represent differences (OG orientation minus the OC orientation) in energies calculated for a single hexamer duplex in the context of the lattice interactions for each of the packing orientations. The crystal packing models are constructed from the parent GG/CC structure. Therefore, the models and calculated energies do not consider the effect of crystal packing forces. Thus, the calculated energies better model all potential interactions expected in the equilibrium between the Z-DNA hexamers in solution and the crystal surface. In addition, the values of  $\Delta\Delta SFE^{\circ}_{ref}$  and  $\Delta\Delta pvdW^{\circ}_{ref}$  are relative to the calculated energies of the reference 4mCC structure. Therefore, although all lattice interactions are considered in the calculation of  $\Delta\Delta SFE^{\circ}_{ref}$  and  $\Delta\Delta pvdW^{\circ}_{ref}$ , the calculated energies represent only the differences in key interactions affected by each specific modification of the 4mCC structure. Thus, both the calculated and observed

energies are expected to reflect the differences in key interactions at the crystal surface which influence the distribution of the duplexes in the crystal.

Preliminary calculations of the differences in lattice van der Waals and solvation free energies for the six asymmetric duplexes. The differences in the van der Waals packing energies ( $\Delta\Delta_{pvdW}^{\text{ref}}$ ) and solvation free energies ( $\Delta\Delta_{SFE}^{\text{ref}}$ ) calculated between the OG minus OC packing orientations for models of the five modified structures and the parent GG/CC structure are given on table 4.3. The differences in lattice energies are reported relative to the reference 4mCC structure and therefore represent the effect of removing the METH4-10 methyl group (GG/CC *minus* 4mCC) and demethylation of the external dinucleotides (hemi-methylated *minus* 4mCC).

From table 4.3, the METH4-10 methyl group introduces the most significant interactions for studying the competition between van der Waals lattice interactions and solvent interactions in determining the orientation of the duplex in the crystal lattice. For all the modified structures, the METH4-10 is predicted to destabilize the OG packing orientation due to the steric collision with a symmetry related GUA12 residue. This methyl group is associated with a destabilization of the OG orientation by approximately  $43 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$  when considering van der Waals interactions (GG/CC *minus* 4mCC). Based on this initial modeling of the crystal packing interactions, it is expected that the OG orientation would never be observed in the modified structures due to the presence of this methyl group. However, the METH4-10 methyl group is expected to stabilize the OG packing orientation by  $0.599 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$  (GG-CC *minus* 4mCC) due to the decreased solvent exposure in the OG orientation *versus* the OC orientation. Thus, in each case, a competition is created between van der Waals and solvation energy terms.

The remaining comparisons on table 4.3 illustrate the effects of demethylation of the external dinucleotides in the four hemi-methylated structures. The effects of

demethylating the external dinucleotides are expected to be less dramatic as compared to the METH4-10 methyl group based on the magnitude of the differences in van der Waals energies. The differences in solvation free energies predict that the relative to the 4mCC structure, the demethylation of the GG hexanucleotide should stabilize the OG orientation while the opposite effect is expected for demethylation of the CC hexanucleotide. The order of stability of the OG orientation is predicted from the  $\Delta\Delta SFE^{\circ}_{\text{ref}}$  to follow the order 4mCC-MC5 > 4mCC-MC1 > 4mCC > GG-MC1 > GG-MC5 > GG/CC.

The studies in this chapter utilize a novel system to examine the relative significance of two competing forces in determining the specific macromolecular interactions which occur in the process of crystallization. The crystal structures of the six asymmetric hexamer duplexes provide information concerning the structural effects of removing a methyl group from the external dinucleotides of the asymmetric duplexes. The consequences of the observed structural effects were apparent in differences in the lattice van der Waals interaction energies between the fully methylated structures, and the hemimethylated structures. The observed distribution of the duplex between the two packing orientations in each crystal structure allowed the comparison of the observed and calculated differences in lattice interaction energies. In this manner, these studies provide information on the balance of van der Waals interaction and solvation free energies in determining the orientation of the duplex in the crystal lattice.

Table 4.3. Predicted differences in solvation free energies ( $\Delta\Delta\text{SFE}^*_{\text{ref}}$ ) and van der Waals lattice interaction ( $\Delta\Delta\text{pvdW}^*_{\text{ref}}$ ) energies for the six asymmetric duplexes relative to the reference 4mCC structure. The differences are calculated between the OG minus OC orientations for models of the modified structures derived from parent GG/CC structure (Schroth et al., 1993). These differences in energies show the effect of removing the METH4-10 (GG/CC *minus* 4mCC) and the effects of removing specific methyl groups from the external dinucleotides (hemi-methylated *minus* 4mCC). The van der Waals energies were calculated with the X-PLOR structure refinement program using the *parm91* force field which is specific for DNA (Brunger, 1992). The solvation free energies differences were obtained from calculated solvent-accessible-surfaces (SAS) within the crystal lattice environment (Connolly, 1983). Solvation free energies were calculated by applying atomic solvation parameters specific for DNA atoms (Kagawa et al., 1989; Kagawa et al., 1993). All energies are in units of kcal mol<sup>-1</sup> duplex<sup>-1</sup>.

Structure	$\Delta\Delta\text{pvdW}^*_{\text{ref}}^1$	$\Delta\Delta\text{SFE}^*_{\text{ref}}^1$
4mCC	0.000	0.000
GG/CC	-43.454	+0.599
GG-MC1	+0.941	+0.036
GG-MC5	+0.851	+0.295
4mCC-MC1	-0.653	-0.097
4mCC-MC5	-0.726	-0.256

1. All modified structures include the effects of the introduced METH4-10 methyl group. Therefore, relative to the *parent* GG/CC structure, the OG orientations are destabilized by approximately 43 kcal mol<sup>-1</sup> duplex<sup>-1</sup> and stabilized by 0.599 kcal mol<sup>-1</sup> duplex<sup>-1</sup> due to the presence of this methyl group.

## Materials and Methods

Synthesis and purification of DNA hexanucleotides. All the oligonucleotides used in the following studies were synthesized using phosphoramidite chemistry on an Applied Biosystems DNA synthesizer located in the Central Services Laboratory of the Center for Gene Research and Biotechnology at Oregon State University. The crude preparations obtained from the DNA synthesizer were judged to be at least 95% pure based on the product of the coupling efficiencies of each step in the synthesis. In all cases, the only further DNA purification required to obtain diffraction quality crystals from these samples was by size exclusion chromatography using a Sephadex G-10 (Sigma Chemical Co., 40-120 micron bead diameter, MW cutoff 700) to remove the blocking agents and precursors remaining after the syntheses. Typically, each DNA preparation was dissolved in distilled de-ionized water, loaded onto a 96 ml (radius=1.3cm, height=18cm) column and eluted with distilled de-ionized water. The eluate was monitored for absorbance at 260nm and 200nm with a Hewlett Packard 8452A diode array spectrophotometer. The first peak eluted from the column, characterized by intense 260nm and 200nm absorbances, was collected as the full length hexamer oligonucleotide samples. The samples were then lyophilized, and redissolved in 100ul of 30mM sodium cacodylate (NaCac) pH 7 buffer. The concentration of the hexanucleotide samples were estimated by UV absorbance measurements at 260nm using the extinction coefficient for single-stranded DNA ( $\epsilon = 25\text{mg/O.D. ml}$ ). The DNA samples used in the crystallization experiments were prepared by combining equal molar quantities of the appropriate hexanucleotide sequences for each of the asymmetric duplexes.

Crystallization experiments. The crystals of all six asymmetric duplexes were obtained at room temperature using the sitting drop vapor diffusion method (Ducruix and Giege, 1992). In the crystallization experiments for Z-DNA hexamers, a sample well containing an aqueous solution of the DNA sample, buffer, salts, and a precipitant was

equilibrated against an aqueous reservoir solution containing a higher concentration of the precipitant. The precipitant used in these studies, 2-methyl-2,4-pentanediol (MPD), has a lower vapor pressure than water. Therefore, the equilibration of the sample and reservoir solutions occurs through the vapor diffusion of the more volatile water. Under the conditions in the Z-DNA crystallization experiments, the equilibration of the two solutions in a closed system slowly reduces the volume of the sample solution until the DNA concentrations solution becomes supersaturated.

All the crystallization experiments in this chapter used gold label magnesium chloride ( $\text{MgCl}_2$  Aldrich Chemical Company), and sodium cacodylate (Sigma Chemical Company) titrated to pH 7.0 with hydrochloric acid (reagent grade, J.T. Baker Chemical Company) as a buffer. Gold labeled and reagent grade 2-methyl-2,4-pentanediol (Aldrich Chemical Company) were used as the precipitant in the sample well and reservoir solutions, respectively. A 20mM stock solution of polyamine was prepared from spermine tetrahydrochloride (Sigma Chemical Company), pH adjusted to 7 with sodium hydroxide (reagent grade, Aldrich Chemical Company) for the crystallization of the hemimethylated duplexes. All aqueous stock solutions used in the sample wells in the crystallization experiments were prepared using glass distilled de-ionized water.

The initial salt conditions used in the crystallization experiments were calculated from the predicted equilibrium CS, by taking into account the reduction of the sample volume during the crystallization process. This was estimated as the ratio of the reservoir MPD concentration over the sample well MPD concentration and therefore estimates the expected factor by which the concentration of each of the sample components will increase upon equilibration of the system.

Crystallization conditions for the five modified asymmetric duplexes. The crystallization conditions for the methylated 4mCC structure were estimated to be similar to the parent GG/CC structure. The initial search conditions for the crystallization of this

sequence were centered about the conditions for the GG/CC structure ranging in equilibrium CS from 0.8M to 9.6M.

The remaining structures were designed to study the effect of single demethylations of the external dinucleotides in the 4mCC structure. The uncertainty in the structure and therefore stability of the hemi-methylated dinucleotides made the prediction of the crystallization conditions by this method less reliable. Thus, an initial broad search of the equilibrium CS was conducted for each of the hemi-methylated sequences. For each hemi-methylated sequence, the equilibrium CS of these initial crystallization experiments ranged over 0.957M to 16.765M. All the sequences studied crystallized within the limits of these initial very broad search conditions.

Data collection. The data for the five asymmetric hexamer duplexes were collected on a Siemens P4 diffractometer with a sealed tube copper X-ray source, using the XSCANS version 2.2 data collection software (Siemens, 1993). All data sets were collected at room temperature for a single crystal mounted in a sealed glass capillary tube using power settings of 40mA and 50kV. The data for the parent GG/CC crystal was collected by Dr. Charles Campana on a Siemens P4R diffractometer with a rotating anode copper X-ray source.

Lorentz, and polarization corrections as well as corrections for X-ray induced radiation decay were applied to the raw intensity data for all crystals (Siemens, 1993). At the power settings used for the data collection, the intensity drop-off was approximately 5-10% over the typical one week data collection period as determined by the intensities of a minimum of 3 standard reflections. The final corrections for absorption of X-rays by the crystal were applied using the SHELXTL PC refinement program (Siemens, 1990). The shape of the crystal was approximated as an ellipsoid with a radius length defined as the average of the three dimensions of the block shaped crystals. An absorption coefficient of  $1.56\text{mm}^{-1}$  (determined for the GG/CC crystal) was used in the absorption



corrections for all data sets. The space groups were determined using the SHELXTL PC refinement program (Siemens, 1990).

General refinement of the crystal structures using X-PLOR. All model building, structure refinement, and molecular mechanics calculations were performed on Silicon Graphics Inc. Personal Iris workstations. DNA models were built utilizing the InsightII molecular modeling software from Biosym Inc. The X-PLOR version 3.1 refinement package (Brunger) was used in the refinement of all the crystal structures. Electron density map sections were calculated and drawn on a Digital Equipment Corporation GPS VaxII workstation using the utilities of the PRLS19 (Hendrickson & Konnert, 1979) refinement program modified to account for the additional atomic occupancy variable in the structures.

Restrained macromolecular structure refinement using the X-PLOR refinement program combines the energy minimization of molecular mechanics with the minimization of the standard residual of the crystallization refinement routines. The total energy ( $E_{Tot}$ ) minimized in the refinement process, given in equation 1, is the sum of an empirical energy function ( $E_{Emp}$ ) and the weighted crystallographic residual ( $E_{Eff}$ ).

$$E_{Tot} = E_{Emp} + W_a E_{Eff} \quad \text{Eq. 1}$$

In equation 1,  $E_{Emp}$  is the restraining energy function which describes the potential energy of the molecule, and  $W_a$  is the weight that scales  $E_{Eff}$  to  $E_{Emp}$ .  $E_{Emp}$  describes the potential energy of the molecule with separate energy terms for intramolecular geometric constraints ( $E_{Geo}$ ), intramolecular non-bonded energies ( $E_{Int}$ ), and intermolecular non-bonded lattice energies ( $E_{Inter}$ ) (equation 2).

$$E_{Emp} = \sum (E_{Geo} + E_{Int} + E_{Inter}) \quad \text{Eq. 2.}$$

In equation 2, the summation is over all atom pairs and four atom sets for these interactions. The *parmnahle* forcefield was used in the refinement of all crystal structures. This forcefield was derived from the CHARMM force field and defines force constants and geometric constants specific for DNA structures (Nillson & Karplus, 1986). The force constants involved in maintaining base planarity in the  $E_{\text{dihedral}}$  were increased by a factor of 10 as advised in the X-PLOR manual (Brünger, 1992).

The  $E_{\text{Eff}}$  term defines the constraints imposed by the crystallographic data (Equation. 3). The  $E_{\text{Eff}}$  energy function is in fact the standard residual between the calculated and observed structure factors.

$$E_{\text{Eff}} = \sum_h w_h [|F_o| - |F_c|]^2 \quad \text{Eq. 3.}$$

In equation 3,  $F_o$  and  $F_c$  are the observed and calculated structure factors respectively. The calculated structure factor ( $F_c$ ) for each reflection ( $h$ ) is a function of 2 terms that are normally not separable in the refinement process (equation 4). These are the thermal factor (or temperature factor  $B_i$ ) and the occupancy of the atom ( $Q_i$ ).

$$F_c(h) = \sum Q_i f(B_i, h) \exp[2\pi i(h \cdot r)] \quad \text{Eq. 4.}$$

(sum over atoms)

In equation 4, the occupancy factor ( $Q_i$ ) describes the fraction of the time an atom or group of atoms is present at a particular position in space ( $r$ ). The term  $f(B_i, h)$ , the atomic scattering factor for atom  $i$  describes the scattering power of the atom. The atomic scattering factor of each atom is adjusted by a temperature factor  $B_i$  which describes the thermal motion of the atom. Standard positional refinement routines generally refine the atomic coordinates and the temperature factors of the asymmetric unit. In addition to

standard positional refinement, the studies presented in this chapter require the refinement of the occupancy factor of selected atoms in order to quantitate the distribution of both packing orientations in the crystal lattice.

Structure and occupancy refinement of  $d(xCGGGxCG) \cdot d(xCGCyCxCG)$ . The initial models for the refinement of the 4mCC structure were constructed in the InsightII molecular modeling software by methylating the parent GG/CC structure (Schroth et al., 1993). In a similar manner, the initial models for the hemi-methylated structures were constructed by demethylating of the appropriate methylated cytosine residue of the 4mCC structure.

The final structures for the six crystals were obtained through a process involving three separate phases of refinement (Table 4.4). Both conventional positional, and simulated annealing refinement routines were used in refining the atomic positions of the models for all structures. The occupancy of specific atoms in the molecule were refined in a separate routine.

The objectives of the initial phase of refinement were to determine the presence of two packing orientations of the hexamer duplex in the crystal lattice and to refine the atomic coordinates of the DNA atoms. The coordinates of the two packing orientations were refined separately for each crystal structure using conventional positional and simulated annealing routines with data from 8.0 to 2.0 Å resolution for data having  $I > 2\sigma(I)$ . For the methylated and demethylated derivatives, the presence of the methyl groups in question were first confirmed in  $2(F_o - F_c)$  difference maps in which the methyl carbons were omitted from the density map calculation. The methyl groups were identified in these maps as strong positive peaks approximately 1.5 Å away from the C5 base carbon atom of the cytosine ring. These methyl groups, once identified, were built into the model using the InsightII molecular modeling software. Strong water molecules were located as intense positive densities in  $2(F_o - F_c)$  difference maps and their positions added to the model. Higher resolution data was included in the refinement until the

resolution limit of each data set was reached. Additional water molecules were located, and added to the models. Before continuing the refinement, the presence of two packing orientations in each of the crystal structures was confirmed in  $2(F_o - F_c)$  difference maps as residual electron densities in the region of the base atoms of the central dinucleotide.

The second phase of refinement involved the refinement of the coordinates of composite models for each crystal structure. The composite models were constructed from the refined coordinates of two separately refined structures for the two packing orientations. The water molecules were removed from the model in order to avoid biasing the refinement of the DNA structure. The two sets of coordinates in the model were assigned as two distinct, non-interacting structures within the lattice. Thus, the OG packing model was refined in the context of the OG crystal packing, and the OC model in the context of the OG crystal packing. Initially, the composite model was refined with the occupancies of each packing orientations set at 50%, thus the total occupancy of the asymmetric duplex was 100%. Initial refinement of the composite models included data

Table 4.4. Protocol for the crystal structure refinement of the asymmetric Z-DNA hexamers.

Phase	Process	Data <sup>1</sup>	Notes
1	Conventional Positional	8.0-2.0Å, I>2σ(I)	Refinement of the individual packing models.
	Simulated Annealing	8.0-2.0Å, I>2σ(I)	
	Conventional Positional	8.0-R.L., I>2σ(I)	Waters located in 2(Fo-Fc) difference maps.
	2(Fo-Fc) Difference Map	8.0-R.L., I>2σ(I)	Identification of atoms of alternative packing.
2	Simulated Annealing	8.0-2.0., I>2σ(I)	Combined individual models from Phase 1 minus the water molecules. Occupancy of each model assigned as 50%.
	Conventional Positional	8.0-R.L., I>2σ(I)	Addition of first shell groove waters.
3	Occupancy/Temp factor	8.0-1.6Å, I>2σ(I)	Occupancies refined for N1, O6 of guanine and N4 of cytosine residues of central dinucleotide.
	Simulated Annealing	8.0-R.L., I>2σ(I)	
	Conventional Positional	8.0-R.L., I>2σ(I)	Further addition of water molecules
	Occupancy/Temp factor	8.0-1.6Å, I>2σ(I)	

1. R.L. indicates that the refinement included data down to the resolution limit of the data set.

from 8.0 to 2.0 Å resolution having  $I > 2\sigma(I)$ . Following simulated annealing and conventional positional refinement routines, the first hydration shell waters were located in  $2(F_0 - F_C)$  difference maps, and added to the structure. The positions of the first shell water molecules, were located in each of the crystal structures as strong positive densities in  $2(F_0 - F_C)$  difference maps. The first shell water molecule are defined as water molecules which directly hydrogen bond to the potential hydrogen bonding atoms of the DNA molecule. For this initial stage of the refinement of the composite model, only the water molecules directly binding to the base atoms of the DNA in the major or minor grooves were candidates for addition to the model. This selection was applied in order to prevent the assignment of DNA density near the backbone atoms as water molecules. The positions of this set of first shell water molecules were refined using conventional positional refinement. These strong water positions were added to aid in the phasing of the base atoms, while avoiding the mis-assignment of backbone density as solvent density. The atomic coordinates of DNA models were further refined with simulated annealing, and conventional positional refinement including reflections to the higher resolution limit of the data sets.

In the final phase of refinement (Table 4.4) , the occupancies of the models were determined using the occupancy refinement routine in X-PLOR. This occupancy refinement was followed by the location and addition of weaker water molecule densities. The atoms of the two external dinucleotides of the asymmetric duplex are chemically equivalent between the two packing orientations and are initially crystallographically indistinguishable. These atoms initially cannot be assigned to either of the two packing orientations and therefore were not used to determine the occupancies. The atoms which are distinguishable between the packing orientations occur at the 3-10 and 4-9 base pairs of the central dinucleotide. The base pairs of this central dinucleotide are unique to each packing orientation as a result of the differences in the hexanucleotide positions in the

asymmetric duplex between the two packing orientations (see Figure 4.3 ). In the OC orientation, the base pairs in the central dinucleotide are CYT 3 - GUA10 and CYT4 - GUA 9. In contrast, in the OG orientation the base pairs of the central dinucleotide are GUA3 - CYT10 and GUA4 - CYT9. Within these base pairs of the central dinucleotide, the atoms of the imidazole ring of the guanine base and the pyrimidine ring atoms nearly overlap with atoms of the alternative packing orientations (see Figure 4.3). The positions which are most distinguishable between the two orientations in the central dinucleotide are the N4 atoms of cytosine residues, and the N1 and O6 of the guanine residues. Only these atoms which have distinguishable positions between the two packing orientations were chosen to represent each of the packing orientations in the occupancy refinement.

The refined occupancy values for the six structures were obtained by an iterative process, alternating occupancy and temperature factor refinements with positional refinement (see phase 3 of Table 4.4). For all structures the maximum resolution limit of the data included in the occupancy refinement was determined by the limits of the weakest data set.

Following the occupancy refinement, the total number of electrons were calculated for the two models from the refined occupancy values of the three selected atoms. The number of electrons associated with each base pair of the central dinucleotide were summed giving the total electrons for each base pair. The occupancy of the packing orientations for each base pair was calculated as the fraction of the total electrons. The occupancy values obtained from the refinement of the central dinucleotide were then assigned to the models and positional refinement was continued. Because the initial positional refinement was conducted with both models assigned as 50% occupied, the positional refinement following the first occupancy refinement included a simulated annealing routine in order to allow the atoms of both models to find positions appropriate for the new occupancy values. The alternating occupancy and standard positional

refinement continued with the addition of water molecule positions until the residual density in  $2(F_O - F_C)$  difference maps could no longer be reliably assigned as water molecules.

For the GG/CC, 4mCC, GG-MC1, 4mCC-MC5, and 4mCC-MC1 crystal structures, the occupancy values for the base atoms of the internal dinucleotide converged using the refined temperature factors for those atoms used in the occupancy refinement. For the GG-MC5 structure, the occupancy values of the central base pairs were initially varied over 7% in the final stages of refinement. This data set was of lower resolution and the electron density information in these regions were not as distinct as the same regions in the higher resolution structures. The consistent refinement of the occupancies and the isotropic temperature factors of these atoms are expected to be increasingly difficult with lower resolution data. Therefore, for the GG-MC5 structure, the average temperature factor for the central base pair for each model was assigned to each of the base atoms of central base pair prior to the occupancy refinement.

Calculations of the observed differences in crystal packing energies ( $\Delta G^{\circ}_{occ}$ ), intramolecular and crystal packing interaction energies. The observed difference in free energy between the two packing orientations ( $\Delta G^{\circ}_{occ}$ ) was calculated from the refined occupancy values for each of the structures assuming a Boltzmann relationship between two states.  $\Delta G^{\circ}_{occ}$  is defined in equation 5.

$$\Delta G^{\circ}_{occ} = -RT \ln[\%OG/\%OC] \quad \text{Eq. 5.}$$

In Equation 5, %OG and %OC are the occupancy values of the OG and OC orientations obtained from the refinement of the six asymmetric duplexes. R is the gas constant (1.987 cal / ° K mol) and T is the temperature (296 °K)



All intramolecular energy terms and intermolecular lattice interaction energies were calculated in X-PLOR using the *parm94* forcefield. The solvation free energies were calculated from the coordinates of the final models from each refinement. The solvent accessible surfaces (SAS) of the hexamer duplex in the context of the crystal lattice was calculated using the Connolly rolling ball method (Connolly, 1983). The thermodynamic solvation parameters for DNA surfaces described in Chapters 2 and 3 were applied to the exposed surface areas to obtain the solvation free energies (SFEs) of the hexamer duplex in the crystalline environment.

In order to better account for the differences interactions between two orientations, the crystal packing models included the packing interactions between the duplexes in the lattice of the alternative packing orientation. In this model, the additional packing interactions were considered for the OG model in the alternative OC lattice, and the OC model in the alternative OG lattice. Highly unfavorable van der Waals contacts were created when placing each model in their respective alternative lattices. In order to account for these interactions, the energies in both the alternative lattice ( $pvdW_{Altdn}$ ) and self-consistent lattice ( $pvdW_{dn}$ ) were calculated for each dinucleotide in the duplex. The contribution of each van der Waals packing interaction term,  $pvdW_{dn}$  and  $pvdW_{Altdn}$ , to the total van der Waals energy of a dinucleotide ( $pvdW_{Tdn}$ ) was scaled by the probability of occurrence of each of the two interactions using a Boltzmann distribution between the two states (Equation 6).

$$P_{Altdn} = \exp [pvdW_{Altdn} / (pvdW_{Altdn} + pvdW_{dn})] \quad \text{Eq. 6.}$$

The total van der Waals lattice interaction energy for each dinucleotide ( $\Delta pvdW_{Tdn}$ ) is defined by equation 7.

$$pvdW_{Tdn} = (P_{Altdn})pvdW_{Altdn} + (1-P_{Altdn})pvdW_{dn} \quad \text{Eq. 7.}$$

The total packing van der Waals energy for the OG or OC orientations ( $pvdW_{OG}$  or  $pvdW_{OC}$ ) is simply the sum of the three component  $pvdW_{Tdn}$  energies for a hexamer duplex.

The SFEs for each packing orientation were also calculated in the context of the self-consistent lattice, and the alternative lattice. The SFEs of each packing orientation included the solvation free energy contributions from both solvent exposed and for buried surfaces. The exposed and buried surface areas were both calculated using the Connolly rolling ball method (Connolly, 1983). The buried surface areas were obtained from the difference between the exposed surface areas of the duplex in the context of the crystal lattice *minus* the exposed surface areas of the isolated hexamer duplex. The free energies for both buried and solvent exposed surfaces were calculated by applying the atomic solvation parameters (ASPs) derived for solvating DNA surfaces (Kagawa et al., 1989). The calculation of the free energy associated with burying surfaces in this manner results in a favorable negative free energy for burying hydrophobic surfaces and an unfavorable positive free energy for burying hydrophilic surfaces.

All solvation free energy terms were tabulated as energies per dinucleotide which included the contributions from both buried and exposed surfaces. As with the packing van der Waals energies, the solvation free energies were calculated in both the self consistent and alternative packing models. The scaling factor applied ( $P_{Altdn}$ ) to the

solvation free energies were the same factors determined by the van der Waals interaction energies for each dinucleotide ( $PA_{ltdn}$ ). The total solvation free energy for each dinucleotide ( $SFET_{dn}$ ) is defined in Equation 8.

$$SFET_{dn} = (1 - PA_{ltdn})SFE_{dn} + (PA_{ltdn})SFE_{Altdn} \quad \text{Eq. 8.}$$

In equation 8,  $SFE_{dn}$  and  $SFE_{Altdn}$  are the solvation free energies per dinucleotide associated with the hexamer duplex in the self-consistent and alternative crystal packings respectively. The total SFEs for the OG and OC orientations ( $SFE_{OG}$  or  $SFE_{OC}$ ) is the sum of the three  $SFET_{dn}$  dinucleotide energies in each hexamer duplex.

The differences in the packing van der Waals energies ( $\Delta pvdW$ ) and SFEs ( $\Delta SFE$ ) are the differences in free energies between the OG minus OC orientations (Equations 9 and 10).

$$\Delta pvdW = pvdW_{OG} - pvdW_{OC} \quad \text{Eq. 9.}$$

$$\Delta SFE = SFE_{OG} - SFE_{OC} \quad \text{Eq. 10}$$

The total difference (OG minus OC) in van der Waals energy ( $\Delta vdw_{Tot}$ ) includes contributions for the differences in intramolecular van der Waals energies ( $\Delta vdw_{int}$ ), and the packing van der Waals energy ( $\Delta pvdW$ ) (Equation 11).

$$\Delta vdw_{Tot} = \Delta vdw_{int} + \Delta pvdW \quad \text{Eq. 11.}$$

*Structural analysis of the Z-DNA crystal structures.* The helical parameters of the final DNA structures were calculated using the Nucleic Acid Structure Evaluation (NASTE) program (Basham et al., unpublished).

## Results

The final chapter of the thesis describes experiments designed to study the relative significance of van der Waals and solvent interactions in determining the orientation of Z-DNA hexamer duplexes in the crystal lattice of the  $P2_12_12_1$  space group. The significance of these forces are studied by monitoring the change in distribution of the duplex between the two distinct packing orientations in response to 5 modifications to the parent structure previously solved by Schroth et al.(1993) to 1.3 Å resolution. The crystal packing interactions were manipulated with the methylation of a single cytosine residue and the demethylation of four specific C5 methylated cytosine residues of this asymmetric hexamer duplex. The distribution of the hexamer of the two packing orientations observed in these crystal structures provides a measure of the difference in free energy between the two packing orientations ( $\Delta G^{\circ}_{occ}$ ). The observed  $\Delta G^{\circ}_{occ}$  are compared to the calculated lattice interaction van der Waals energies ( $\Delta\Delta p v W^{\circ}_{ref}$ ) and solvation free energies ( $\Delta\Delta SFE^{\circ}_{ref}$ ). In addition, the observed structural differences indicate differences in the manner in which each duplex optimizes crystal packing interactions.

Crystallization of the six asymmetric hexamer duplexes. The six asymmetric hexamers studied crystallized as Z-DNA under the conditions reported on table 4.5. The initial conditions and the equilibrium cationic strength (CS) which yielded diffraction quality crystals, and the time required for the appearance of each crystal are listed on table 4.5. The two hexamer sequences with fully methylated external dinucleotides crystallized at the extremes of the equilibrium CS values. The parent GG/CC hexamer crystallized under the lowest salt concentrations with an equilibrium CS value of 0.7 M while the 4mCC hexamer required an equilibrium CS of 6.24 M. The differences in the equilibrium CS between these fully methylated hexamers likely reflect the presence of the polyamines in the crystallization set-ups. Polyamines were not considered in the

Table 4.5. The crystallization conditions used to obtain diffraction quality crystals of the six asymmetric hexamer duplexes. The table reports the initial concentrations for the DNA, buffer, MgCl<sub>2</sub>, polyamine, and precipitant. In all experiments, the buffer was a sodium cacodylate pH7, the polyamine (PA) was spermine, and the precipitant was 2-methyl-2,4-pentanediol (MPD). The well (MPD<sub>w</sub>) and reservoir (MPD<sub>r</sub>) concentrations of precipitant are reported as percentages of the MPD/Water volume to volume ratios. The initial DNA concentration are reported as single stranded DNA concentrations.

Structure	ssDNA(mM)	Buffer (mM)	MgCl <sub>2</sub> (mM)	PA (mM)	MPD <sub>w</sub> (%)/MPD <sub>r</sub> (%)	Time (days)	CS(M) <sup>1</sup>
GG/CC	1.4	14	21	2.8	1.4 / 10.0	10	0.7
4mCC	1.4	30	30	0.0	0.5 / 20.8	28	6.2
GG-MC1	1.6	31	8	2.4	0.4 / 16.3	55	2.6
4mCC-MC1	1.5	30	4	2.4	0.4 / 17.1	30	2.0
GG-MC5	1.4	24	4	2.4	0.4 / 19.0	37	1.9
4mCC-MC5	0.7	30	5	3.0	1.0 / 20.8	35	1.0

1. The values for the cationic strength (CS) were estimated from the equilibrium magnesium and sodium concentrations by the relationship  $CS = \sum .22 [Mg^{2+}] + 1^2[Na^+]$ .

calculation of the equilibrium CS although they are known to stabilize the Z-conformation (Rich et al., 1984). The presence of spermine in the crystallization solutions would therefore reduce the magnesium concentration required to crystallize a sequence as Z-DNA. The remaining hexamer sequences each have a single hemi-methylated external dinucleotide, and crystallized at slightly higher equilibrium CS (ranging from 1.04 M to 2.57 M) as compared the GG/CC hexamer duplex.

Unit cell determinations, and data collection. The six asymmetric hexamer duplexes crystallized in the orthorhombic  $P2_12_12_1$  space group, and have unit cell parameters characteristic of crystals of Z-DNA hexamers (table 4.6). Therefore, the crystals of these asymmetric hexamer duplexes are isomorphous, indicating that the structural changes resulting from the modifications to the parent DNA strands were not large enough to completely disrupt the crystal lattice of these Z-DNA hexamers. This allows the study of the crystal lattice interactions with direct comparisons between the structures of these asymmetric duplexes.

The number of reflections observed *versus* the total possible reflections (percent completeness) for each data set are given on table 4.7. The table reports the percent completeness for data between 8 to 1.3 Å resolution. The X-ray diffraction data for each structure represent data sets collected from a single crystal of each sequence. The six crystals diffracted to better than 1.6 Å resolution, as judged by the highest resolution shell that is approximately 40% complete for reflections with intensities ( $I$ ) greater than  $2\sigma(I)$ . The high resolution data allows the determination of the orientation of the hexamer in the lattice, and the occupancy of each orientation.

Refinement of the six asymmetric hexamer duplex crystal structures. Table 4.8 reports the resolution limits, number of reflections, the final number of assigned waters, and final R-factors in the refinement the six crystal structures. In addition, table 4.8 reports the root mean square deviations in the bond distances ( $RMS_{bond}$ ) and angles ( $RMS_{angle}$ ) for the final OG and OC models. The refined structures in all cases fit

Table 4.6. Unit cell parameters and space groups determined for the six crystals of asymmetric Z-DNA hexamers. The cell lengths  $a$ ,  $b$  and  $c$  are reported in Å units, and the interaxial angles  $\alpha$ ,  $\beta$  and  $\gamma$  are reported degree units. The space group was determined using the *xprep* routine in the SHELXTL PC refinement package (Siemens, 1990).

Structure	$a$	$b$	$c$	$\alpha$	$\beta$	$\gamma$	Space Group
GG/CC	17.865	30.822	44.797	90.0	90.0	90.0	$P2_12_12_1$
4mCC	18.012	30.770	44.62	90.0	90.0	90.0	$P2_12_12_1$
GG - MC5	17.991	30.775	44.944	90.0	90.0	90.0	$P2_12_12_1$
4mCC - MC5	18.026	30.823	45.009	90.0	90.0	90.0	$P2_12_12_1$
GG - MC1	17.937	30.791	44.940	90.0	90.0	90.0	$P2_12_12_1$
4mCC - MC1	17.852	30.813	45.110	90.0	90.0	90.0	$P2_12_12_1$

Table 4.7. Comparison of the completeness of the data sets collected for the GG/CC, 4mCC, 4mCC-MC5, GG-MC1, GG-MC5, and the 4mCC-MC1 crystals. The percent completeness is fraction of the total possible number of reflections which are observed for each of the reported resolution shells. The % completeness for each resolution shell, and accumulated over all shells were calculated using reflections having intensities (I) better than  $2\sigma(I)$ . The high resolution limit used in the refinement of each structure is highlighted with bold characters.

Resolution range (Å)	GG/CC		4mCC		4mCC-MC5	
	Shell(%)	Accum.(%)	Shell(%)	Accum.(%)	Shell(%)	Accum.(%)
8.00 - 2.57	97.10	97.10	92.49	92.49	91.86	91.86
2.57 - 2.05	92.62	94.91	79.30	86.04	81.89	86.99
2.05 - 1.80	88.00	92.67	72.99	81.83	71.46	81.98
1.80 - 1.64	82.87	90.30	64.66	77.67	58.63	76.32
1.64 - 1.52	72.29	86.74	46.30	71.48	<b>44.54</b>	<b>70.07</b>
1.52 - 1.43	55.76	81.83	<b>38.81</b>	<b>66.20</b>	30.06	63.55
1.43 - 1.36	50.32	77.45	35.19	61.90	25.09	58.24
1.36 - 1.30	<b>44.37</b>	<b>73.39</b>	27.42	57.72	5.17	51.79
<hr/>						
Resolution range (Å)	GG-MC1		GG-MC5		4mCC-MC1	
	Shell(%)	Accum.(%)	Shell(%)	Accum.(%)	Shell(%)	Accum.(%)
8.00 - 2.57	85.37	85.37	89.32	89.32	91.60	91.60
2.57 - 2.05	66.63	76.24	74.76	82.22	85.61	88.68
2.05 - 1.80	49.33	67.49	58.08	74.42	75.96	84.58
1.80 - 1.64	<b>41.25</b>	<b>61.14</b>	<b>44.35</b>	<b>67.12</b>	66.67	80.21
1.64 - 1.52	26.39	54.30	33.46	60.49	49.57	74.20
1.52 - 1.43	12.88	47.62	26.32	54.95	<b>40.30</b>	<b>68.75</b>
1.43 - 1.36	00.00	40.99	22.63	50.48	30.95	63.46
1.36 - 1.30	00.00	36.02	4.79	44.92	23.60	58.53



Table 4.8. Parameters for the refinement of the crystal structures of the six asymmetric hexamer duplexes. The table reports the final R-factors, occupancies, and the number of assigned water molecules for each crystal structures. The occupancy was calculated as the average of the occupancies for the 3-10 and 4-9 base pairs. The free energy difference between the two packing orientations was calculated assuming an equilibrium between two distinct states and is the energy differences relative the G packing orientations. The RMSbond and RMSangle are the root mean square deviations from the equilibrium values calculated in the X-PLOR refinement program using the *parm94* forcefield. The deviations were calculated for the final structures of the G and C models omitting the hydrogen atoms.

Table 4.8.

	GG/CC	4mCC	GG-MC5	4mCC-MC5	GG-MC1	4mCC-MC1
Resolution Range (Å)	8.0 - 1.3	8.0 - 1.4	8.0 - 1.6	8.0 - 1.5	8.0 - 1.6	8.0 - 1.4
Number of Reflections	4741	3250	2229	2912	2024	3374
Number of Waters(solvent)	55	49	31	56	16	48
R-factor <sup>2</sup>	19.42	19.56	19.14	19.10	18.63	17.25
RMS <sub>bond</sub> (OG/OC) (Å)	0.02/0.01	0.02/0.02	0.02/0.02	0.02/0.01	0.02/0.02	0.02/0.02
RMS <sub>angle</sub> (OG/OC) (°)	3.4/3.3	3.9/3.5	3.8/4.0	3.6/3.7	4.0/3.9	3.5/3.5
Occupancy <sup>3</sup> (%OG / %OC)	68/32±1	50/50±1	43/57±3	70/30±1	47/53±0	54/46±1
$\Delta G^{\circ}_{occ}$ <sup>1</sup> (kcal/ mol duplex)	-0.44 ± 0.03	0.00 ± 0.02	+0.29 ± 0.09	-0.50 ± 0.05	+0.07 ± 0.00	-0.09 ± 0.03

1.  $\Delta G^{\circ}_{occ} = -RT \ln (\%OG/\%OC)$  where R is the gas constant (1.987 cal/ °K mol ), and T is the temperature (296 °K).

2.  $R\text{-factor} = R = \frac{\sum_h ||F_o| - |F_c||}{\sum_h |F_o|}$  in which  $|F_o|$  is the observed structure factor, and  $|F_c|$  is the calculated structure factor

3. Confidence limits are defined by the standard deviation of the average.

reasonable well to the observed data as indicated with final R-factors less than 20%. Spermine molecules were not observed in any of the six crystal structures, and the unambiguous assignment of a magnesium water complex was possible only in the GG/CC structure (Schroth et al., 1993).

A comparison of the characteristics of the electron density at the external dinucleotides and the central dinucleotides indicate that the duplexes distribute between two discrete orientations in the crystal lattice. Figure 4.4 shows a comparison of 2Fo-Fc electron density maps calculated for the 2-11, and 3-10 base pairs for the GG/CC structure (panel A and B), and the 3-10 base pairs of the 4mCC structure (panel C), and the GG-MC5 (panel D) structure. The electron density maps in this figure were calculated with the atomic coordinates of the OG model for each structure. The electron densities of the 2-11 base pair (panel A) has the typical features of the 1-12, 5-8, and 6-7 base pairs of the external dinucleotides. These sites are characterized by well defined electron densities associated with the DNA atoms and the absence of observed density in the intervening spaces between the atoms of the base pairs. The electron densities of the 3-10 base pair (panel B) of the same GG/CC structure are also distinct for a majority of the DNA atoms. However, in contrast to the electron densities of the 2-11 base pair, the 3-10 base pair has residual electron density between the base atoms of the OG model (thick bonds). This residual electron density is observed for the 3-10 base pair of the 4mCC (panel C) and the GG-MC5 structure (panel D) as well as in the remaining crystal structures. Residual electron density is also observed between the base atoms of the 4-9 base pair in all structures. Thus, the unassigned density in these crystals are localized in the region of the central dinucleotide. The comparison of the electron densities of the base pairs of the external and central dinucleotides suggests that additional scatterers not included in the OG models are present at the central dinucleotides but not at the external dinucleotides.

**Figure 4.4.** Comparison of 2Fo-Fc electron densities of the fully occupied 2-11 base pair and the partially occupied 3-10 base pair. The density maps were calculated for the 2-11, and 3-10 base pairs for the GG/CC structure (panels A and B), and the 3-10 base pairs of the 4mCC (panel C), and the GG-MC5 (panel D) structures. The density maps were calculated with the atomic positions of the OG packing orientations. The models for both packing orientations are drawn in the figure with the OG models drawn with thick bonds, and the OC models drawn with the thin bonds. Each panel shows a 2.5Å thick section along the crystallographic c-axis. The positions of the atoms for the OC models in all structures are highlighted with filled circles.

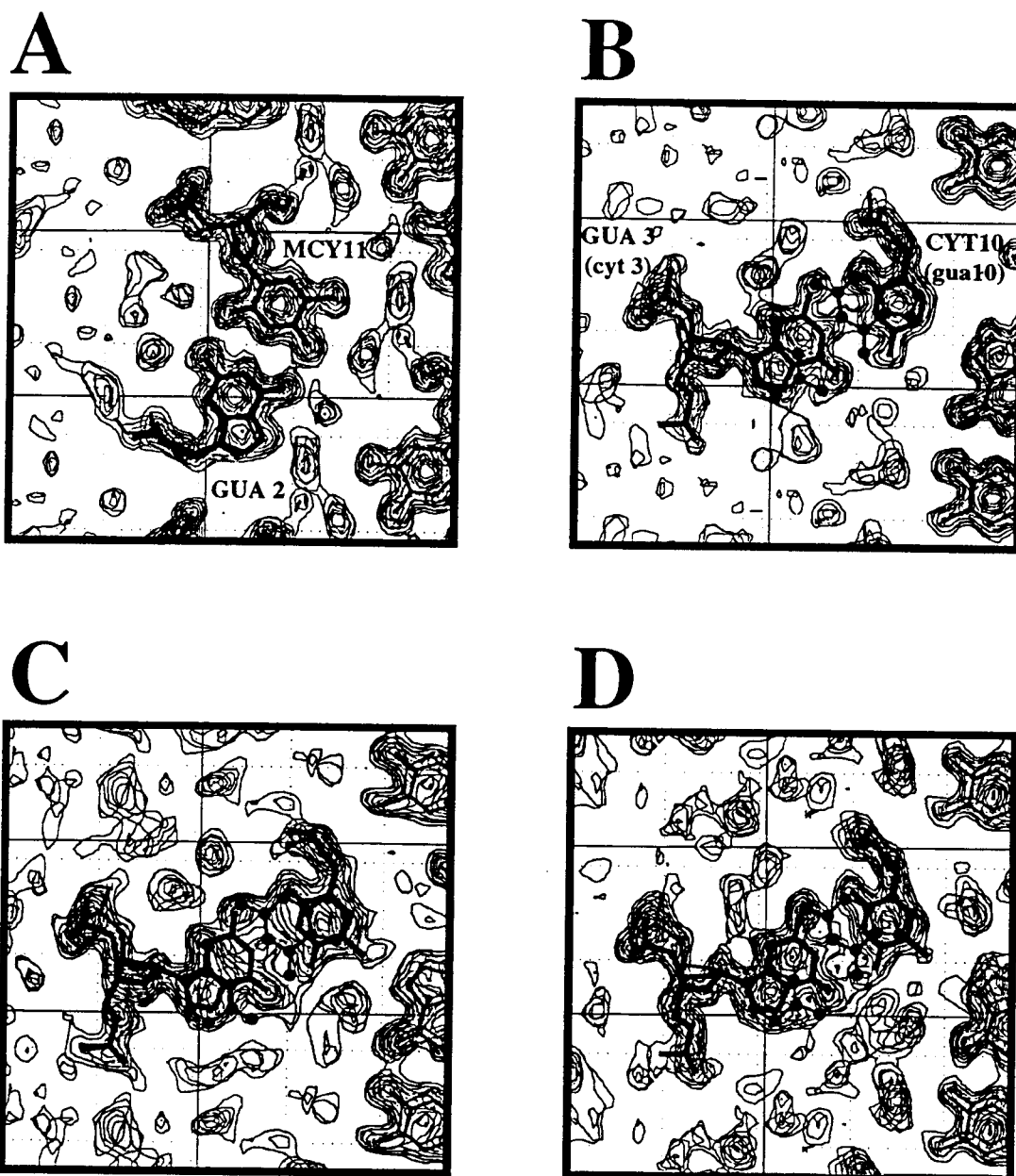


Figure 4.4

The ability of a single model to phase the external dinucleotides but not the central dinucleotide is consistent with the presence of two discrete packing orientations related by the interchanging of the strand assignments in the asymmetric unit of the crystal structures. The hexanucleotides which comprise the hexamer duplex in all structures are chemically equivalent at the two external dinucleotides (sans methyl groups of the cytosine residues). Thus, the residues of the external dinucleotides would not be altered by the interchanging the hexanucleotides in asymmetric unit, either strand assignment would result in external d(CG) dinucleotides. The hexanucleotides differ in the two central residues with two central cytosine (CYT) residues for CC strand and two guanine (GUA) residues for the GG strand. The two packing orientations would therefore be distinguishable at the central dinucleotide with overlapping coordinates of CYT and GUA residues at the 3-10 and 4-9 base pairs of the asymmetric unit. This is consistent with the observations from the  $2F_0-F_c$  electron density maps sections in figure 4.4.

A quantitative difference in the distribution of the duplex in the OG orientation is observed with a comparison of the 3-10 base pair in the GG/CC structure, the 4mCC, and the GG-MC5 (figure 4.5). Figure 4.5 shows  $2(F_0 - F_c)$  difference electron density map sections of the central 3-10 base pair of the GG/CC (panel A), 4mCC (panel B), and the GG-MC1 (panel C) crystal structures. The maps were calculated using the atomic coordinates of OG models and the assigned water molecules. The difference in the OG occupancy is evident in the increased residual electron density in the intervening spaces between the base atoms of the OG models (panels B through D for the GG/CC through GG-MC5 structures). The electron densities of the base pairs shown in figures 4.4 and 4.5 strongly suggest that a better description of the crystal structures for these asymmetric hexamer duplexes is a composite model which consists of a distribution of the hexamer duplex between the two packing orientations.

**Figure 4.5.** Comparison of 2(Fo-Fc) difference electron density maps of the 3-10 base pairs of the GG/CC, 4mCC, and GG-MC5 crystal structures. The GG/CC structure is shown in panel A, the 4mCC structure in panel B, and the GG-MC5 structure in panel C. Each section shows a 2.5 Å thick slice along the crystallographic c-axis with contouring of the positive difference electron density. The occupancy ratio of the OG:OC models in each of the structures are given at the bottom of each of the panels. The maps were calculated with the OG models assigned as 100% occupied, therefore the positive electron density indicates the presence of additional density which is not accounted for by the OG model. A comparison of these sections shows qualitatively that as there is increased residual positive density as the occupancy of the OG model decreases. In addition, the residual density in the region of atoms of the central base pair very nearly fit the atom positions of the OC models. Each panel shows a 2.5 Å thick section along the crystallographic c-axis. Residues GUA 3, and CYT10 (or MCY10 for the 4mCC and GG-MC5 structures) of the OG model are drawn with thicker bonds while the CYT 3 and GUA10 residues of the OC models are drawn with thin bonds. The positions of the atoms for the OC models in all structures are highlighted with filled circles.

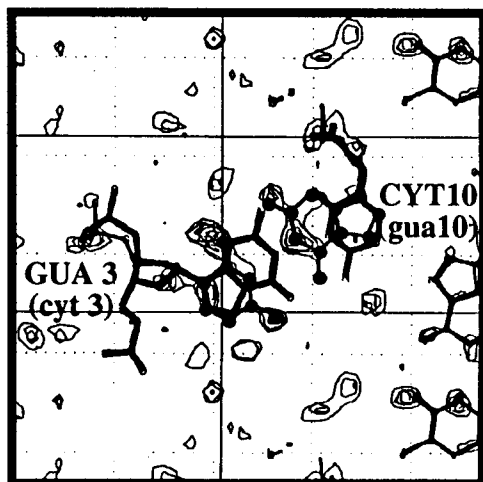
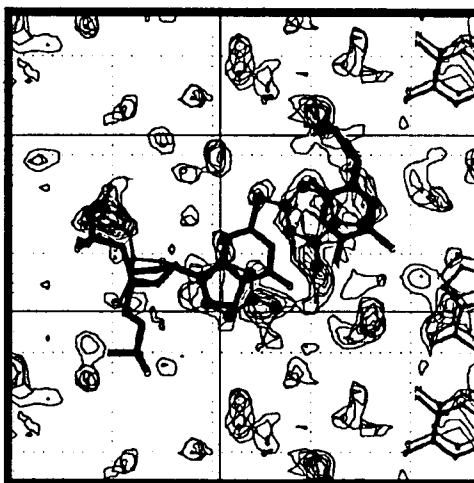
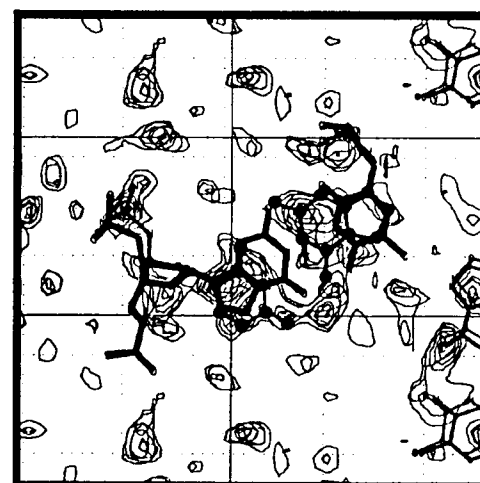
**A****67:33****B****50:50****C****43:57**

Figure 4.5



**Occupancy refinement.** To study the effect of the modifications of the duplex structure on the thermodynamics of helix packing in the crystal lattice, a quantitative measure of the distribution of the hexamer duplex between the two packing orientations was obtained with a refinement of the occupancy of the OG and OC models in each of the structures. Following the identification of two packing orientations in the crystal structures, the two separately refined models of both packing orientations were combined and re-refined as a composite model (see steps 2 and 3 of table 4.4). In the initial positional refinement of the composite model, the OG and OC models were each assigned as 50% occupied (ie. both OG and OC models contributed equally to the composite model). Following the initial positional refinement routines of the composite model, select atoms of the central dinucleotides were used in occupancy refinement of the two packing models. The base atoms of the 3-10, and 4-9 base pairs of the central dinucleotide in the hexamer duplex are the only atoms which can be absolutely assigned to either OG and OC models. Of these atoms, the O6 and N1 atoms of the guanine residues, and N4 atom of the cytosine residues were the most distinguishable between the two packing orientations (figure 4.5). Therefore, only these three atoms per base pair were considered to be reliable indicators of the strand sense in the occupancy refinement.

The final refined occupancy values for each of the two packing orientations in the crystal structures are listed on table 4.8. The distribution of the duplex in the OG orientation varied from 43% to 70% in the six crystal structures. The occupancies reflect a difference in lattice interaction energy ( $\Delta G^{\circ}_{\text{occ}}$ ) between the OG and OC orientations ranging from +0.29 to -0.50 kcal mol<sup>-1</sup> duplex<sup>-1</sup>. These values are an indicator of the relative thermodynamic stability of the duplex in the OG orientation in each of the structures. The refined occupancy values are not correlated with the time required for the appearance of crystals (compare to table 4.5). This suggests that the occupancies determined for these crystals are not likely to be solely a function of the differences in the rates of crystal growth. The observations are consistent with the generally accepted

model which considers the crystal growth process as being a dynamic equilibrium between the DNA molecules in solution and at the growing crystal surfaces (Stout & Jensen, L. H. 1989). Therefore, the observed differences in occupancies in these crystal structures allow the study of the thermodynamics of crystal packing interactions.

Comparisons of the calculated  $\Delta\Delta SFE^*_{ref}$  and  $\Delta\Delta pvdW^*_{ref}$  with the observed differences in lattice interactions ( $\Delta G^*_{occ}$ ). The observed  $\Delta G^*_{occ}$  and the calculated differences in lattice energies predicted in the initial modeling of the structures are listed on table 4.9. The observed  $\Delta G^*_{occ}$  of the 4mCC reference structure is  $0.00 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$ , therefore the  $\Delta G^*_{occ}$  values reported for the remaining five structures represent the effects associated with removing specific methyl groups from the 4mCC reference structure. The  $\Delta G^*_{occ}$  values obtained from the final refined occupancy values show that the stability of the OG orientation in these crystals follow the order 4mCC-MC5 > 4mCC-MC1 > 4mCC > GG-MC1 > GG/CC > GG-MC5. The relative significance of van der Waals and solvent interactions in determining the orientation of the duplex in the lattice are examined in comparisons of  $\Delta G^*_{occ}$  to the calculated solvation free energies ( $\Delta\Delta SFE^*_{ref}$ ) and van der Waals interaction energies ( $\Delta\Delta pvdW^*_{ref}$ ). Both calculated energies represent the differences in interactions at the crystal surface relative to those of the reference 4mCC structure. These calculated energies do not include the effects of crystal packing distortions of the DNA structure or the on lattice interactions. Therefore, the calculated and observed energies reflect the effects of differences in key interactions on the distribution in the crystal as determined by the equilibrium between the Z-DNA hexamers in solution and the growing crystal surface.

The comparisons of  $\Delta\Delta SFE^*_{ref}$  and  $\Delta G^*_{occ}$  for the six crystal structures suggest that solvent interactions play a significant role in determining the orientation of the duplex. The  $\Delta\Delta SFE^*_{ref}$  energies calculated for the hemi-methylated structures are able to account for the differences in  $\Delta G^*_{occ}$  resulting from each demethylation at the external dinucleotides of the reference 4mCC structure (figure 4.6). A linear least squares fit of

these data results in a slope of 0.645 with a y-intercept of 0.019 ( $R=0.94$ ). Thus, for the five structures with similar steric constraints imposed by METH4-10 group, the observed differences in  $\Delta G^{\circ}_{occ}$  are explainable in terms of the differences in solvent interactions relative to the reference 4mCC structure.

The presence of the METH4-10 group of the 4mCC, 4mCC-MC5, 4mCC-MC1, GG-MC5, and the GG-MC1 structures which is absent in the parent GG/CC structure is expected to favor the OC orientation. This results from an unfavorable steric interactions (equivalent to  $43 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$ ) between this methyl group and the GUA12 residue of a symmetry related hexamer duplex (see bottom of table 4.3). With the large magnitude of this steric interaction, the OG orientation is not expected in any of the modified structures. The observation of a distribution between the two orientations in all structures with the METH4-10 group further supports the significance of solvent interactions in determining the orientation of the duplexes in the lattice.

The significance of van der Waals interactions in determining the orientation of the duplex in the lattice is not supported in the comparison of the calculated  $\Delta\Delta pvdW^{\circ}_{ref}$  and the  $\Delta G^{\circ}_{occ}$ . The observed  $\Delta G^{\circ}_{occ}$  for the five structures with the METH4-10 group are not well correlated with the differences in  $\Delta\Delta pvdW^{\circ}_{ref}$  (table 4.9). In addition, the calculated  $\Delta\Delta pvdW^{\circ}_{ref}$  are relatively small (approximately 2%) in magnitude as compared to the  $43 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$  introduced by the single METH4-10 methyl group. In comparison, the effects of a single methylation on  $\Delta\Delta SFE^{\circ}_{ref}$  total  $0.6 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$ . The calculated differences in  $\Delta\Delta SFE^{\circ}_{ref}$  for the 5 structures range in values between 0.3 and  $-0.3 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$  and thus account for the solvent effects equivalent to a full methyl group. This comparison suggests that the differences in  $\Delta\Delta pvdW^{\circ}_{ref}$  are less able to explain the differences in the distribution of the duplex in these crystal structures.

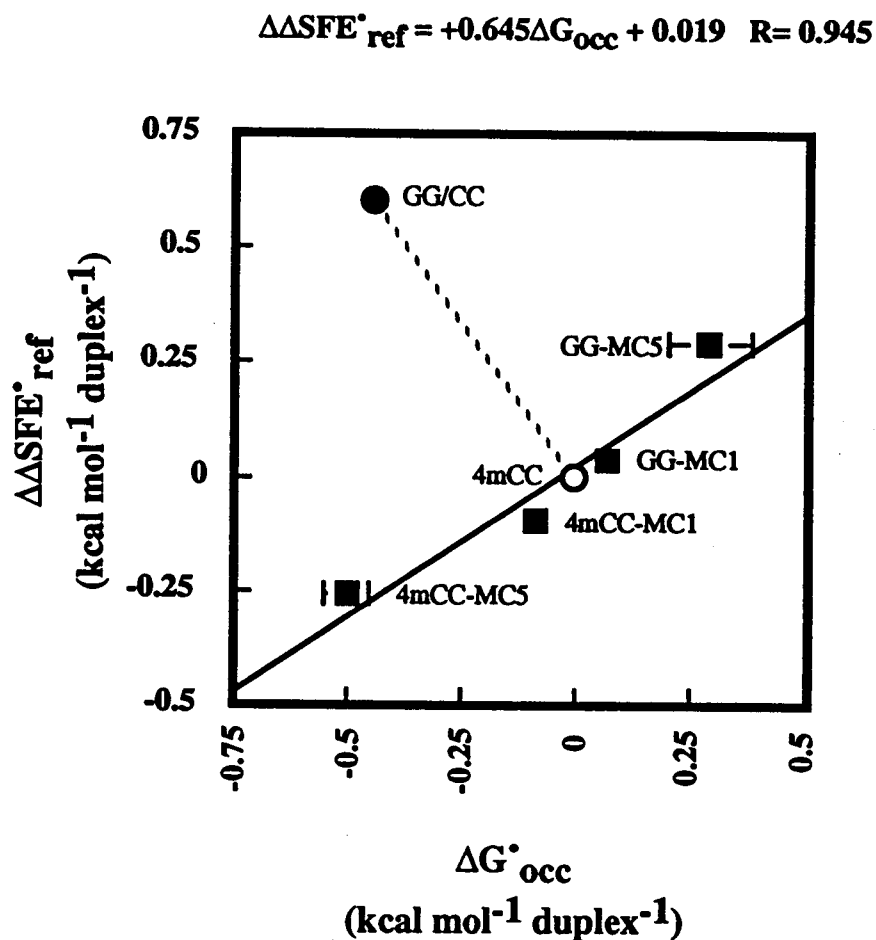


Figure 4.6. Plot comparing the difference in solvation free energies between the OG and OC models relative to the reference 4mCC structure and observed difference in lattice interaction energies ( $\Delta G^*_{\text{occ}}$ ). The open circle represents the energies for the reference structure. The closed circle represents the energies for the parent GG/CC structure. The closed squares represent the energies for the remaining hemi-methylated structures.

The  $\Delta\Delta SFE^{\circ}_{\text{ref}}$  for the GG/CC structure is not consistent with the trend relating the  $\Delta\Delta SFE^{\circ}_{\text{ref}}$  and  $\Delta G^{\circ}_{\text{occ}}$  established by the reference and hemi-methylated structures. The inconsistency between the calculated  $\Delta\Delta SFE^{\circ}_{\text{ref}}$  for this structure reflect the differences in crystal packing interactions associated with the METH4-10 group. The effects of removing the steric clash on the stability of the OG orientation is illustrated with dashed line between the plotted energies for the GG/CC and 4mCC structures (figure 4.6). The differences in  $\Delta\Delta SFE^{\circ}_{\text{ref}}$  is  $+0.6 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$  and the difference in  $\Delta G^{\circ}_{\text{occ}}$  is  $+0.44 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$  between the parent and reference structures. Thus, based on the relationship between  $\Delta\Delta SFE^{\circ}_{\text{ref}}$  and  $\Delta G^{\circ}_{\text{occ}}$ , the destabilizing effect of this methyl group is estimated to be  $+0.74 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$ . This energy likely reflects the penalty associated with the steric clash of the METH4-10 group, and is greatly reduced from the expected  $43 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$  calculated in the initial modeling of the structures. This comparison shows that the van der Waals energies were overestimated in the modeling which results from the inability to predict the flexibility of the duplexes.

The details of each crystal structure provide information on the effect of crystal packing forces and short range interactions on the DNA duplex structure. The crystal lattice interactions are observed to differ for the structures with fully *versus* hemi-methylated external dinucleotides. The effect of the METH4-10 methyl group on the crystal lattice interactions and the internal stability of the asymmetric duplex are discussed in a comparison of the reference 4mCC and the parent GG/CC structures. The effect of removing methyl groups from MCY residues of the external dinucleotides will be discussed with comparisons of reference 4mCC structure with the four hemi-methylated structures. The hemi-methylation effects will be discussed in terms of the differences in packing interactions and helical characteristics observed in these structures.

Table 4.9. Calculated differences in internal and lattice van der Waals energies, and solvation free energies between the OG and OC orientations for the crystal structures of the six asymmetric hexamer duplexes.

Structure	$\Delta\Delta\text{SFE}^*_{\text{ref}}$	$\Delta\Delta\text{pvdW}^*_{\text{ref}}$	$\Delta\text{SFE}$	$\Delta\text{vdW}_{\text{int}}$	$\Delta\text{pvdW}$	$\Delta\text{vdW}_{\text{Tot}}$	$\Delta\text{G}^*_{\text{occ}}$
GG/CC	+0.60	-43.46	+2.2	+2.7	-4.1	-1.4	-0.44
4mCC	0.00	0.00	-3.6	+11.1	+6.7	+17.8	0.00
GG-MC1	0.04	+0.94	-2.1	-0.5	-2.9	-3.4	+0.07
GG-MC5	0.30	+0.85	+2.3	+6.3	-5.5	+0.8	+0.29
4mCC-MC1	-0.10	-0.65	+0.4	-5.5	-5.5	-11.0	-0.09
4mCC-MC5	-0.26	-0.73	+0.2	+4.1	-12.0	-7.9	-0.50

$\Delta\Delta\text{SFE}^*_{\text{ref}}$  and  $\Delta\Delta\text{pvdW}^*_{\text{ref}}$  are the differences in energies (OG minus OC) relative to the 4mCC reference structure calculated in the original modelling of the structures (table 4.3).

$\Delta\text{SFE}$  - The difference in SFE (OG minus OC) in the crystal lattice.

$\Delta\text{vdW}_{\text{int}}$  - The difference in internal van der Waals energies (OG minus OC).

$\Delta\text{pvdW}$  - The difference in lattice van der Waals energies (OG minus OC).

$\Delta\text{vdW}_{\text{Tot}}$  -  $\Delta\text{vdW}_{\text{int}} + \Delta\text{pvdW}$ .

$\Delta\text{G}^*_{\text{occ}}$  - The difference in free energy between OG and OC (OG over OC) as calculated from the differences in the distribution of the hexamer duplex in the crystal lattice of each structure.

Comparison of the parent GG/CC and the reference 4mCC structures. The occupancy of OG orientation decreases from 68% to 50% in the GG/CC *versus* 4mCC structures respectively showing that the METH4-10 methyl group has an overall destabilizing effect on the OG packing orientation. This result is qualitatively consistent with the initial modeling which predicted the destabilization of the OG orientation due to unfavorable van der Waals interactions.

The existence of unfavorable van der Waals contacts in the 4mCC crystal structure are supported in a comparison of the internal and packing interactions calculated for the parent GG/CC and reference 4mCC crystal structures (table 4.9). The OG orientation of the 4mCC structure has a relatively unfavorable ( +6.7 kcal mol<sup>-1</sup> duplex<sup>-1</sup>) positive van der Waals lattice interaction energy ( $\Delta pvdW$ , OG minus OC). The decreased magnitude of the van der Waals packing interaction energies in the 4mCC structure (less than 43 kcal mol<sup>-1</sup> duplex<sup>-1</sup>) when comparing initial model calculations (table 4.9) to the energies derived from the crystal structure indicate the DNA structure has adjusted to accommodate the large unfavorable steric interaction. The difference in internal van der Waals energies of the OG packing orientation in this 4mCC structure *versus* the parent GG/CC structure increases (2.7 to 11.1 kcal mol<sup>-1</sup> duplex<sup>-1</sup>). This supports the existence of the steric penalty against the OG orientation in the 4mCC structure (table 4.10).

It could be argued that the inherent structure of the 4mCC duplex somehow resulted from the avoidance of the unfavorable contact in the crystal packing. However, a comparison of the structure of the MCY4 residue in the OC *versus* MCY10 residue in the OG orientation indicate structural perturbations result from lattice interactions when the hexamer is in the OG orientation. The methyl group environment of the MCY4 residue in the OC orientation is relatively free of crystal lattice contacts and therefore is expected to represent the unperturbed structure of this residue. In the OG orientation, the methyl group of the MCY10 residue is observed to be in vdW contact with the N7 atom of GUA12 (carbon to nitrogen distance 2.92 Å). When the MCY4 residue in the OC

orientation is modeled in the context of the MCY10 residue in the OG orientation, an unfavorable steric interaction between the methyl group and the N7 of the adjacent GUA12 is created (MCY4 methyl carbon to GUA12 N7 nitrogen distance is 2.42 Å). This closer approach results in an estimated 10 kcal mol<sup>-1</sup> residue<sup>-1</sup> increase in the van der Waals packing interaction energy. The evidence suggests that methylation creates an unfavorable contact in the OG orientation which is accommodated at the expense of less favorable intramolecular interactions and a decreased occupancy of the OG orientation.

Comparisons of the 4mCC and the hemi-methylated structures. The remaining modified structures examine the effect of removing a single methyl group from the external dinucleotides from the reference 4mCC structure. The structures with hemi-methylated d(CG) dinucleotides were found to have different helical characteristics as compared to the 4mCC structure. The structural differences were observed to affect the packing interactions in the crystal lattice. Removing a methyl group from the external dinucleotide in all cases resulted in negative differences in  $\Delta pvdW$  energies and lower  $\Delta vdW_{int}$  as compared to the reference 4mCC structure (table 4.9). These differences in the packing interaction energies (table 4.9), and the internal energies indicate that each structure adapted to the introduced structural perturbations and optimized the lattice interactions in different manners. These differences are apparent in comparisons of the atomic positions and helical parameters between the fully and hemi-methylated structures.

The differences in the structures is shown with a comparison of the RMS deviation between the base atoms for the dinucleotides in the hexamer duplex (figure 4.7). Panels A and C are comparisons of the OG models, and panels B and D are comparisons of the OC models. The open circles in all panels are comparisons between the GG/CC and 4mCC structures which have fully methylated external dinucleotides. From panels A and C, the comparisons of OG models of the fully methylated structures show that the deviations range from 0.36 to 0.43 Å over the three dinucleotides. This comparison



shows that the introduced steric collision between residue MCY10 and GUA12 was accommodated in the fully methylated structures with relatively small structural changes

In contrast, greater structural differences are apparent when comparing the 4mCC structure to the hemi-methylated structures. This is evident with larger deviations ( $> 0.60$  Å) associated with the demethylation of the 5-8/6-7 dinucleotide (panel A) and for the demethylation of the 1-12/2-11 dinucleotide ( $> 0.45$  Å) (panel C). These comparisons indicate greater structural differences at both external dinucleotides of the hemi-methylated OG structures with smaller differences at the central out-of-alternation dinucleotide.

In similar comparisons of the OC orientation of the 4mCC and hemi-methylated structures (panels B and D), the hemi-methylated structures differ from fully methylated structures based on larger deviations at the external dinucleotides. Steric collisions are not expected with the hexamer in the OC orientations, thus these differences likely reflect the inherent differences in the structures of the hemi- and fully methylated duplexes.

The structural adjustments are apparent at both external dinucleotides whether fully or hemi-methylated. This is not unexpected when considering the interactions within the crystal lattice. In the crystal lattice, the hexamer duplexes are stacked end to end

Figure 4.7. Comparison of the final structures of the fully and hemi-methylated structures with root mean square (RMS) deviations calculated for dinucleotides of the OG and OC models. Panels A and C are comparison of the OG packing models, and panels B and D are comparisons of the OC packing models. Panels A and B are comparisons showing the effect of hemi-methylation at the 5-8/6-7 dinucleotide. Panels C and D are comparisons showing the effect of hemi-methylation at the 1-12/2-11 dinucleotide. All comparisons are with the reference 4mCC structure and therefore represent the effects of demethylation of specific MCY residues in the asymmetric duplex. The filled squares (■) and cross filled squares (▣) are comparisons of the 4mCC-MC1 and GG-MC1 structures with the reference 4mCC structure. The filled diamonds (◆) and cross filled diamonds (◈) are comparisons of the GG-MC5 and 4mCC-MC5 structures with the reference 4mCC structure. The open circles (○) in all panels are the comparison between the parent GG/CC and reference 4mCC structures. The RMS deviation was calculated between the atomic coordinates of non-hydrogen base atoms of the dinucleotides and excluded the atoms of the methyl groups. The RMS deviations were calculated using the X-PLOR refinement routine (Brunger, 1992).

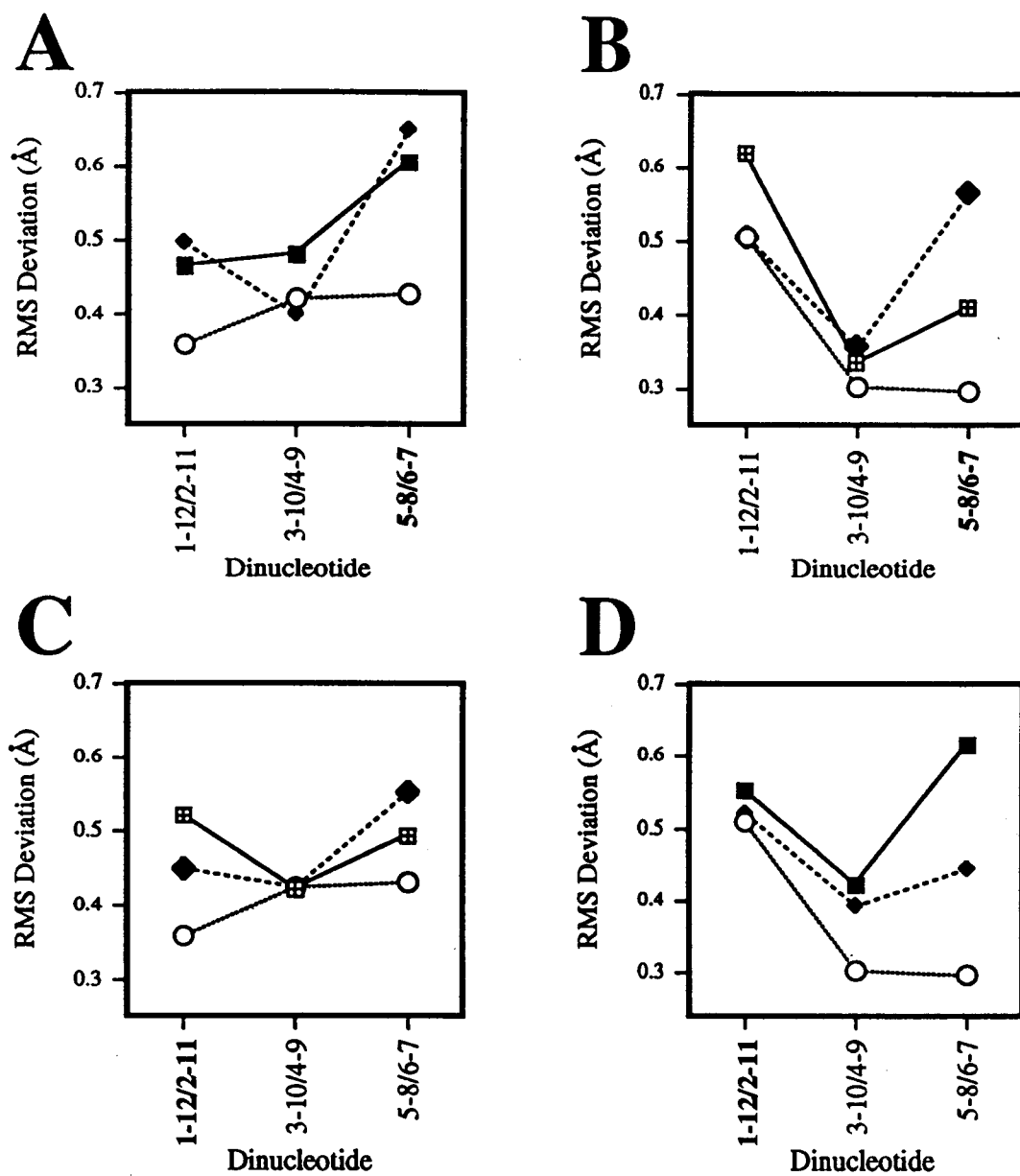


Figure 4.7

generating continuous helices of Z-DNA (figure 4.2). In this packing arrangement, the opposing ends of the hexamer duplex are interacting through strong base stacking interactions. Thus, the consequences of adjustments at one end of the hexamer duplex are expected at both ends of the duplex. The structural modifications of either external dinucleotide are likely to significantly affect other non-stacking lattice interactions. In particular, the steric interaction between the METH4-10 group in the OG orientation and the end residue GUA12 of the external dinucleotide is expected to be very sensitive to any adjustments at this interface. The differences in the duplex structure influence the ability of each structure to optimize the lattice interactions. This is consistent with the observed differences in  $\Delta pvdW$  between the fully and hemi-methylated structures (table 4.9). These adjustments are expected to be apparent in changes in the helical parameters.

The differences between the fully methylated structures, and the hemi-methylated structures are most apparent in the differences in the twist angles between the structures of the OG models (table 4.10). The average helical twist for all hemi-methylated CG steps in the OG models is  $-15.7 \pm 0.3^\circ$  as compared to  $-11.6 \pm 1.0^\circ$  for the same step of the fully methylated structures. Thus, the end base pairs of the hemi-methylated dinucleotides are underwound by approximately  $4^\circ$  as compared to the fully methylated dinucleotides of the OG models.

In the OC models, a large difference observed at the CG base step between the 5-8/6-7 base pair between the fully and hemi-methylated structures. The average twist angle for this step is  $-17.8 \pm 0.7^\circ$  for the fully methylated structures *versus*  $-12.5 \pm 1.3^\circ$  for the hemi-methylated structures. Thus, hemi-methylation results in an overwinding of the 5-8/6-7 base pairs by approximately  $5^\circ$  relative to the fully methylated OC structures. It is not unexpected that these parameters should be different when comparing the hemi- and fully methylated structures. The twist angles for the crystal structures of the fully methylated  $d(m^5CG)_3$  hexamer, and the unmethylated  $d(CG)_3$  show characteristic differences for these dinucleotides (Fujii et al., 1982). The twist angle for CG step was

$-13 \pm 1^\circ$  and  $-46 \pm 1^\circ$  for the GC base steps of the fully methylated structure versus  $-8 \pm 1^\circ$  and  $-51 \pm 2^\circ$  for the same steps in the unmethylated structure. In addition, the structure and thermodynamic stability of hemi-methylated d(CG) dinucleotides as Z-DNA have been shown to be unique in comparison to either fully or unmethylated d(CG) dinucleotides (Bononi & Ho; unpublished).

In summary, two general methyl effects were observed in the structures of the asymmetric duplexes. The first involved the accommodation of a steric clash associated with the METH4-10 group through decreased occupancy of the OG orientation, and by internal structural perturbations. The second effect was observed in the demethylation of the d(m<sup>5</sup>CG) external dinucleotides. The demethylation of these dinucleotides allowed the optimization of van der Waals lattice interactions in the presence of the METH4-10 methyl group. These adjustments are apparent as structural differences between the hemi-methylated and fully methylated duplexes.

Table 4.10. Twist angles between base pair steps calculated for the OG and OC models of the six crystal structures. The twist angles in bold type are the angles in which at least one of the base pairs in the step is a the demethylated cytosine residue relative to the 4mCC structure. The twist angles were calculated with the NASTE program (Basham unpublished)

Base Step	GG/CC	4mCC	GG-C1	GG-C5	4mCC-C1	4mCC-C5
OG models:						
CG/GC <sup>1</sup>	-12.1	-10.1	<b>-16.2</b>	-11.5	-13.8	<b>-15.4</b>
GC/GC	<b>-49.2</b>	-49.1	-47.3	-50.3	-48.0	<b>-44.1</b>
GC/GC	<b>-11.6</b>	-12.9	-12.7	-11.7	-12.8	-10.7
GC/CG	-48.4	-46.8	-43.7	<b>-46.5</b>	-44.5	-51.1
CG/GC <sup>1</sup>	-12.1	-11.9	-15.3	<b>-15.7</b>	<b>-15.6</b>	-14.0
OC models:						
CG/GC	-12.7	-13.4	-10.2	<b>-15.0</b>	<b>-12.3</b>	-11.4
CG/CG	<b>-44.2</b>	<b>-43.2</b>	<b>-42.2</b>	<b>-41.2</b>	<b>-47.0</b>	
CG/CG	<b>-12.8</b>	-12.8	-12.8	-10.8	-12.0	-15.8
CG/GC	<b>-45.6</b>	-49.6	-51.0	-50.2	-52.2	<b>-44.8</b>
CG/GC <sup>2</sup>	-18.3	-17.3	<b>-12.7</b>	-12.0	-11.2	<b>-14.2</b>

1. Ave. twist angle for the fully methylated CG/GC step for OG models is  $-11.6 \pm 1.0^\circ$ , versus  $-15.7 \pm 0.3^\circ$  for the *hemi*-methylated steps.
2. Ave. twist angle for the fully methylated CG/GC step (base pairs 5-8/6-7) for the OC models is  $-17.8 \pm 0.7^\circ$  versus  $-12.5 \pm 1.3^\circ$  for the *hemi*-methylated steps.

## Discussion

The experiments in this chapter utilize a novel system for examining the contribution of van der Waals and solvent interactions in DNA assembly. These two competing forces were studied in the crystal structures of six asymmetric Z-DNA hexamer duplexes.

The crystal structures of five modified sequences of the asymmetric hexamer duplexes were solved to near atomic resolution ( $> 1.6 \text{ \AA}$ ). In these crystal structures, the asymmetric hexamer duplexes adopts two discrete and distinguishable orientations in the crystal lattice. The distribution between the two orientations in the five modified duplexes and the parent duplex crystal were found to range from 70:30 to 43:57 in the ratio of OG to OC orientation and reflect a range of  $-0.50$  to  $+0.29 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$  difference in interaction energy ( $\Delta G^*_{\text{occ}}$ ) between the two orientations. These variable occupancies were independent of the time required for the appearance of crystals in the crystallization experiments, and were therefore considered to reflect the relative thermodynamic stability of each orientation.

The large difference in  $\Delta\Delta pvdW^*_{\text{ref}}$  between the parent structure and the remaining structures shows the large steric penalty against the OG orientation due to the METH4-10 methyl group. The large magnitude of the steric penalty ( $\sim 43 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$ ) suggests that if the distribution of the duplex in the crystal lattice is dependent only on the differences in van der Waals interactions, that there would not be a distribution between the two orientations in any of the modified structures. The duplex in the OG orientation is expected to never be observed in the structures with the METH4-10 methyl group. The fact that a distribution between the two orientations is found in every structure reflects the relative significance of the solvent interactions in determining the orientation of the duplex in the lattice. The unfavorable solvent interactions associated with the exposure of the

METH4-10 methyl group were significant enough to force the duplex to accept what is expected to be a very large unfavorable steric interaction as well as compromise the internal interactions.

The  $\Delta G^{\circ}_{\text{occ}}$  values for the reference and hemi-methylated structures agrees with the predictions based on the  $\Delta\Delta\text{SFE}^{\circ}_{\text{ref}}$  calculated from the models of each of the structures (table 4.8). The  $\Delta G^{\circ}_{\text{occ}}$  is correlated with the  $\Delta\Delta\text{SFE}^{\circ}_{\text{ref}}$  for the hemi-methylated and 4mCC structures suggesting that the orientation of the duplex is determined primarily by the differences in solvent interactions (figure 4.6). A linear least squares fit of these data results in a slope of 0.645 with a y-intercept of 0.019 ( $R=0.94$ ). The correlation between the calculated  $\Delta\Delta\text{SFE}^{\circ}_{\text{ref}}$  and the observed  $\Delta G^{\circ}_{\text{occ}}$  for these structures further supports the significance of solvent interactions in the assembly of Z-DNA hexamers in the process of crystallization.

The details of the 4mCC structure provide evidence that the METH4-10 methyl group introduces an unfavorable steric interaction in the OG orientation involving the methyl group at residue MCY10 and the symmetry related GUA12 residue. The introduction of the METH4-10 group results in a decrease in the relative stability of the OG orientation by  $0.44 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$ . The existence of an unfavorable steric interaction (approximately  $10 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$ ) is confirmed with the modeling of the unperturbed OC packing model in the OG packing orientation. In addition, perturbations to the OG packing orientations are apparent with the increase in van der Waals packing and structural penalties ( $\Delta\text{vdW}_{\text{Tot}} = +17.8 \text{ kcal mol}^{-1} \text{ duplex}^{-1}$ ).

In comparing the packing interactions between the reference 4mCC structure and hemi-methylated structures, differences were observed in the van der Waals interaction and internal energies. The structural differences between the fully and hemi-methylated duplexes and are apparent as larger RMS deviations in comparisons of the fully versus hemi-methylated d(CG) external dinucleotides. The RMS deviations reflect differences in the characteristic twist angles observed for the fully versus hemi-methylated structures.



These differences in structure and interaction energies reflect the variable ability of each structure to accommodate the steric clash associated with the METH4-10 group and optimize the lattice interactions.

The comparisons of  $\Delta G^{\circ}_{occ}$  and the calculated solvation free energies based on the initial modeling of the structures suggest that the solvent interactions significantly influence the distribution of asymmetric duplex in the crystal lattice at the surface of the growing crystal. In addition, the comparisons between  $\Delta\Delta SFE^{\circ}_{ref}$  and  $\Delta G^{\circ}_{occ}$  provide a relationship between solvent effects and observed stability of the OG orientation which can be tested. The effect of removing two methyl groups of external dinucleotides from the 4mCC structure on the stability of the OG orientation are predictable from calculated differences in SFEs (figure 4.6). If the effect of removing two methyl groups are in fact additive, the range of the  $\Delta G^{\circ}_{occ}$  are expected to reflect the effects associated with solvating two methyl groups. In addition, a similar relationship (slope) between  $\Delta\Delta SFE^{\circ}_{ref}$  and  $\Delta G^{\circ}_{occ}$  is expected with the removal of the methyl groups from the parent GG/CC structure. These studies would simply test the relationship between calculated  $\Delta\Delta SFE^{\circ}_{ref}$  and the observed  $\Delta G^{\circ}_{occ}$ .

Empirical evidence for the significance of the hydrophobic effect could be obtained by monitoring the distribution of the duplex as a function of the presence of various anions. The efficacy of the hydrophobic effect is expected to follow the order acetate >  $Cl^{-}$  >  $Br^{-}$  >  $I^{-}$  (Collins & Washabaugh, 1985; Creighton, 1984). It is expected that the occupancy of the orientation which is favored by solvent effects would increase. This would result in wider range of  $\Delta G^{\circ}_{occ}$  (for the sequences studied) and therefore a shallower slope in the plot relating  $\Delta\Delta SFE^{\circ}_{ref}$  to  $\Delta G^{\circ}_{occ}$  for anions which enhance the hydrophobic effect (acetate). A steeper slope in this plot would be expected in the presence of the anions  $Br^{-}$  and  $I^{-}$ .

## Chapter 5

### Discussion

The studies presented in this thesis have examined the relative significance of the hydrophobic effect or solvent interactions in two relatively simple macromolecular systems. The contribution of solvent interactions to the stability of DNA secondary structure was examined in the well characterized B- to Z-DNA transition. The observed relative stability of APP sequences as Z-DNA was shown to be predictable with a quantitative analysis of the differences in solvent interactions between the B and Z-conformations. Thus, these studies indicate that solvent interactions and the hydrophobic effect play a significant role in determining the stability of Z-DNA. The significance of solvent interactions in DNA assembly was studied in the crystal structures of six asymmetric Z-DNA hexamer duplexes. This final study provided evidence that the packing of the duplexes, which is determined by specific DNA-DNA interactions, is also strongly influenced by solvent interactions.

Solvent interactions were estimated as solvation free energies derived from calculated solvent-accessible surface areas. The calculated differences in solvation free energies for sequences in the B- versus Z-DNA conformation were compared to experimentally determined transition energies. The comparison showed that the calculated and observed energies were of the same order of magnitude. In addition, the calculated SFEs are able to account for the trend observed for the experimentally determined transition energies ( $\Delta G^{\circ}T(B-Z)$ ). The comparison of the  $\Delta G^{\circ}T(B-Z)$  for these dinucleotides as Z-DNA and the calculated differences in SFEs provided a relationship for predicting the relative stability of APP sequences as Z-DNA. This relationship suggests that solvent interactions account for approximately 70% of the observed sequence dependent stability of Z-DNA. In addition, the predicted stability of

APP hexanucleotide sequences as Z-DNA were shown to be related to the cationic strength required for that sequence to crystallize as Z-DNA. Thus, the ability of APP sequences to adopt the Z-conformation as predicted by SFE calculations are testable with comparisons of the cationic conditions required for Z-DNA crystal formation. The relationship between SFEs and the stability of APP sequences as Z-DNA have been used to examine substituent effects the stability of Z-DNA.

The major groove methylation effects were studied in comparisons of differences in the relative stabilities of d(TA) *versus* d(UA) dinucleotides. The methylation effect is attributed to the differences in the environments in the B- and Z-DNA conformations. In the B conformation, the methyl group is highly exposed in the major groove surface. In the Z conformation, the methyl group sits in a pocket which decreases the overall exposure of the methyl group, and buries both hydrophobic and hydrophilic base and ribose surfaces. The differences in the pocket characteristics accounted for stabilization effect of the methyl group of d(m<sup>5</sup>CG) dinucleotides and the destabilizing effect of the methyl group of the d(TA) dinucleotides. The contrary methylation effects were found to be related to the differences in the methyl environments between d(m<sup>5</sup>CG) versus d(TA) dinucleotides in the Z conformation. The methyl group of the d(m<sup>5</sup>CG) dinucleotide buries a larger hydrophobic base surface area, and a smaller in hydrophilic ribose oxygen surface area relative to the methyl group of d(TA) dinucleotide. The removal of the d(TA) methyl group to form d(UA) dinucleotides was predicted to stabilize the Z conformation by increasing the exposure of hydrophilic base and ribose groups in the Z-DNA major groove. This prediction was tested and supported in crystallization experiments with the d(UA) containing sequence d(m<sup>5</sup>CGUAm<sup>5</sup>CG). This sequence was found to crystallize under 2-fold lower conditions as compared to the analogous d(m<sup>5</sup>CGTAm<sup>5</sup>CG) sequence suggesting that d(UA) dinucleotides are more stable as Z-DNA than d(TA) dinucleotides.

The SFE analysis of the stability of Z-DNA was extended to include minor groove substituent effects. The contribution of the N2 amino group of the purine bases to the stability of Z-DNA was examined by de-aminating the guanine base in d(CG) dinucleotides to form inosine bases in d(CI) dinucleotides, and conversely the amination of the adenine base in d(TA) dinucleotides to form 2-aminoadenine bases in d(TA') dinucleotides. Based on the analysis of solvent accessible surfaces the N2 amino group is shown to decrease the exposure of hydrophobic surfaces, and increase the exposure of hydrophilic surfaces in the Z-conformation. Thus, the SFE calculations predicted that the presence of the N2 amino group on purine bases should stabilize the Z-conformation. The comparison of crystallization conditions which yielded diffraction quality crystal of sequences containing d(TA) versus d(TA') and d(CG) versus d(CI) dinucleotides supported the predictions of sequence dependent Z-DNA stability based on SFE calculations.

The studies presented in chapter 4 described a novel system for studying the balance of forces in DNA-DNA interactions. The crystallographic study examined the balance of opposing van der Waals interactions and the solvent interactions in determining the orientation of Z-DNA hexamers in the  $P2_12_12_1$  crystal lattice. The distribution of the duplex between two discrete and distinguishable orientations in the crystal structures of six asymmetric duplexes was analyzed quantitatively. The observed distributions in these structures reflected the difference in lattice interaction energies between the two packing orientations ( $\Delta G^\circ_{\text{occ}}$ ). The experimentally derived  $\Delta G^\circ_{\text{occ}}$  were compared to differences in calculated SFEs ( $\Delta \Delta \text{SFE}^\circ_{\text{ref}}$ ) and van der Waals ( $\Delta \Delta \text{pvdW}^\circ_{\text{ref}}$ ) lattice interaction energies. The calculated energies reflect the differences in potential lattice interactions at the growing crystal surface. The comparisons of the observed  $\Delta G^\circ_{\text{occ}}$  and the calculated  $\Delta \Delta \text{SFE}^\circ_{\text{ref}}$  and  $\Delta \Delta \text{pvdW}^\circ_{\text{ref}}$  for the six structures indicate that solvent interactions were significant enough to overcome what was expected to be a very large steric interaction. This supports the significance of solvent interactions in determining the orientation of the

duplex in the crystal lattice. In addition, a linear relationship was observed between the calculated  $\Delta\Delta SFE^{\circ}_{\text{ref}}$  and the experimentally determined  $\Delta G^{\circ}_{\text{occ}}$  for five structures which have similar steric crystal packing constraints. The differences in van der Waals interactions ( $\Delta\Delta pvdW^{\circ}_{\text{ref}}$ ) did not correlate with the observed  $\Delta G^{\circ}_{\text{occ}}$ . These results suggest that solvent interactions play a significant role in determining the orientation of the Z-DNA hexamers in the crystal lattice. Thus, these studies indicate that solvent interactions significantly influence the assembly of Z-DNA hexamers in the process of crystallization.

In conclusion, the computational and experimental analyses of the B- to Z-DNA transition and Z-DNA crystallization presented in this thesis have provided relatively clear support for the significance of solvent interactions in DNA folding and DNA assembly. The correlation of calculated SFEs with the observed sequence dependent stability of Z-DNA and with the differences in the helix packing in crystals of Z-DNA suggests that solvent interactions contribute significantly to the folding and stability macromolecules.

## Bibliography

- Alden, C. J., & Kim, S. H., (1979) "Solvent-accessible Surfaces of Nucleic Acids." *J. Mol. Biol.* **132**, 411-434.
- Anfinsen, C.B. (1973) "Principles that Govern the Folding of Protein Chains." *Science* **181**, 223-230.
- Anshelevich, V. V., Vologodskii, A. V., & Frank-Kamenetskii, M.D. (1988) "" *J. Biomol. Struct. Dynam.* **6**, 247-259.
- Arnott, S., Smith, P. J. C., & Chandrasekaran, D. L. (1976) in *Handbook of Biochemistry and Molecular Biology* (Fasman, G. D., Ed.) pp 411-422, CRC Press, Cleveland, OH.
- Azorin, F., Hahn, R., & Rich, A. (1984) "Restriction Endonucleases Can Be Used to Study B-Z Junctions in Supercoiled DNA." *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5714-5718.
- Behe, M. & G. Felsenfeld (1981) "Effects of Methylation on a Synthetic Polynucleotide: The B-Z transition in Poly(dG-dm<sup>5</sup>C)." *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1619-1623.
- Brünger, A (1992) *X-PLOR Manual, Version 3.1* Yale University, New Haven, USA.
- Chen, H. H., Behe, M. J., & Rau, D. C. (1984) "Critical Amount of Oligovalent Ion Binding Required for the B-Z Transition of poly(dG-dm<sup>5</sup>C)." *Nucleic Acids Res.* **12**, 2381-2389.
- Chevrier, B., Dock, A. C., Hartmann, B., Leng, M., Moras, D., Thuong, M. T., & Westhof, E. (1986) "Solvation of the Left-handed Hexamer d(5BrC-G-5BrC-G-5BrC-G) in Crystals Grown at Two Temperatures." *J. Mol. Biol.* **188**, 707-719.
- Cieplak, P., & Kollman, P. A. (1988) "Calculation of the Free Energy of Association of Nucleic Acid Bases in Vacuo and Water Solution" *J. Am. Chem. Soc.* **110**, 3734-3739.
- Coll, M., Fita, I., Lloveras, J., Subirana, J.A., Bardella, F., Huynh-Dinh, T., & Igoleu, J. (1988) "Structure of d(CACGTG), a Z-DNA Hexamer Containing AT Base Pairs." *Nucleic Acids Res.* **16**, 8695-8705.
- Coll, M., Saal, D., Frederick, C. A., Aymami, J., Rich, A. & Wang, A. H.-J. (1989) "Effects of 5-fluorouracil/guanine wobble base pairs in Z-DNA: Molecular and Crystal Structure of d(CGCGFG)." *Nucleic Acids Res.* **17**, 911-923.
- Coll, M., Wang, A. H. J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1986) "Crystal Structure of a Z-DNA Fragment Containing Thymine/2-Aminoadenine Base Pairs." *J. Biomol. Struct. Dynam.* **4**, 157-172.
- Collins, K. D. & Washabaugh, M. W. (1985) "The Hofmeister Effect and the Behaviour of Water at Interfaces." *Q. Rev. Biophys.* **18**, 323-422.

- Connolly, M. L. (1983) "Solvent-Accessible Surfaces of Protein and Nucleic Acids." *Science* **221**, 709-713.
- Creighton, T. E. (1984) *Proteins: Structures and Molecular Principles* W. H. Freeman and Co., New York.
- Dang, L. X., D. A. Pearlman & Kollman, P. A. (1990) "Why do A•T Base Pairs Inhibit Z-DNA Formation." *Proc. Natl. Acad. Sci. USA* **87**, 4630-4634.
- Depew, R. E., & Wang, J. C. (1975) "Conformational Fluctuations of DNA Helix." *Proc. Nat. Acad. Sci. U.S.A.* **72**, 4275-4279.
- Devarajan, S., & Shafer, R. H. (1986) "Role of Divalent Cations on DNA Polymorphism Under Low Ionic Strength Conditions." *Nucleic Acids Res.* **14**, 5099-5109.
- Dickerson, R. E. (1992) "DNA Structure from A to Z." in *Methods in Enzymology*. **211**, (Lilley, D. M. J. & Dahlberg, J. E. eds.) pp 67-111.
- Dickerson, R. E., & Drew, H. R. (1981) "Structure of a B-DNA Dodecamer II: Influence of Base Sequence on Helix Structure." *J. Mol. Biol.* **149**, 761-786.
- Dill, K. A. (1990) "Dominant Forces in Protein Folding." *Biochemistry* **29**, 7134-7155.
- Drew, H. R., & Dickerson, R. E. (1981) "Structure of a DNA dodecamer. III. Geometry of Hydration" *J. Mol. Biol.* **151**, 535-556.
- Edelhoch, H. & Osborne, J. C. Jr. (1976) "The Thermodynamic Basis of the Stability of Proteins, Nucleic Acids, and Membranes." *Adv. Prot. Chem.* **30**, 183-250.
- Egli, M.; Williams, L. D., Gao, Q. & Rich, A. (1991) "Structure of the Pure-Spermine Form of Z-DNA (Magnesium Free) at 1-Å Resolution." *Biochemistry* **30**, 11388-11402.
- Eisenberg, D. & MacLachlan, A. D. (1986) "Solvation Energy in Protein Folding and Binding." *Nature* **319** 199-203.
- Ellison, M. J., Feigon, J., Kelleher, R. J., III, Wang, A. H. -J., Habener, J. F., & Rich, A. (1986) "An Assessment of the Z-DNA Forming Potential of Alternating dA-dT Stretches in Supercoiled Plasmids." *Biochemistry* **25**, 3648-3655.
- Ellison, M. J., Fenton, M. J., Ho, P. S., & Rich, A. (1987) "Long-range Interactions of Multiple DNA Structural Transitions Within a Common Topological Domain." *EMBO J.* **6**, 1513- 1522.
- Feuerstein, B. G., Marton, L. J., Keniry, M. A., Wade, D. L., & Shafer, R. H. (1985) "New DNA Polymorphism: Evidence for a Low Salt, Left-handed Form of Poly (dG-m<sup>5</sup>dC)." *Nucleic Acids Res.* **13**, 4133-4141.
- Fisher, E. F., & Caruthers, M. H. (1979) "Studies on Gene Control regions XII: The Functional Significance of a Lac Operator Constitutive Mutation." *Nucleic Acids Res.* **7**, 401-416.

- Franklin, R. E., & Gosling, R. G. (1953a) "The Structure of Sodium Thymonucleate Fibres I: The Influence of Water Content." *Acta Cryst.* **6**, 673-677.
- Franklin, R. E., & Gosling, R. G. (1953b) "The Structure of Sodium Thymonucleate Fibres II: The Cylindrical Symmetrical Patterson Function." *Acta Cryst.* **6**, 678-685.
- Frank-Kamenetskii, M. D., & Vologodskii, A. V. (1984) "Thermodynamics of the B-Z Transition in Superhelical DNA." *Nature* **307**, 481-482.
- Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J., & Rich, A. (1982) "Molecular Structure of d(m<sup>5</sup>CG)<sub>3</sub>: The Role of the Methyl Group on 5-methyl Cytosine in Stabilizing Z-DNA." *Nucleic Acids Res.* **10**, 7879-7892.
- Gao, Y. -G. Sriram, M., & Wang, A. H. -J. (1993), "Crystallographic Studies of Metal Ion-DNA Interactions: Different Binding Modes of Cobalt(II), Copper(II) and Barium(II) to N7 of Guanines in Z-DNA and a Drug-DNA Complex." *Nucleic Acids Res.* **21**, 4093-4101.
- Garel, J. -P., Jordan, J. -C. & Mandel, P. (1972) "Coefficients de Partage d'Oligoribonucleotides Dans Les Systemes Solvants Salins." *J. Chromatogr.*, **67**, 277-290.
- Geige, R. & Ducruix, A. (1992) "An Introduction to the Crystallogenesis of Biological Macromolecules." in *Crystallization of Nucleic Acids and Proteins: A Practical Approach* (Ducruix, A. & Geige, R. eds.) pp 1-18, Oxford University Press, New York.
- Geige, R., Lorber, B., & Theobald-Dietrich, A (1994) "Crystallogenesis of Biological Macromolecules: Facts and Perspectives." *Acta Cryst.* **D50**, 339-350.
- Geierstanger, B. H., Kagawa, T. F., Chen, S. -L., & Ho, P. S. (1991) "Base-Specific Binding of Copper(II) to Z-DNA: The 1.3 Å Single Crystal Structure of d(m<sup>5</sup>CGUAm<sup>5</sup>CG) in the Presence of CuCl<sub>2</sub>." *J. Biol. Chem.* **266**, 20185-20191.
- 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-Å Atomic Resolution." *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 Cation." *Biochemistry* **24**, 237-240.
- Hansch, C., & Leo, A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York.
- 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.



- Hendrickson, W. A. & Konnert, J. (1979) *Biomolecular Structure, Conformation, Function and Evolution* Vol 1. (Srinivasan, R. ed.) pp43-57 Pergamon, Oxford.
- 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., Frederick, C. A., Saal, D., Wang, A. H. -J., & Rich, A (1987) "The Interactions of Ruthenium Hexaammine with Z-DNA: Crystal Structure of a  $\text{Ru}(\text{NH}_3)_6^{3+}$  Salt of d(CGCGCG) at 1.2 Å Resolution." *J. Biomol. Struct. Dynam.* 4, 521-534.
- 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., Quigley, G.J., Tilton, R., & Rich, A. (1988) "Hydration of Methylated and Nonmethylated B-DNA and Z-DNA." *J. Phys. Chem.* 92, 939-945.
- Hunter, C. A. "Sequence-dependent DNA Structure: The Role of Base Stacking Interactions." *J. Mol. Biol.* 230, 1025-1054.
- Irikura, K. K., Tidor, B., Brooks, B. R., & Karplus, M. (1985) "Transition from B to Z DNA: Contribution of Internal Fluctuations to the Configurational Entropy Difference." *Science* 229, 571-572.
- Johnson, L. (1994) "Structure-based Drug Design." *International Union of Crystallographers Newsletter* 2, 5-8.
- Jovin, T. M., McIntosh, L.P., Arndt-Jovin, D. J., Zarling, D. A., Robert-Nicoud, M., van de Sande, J. H., Jorgensen K. F., & Eckstein, F. (1983) "Left-handed DNA: From Synthetic Polymers to Chromosomes." *J. Biol. Struct. Dynam.* 1, 21-57.
- Jovin, T. M., Soumpasis, D.M., & McIntosh, L.P. (1987) "The Transition Between B-DNA and Z-DNA." *Ann. Rev. Phys. Chem.* 38, 521-560.
- Kagawa, T. F., Geierstanger, B. H., Wang, A. H. -J., & Ho, P. S. (1991) "Covalent Modification of Guanine Bases in Double-stranded DNA: The 1.2 Å Z-DNA Structure of d(CGCGCG) in the Presence of  $\text{CuCl}_2$ ." *J. Biol. Chem.* 266, 20175-20184.
- Kagawa, T. F., Howell, M. L., Tseng, K., & Ho, P. S. (1993) "Effects of Base Substituents on the Hydration of B- and Z-DNA: Correlations to the B- to Z-DNA Transition." *Nucleic Acids Res.* 21, 5978-5986.
- 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 transition as a Model." *Biochemistry* 28, 6642-6651.
- Kauzmann, W. (1959) "Some Factors in the Interpretation of Protein Denaturation." *Adv. Prot. Chem.* 14, 1-63.

- Klotz, I. M. (1960) "Non-Covalent Bonds in Protein Structure" *Prot. Structure and Function* 13, 25-48.
- Kollman, P., Weiner, P., Quigley, G. J., & Wang, A. H. -J. (1982) "Molecular Mechanism Studies of Z-DNA: A Comparison of the Structural and Energetic Properties of Z- and B-DNA." *Biopolymers* 21, 1945-1969.
- Krueger, W. C., & Prairie, M. D. (1985) "A Low-salt Form of Poly (dG-m<sup>5</sup>dC)•Poly(dG-m<sup>5</sup>dC)." *Biopolymers* 24, 905-910.
- Langmuir, I. (1925) "The Distribution and Orientation of Molecules." *Colloid. Symp. Monogr.* 3, 48-75.
- Latha, P. K., & Brahmachari, S. K. (1985) "B to Z Transition in DNA & Their Biological Implications." *FEBS Lett.* 182, 315-318.
- Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M. A., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., Thoma, C. D., Shaw, W. V., & Harrison, P. M. (1991) "Solving the Structure of Human H Ferritin by Genetically Engineering Intermolecular Crystal Contacts." *Nature* 349, 541-544.
- Makhatadze, G. I. & Privalov, P. L. (1993) "Contribution of Hydration to Protein Folding Thermodynamics." *J. Mol. Biol.* 232, 639-659.
- Matthews, B. W. (1987) "Genetic and Structural Analysis of the Protein Stability Problem." *Biochemistry* 26, 6885-3758.
- McElroy, H. E., Sisson, G. W., Schoettlin, W. E., Aust, R. M., & Villafranca, J. E. (1992) "Studies on Engineering Crystallizability by Mutation of Surface Residues of Human Thymidylate Synthase." *J. Cryst. Growth* 122, 265-272.
- Melander, W., & Horvath, C. (1977) "Salt Effects on Hydrophobic Interactions in Precipitation and Chromatography of Proteins: An Interpretation of the Lyotropic Series." *Arch. Biochem. Biophys.* 183, 200-215.
- Mirkin, S. M., Lyamichev, V. I., Kumarev, V. P., Kobzev, V. F., Nosikov, V. V. & Vologodskii, A. V. (1987) "The Energetics of the B-Z Transition in DNA." *J. Biomol. Struct. Dynam.* 5, 79-88.
- Mittl, P. R. E., Berry, N. S., Scrutton S. & Perham, R. N. (1994) "A Designed Mutant of the Enzyme Glutathione Reductase Shortens the Crystallization Time By a Factor of Forty." *Acta. Cryst. D* 50, 228-231.
- Moras, D., & Bergdoll, M. (1988) "Packing and Molecular Interactions in tRNA Crystals." *J. Cryst. Growth* 90, 283-294.
- Muller, N. (1990) "Search for a Realistic View of Hydrophobic Effects." *Acc. Chem. Res.* 23, 23-28.
- Murphy, K. P., Pivalov, P.L., Gill, S.J., (1990) "Common Features of Protein Unfolding and Dissolution of Hydrophobic Compounds." *Science* 247, 559-561.

- Narasimhan, V., & Bryan, A. M. (1984) "Conformational Flexibility of poly (dG-m<sup>5</sup>C) Under Very Low Salt Conditions." *Experientia* **40**, 827-828.
- Nillson and Karplus, (1986) *J.Comp Chem* **4**, 187-217
- Nordheim, A., Lafer, E. M., Peck, L. J., Wang, J. C., Stollar, B. D., & Rich, A. (1982) "Negatively Supercoiled Plasmids Contain Left-handed Z-DNA Segments as Detected by Specific Antibody Binding." *Cell* **31**, 309-318.
- Oxender, D.L. & Fox, C.F. eds. (1987) *Protein Engineering* Alan Liss, New York.
- Peck, L. J., & Wang J.C. (1983) "Energetics of B- to Z- Transition in DNA." *Proc. Natl. Acad. Sci., USA* **80**, 6206-6210.
- Perutz, M. F., Kendrew, J. C., & Watson, H. C. (1965) "Structure and Function of Haemoglobin II: Some Relations Between Polypeptide Chain Configuration and Amino Acid Sequence." *J. Mol. Biol.* **13**, 669-678.
- Pohl, F. M. (1979) "" *Nature* **260**, 365-366.
- Pohl, F. M., & Jovin, T. M. (1972) "Salt-induced Co-operative Conformational Change of a Synthetic DNA: Equilibrium and Kinetic Studies with poly(dG-dC)." *J. Mol. Biol.* **57**, 375-395.
- Pratt, L. R. & Chandler, D. (1977) "Theory of the Hydrophobic Effect." *J. Chem. Phys.* **67**, 3683-3704.
- Privalov, P.L., Makhatdze, G.I. (1992) "Contribution of Hydration and Non-Covalent Interactions to the Heat Capacity Effect on Protein Unfolding." *J. Mol. Biol.* **224**, 715-723.
- Privalov, P. L., & Makhatadze, G. I. (1993) "Contribution of Hydration to Protein Folding Thermodynamics II. The Entropy and Gibbs Energy of Hydration." *J. Mol. Biol.* **232**, 660-679.
- Quigley, G. J. (1991) "Understanding the Crystal Packing of Z-DNA." *J. Cryst. Growth* **110**, 131-136.
- Rich, A., Nordheim, A., & Wang, A. H. -J. (1984) "The Chemistry and Biology of Left-handed Z-DNA." *Ann. Rev. Biochem.* **53**, 791-846.
- Rose, G. D., & Wolfenden, R. (1993) "Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding." *Annu. Rev. Biophys. Biomol. Struct.* **22**, 381-415.
- Saenger, W. (1984) "Principles of Nucleic Acid Structure." (Cantor, C. R. Ed) Springer-Verlag, New York.
- Saenger, W., Hunter, W. N., & Kennard O. (1986) "DNA Conformation is Determined by Economics in the Hydration of Phosphate Groups." *Nature* **324**, 385-388.

- Salemme, F. R., Genieser, L., Finzel, B. C., Hilmer, R. M., & Wendoloski, J. J. (1988) "Molecular Factors Stabilizing Protein Crystals." *J. Cryst. Growth* **90**, 273-282.
- Schellman, J. A. (1987) "The Thermodynamic Stability of Proteins." *Ann. Rev. Biophys. Biophys. Chem.* **16**, 115-137.
- Schneider, B., Cohen, D. M., Schleifer, A. R., Srinivasan, A. R., Olsen W. K., & Berman, H.M. (1993) "A Systematic Method for Studying the Spatial Distribution of Water Molecules Around Nucleic Acid Bases." *Biophys.J.* **65**, 2291-2303.
- Schneider, B., Ginell, S.L., Jones, R., Gaffney, B. & Berman, H.M. (1992) "Crystal and Molecular Structure of a DNA Fragment Containing a 2-aminoadenine Modification: The Relationship Between Conformation, Packing, and Hydration in Z-DNA Hexamers." *Biochemistry* **31**, 9622-9628.
- Schroth, G. P., Kagawa, T. F. & Ho, P. S. (1993) "Structure and Thermodynamics of Nonalternating C•G Base Pairs in Z-DNA: The 1.3Å Crystal Structure of the Asymmetric Hexanucleotide d(m<sup>5</sup>CGGGm<sup>5</sup>CG)•d(m<sup>5</sup>CGCCm<sup>5</sup>CG)." *Biochemistry* **32**, 13381-13392.
- Siemens Analytical X-ray Instruments Inc. (1990) *SHELXTL PC Release 4.1* Madison Wisconsin U. S. A.
- Siemens Analytical X-ray Instruments Inc.(1993) *XSCANS Software Users Guide Version 2.0* Madison, Wisconsin U.S.A.
- Sinden, R. R. & Wells, R.D. (1992)*Curr. Opin. Biotechnol.* **3**, 612-622.
- Singh, S. B., Pearlman, D. A., & Kollman, P. A (1993) "Free Energy Component Analysis: Application to the 'Z-Phobicity' of A•T Base Pairs." *J. Biomol. Struct. Dynam.* **11**, 303-311.
- Singleton, C. K. (1987) "Z-DNA in Solution" in Topics in Nucleic Acid Structure (Neidle, S. Ed) London MacMillan.pp 163-191
- Stout, G. H. & Jensen, L. H. (1989) "X-ray Structure Determination: A Practical Guide." 2nd Ed. John Wiley and Sons, New York.
- Soumpasis, D. M. (1984) "Statistical Mechanics of the B to Z Transition of DNA: Contribution of Diffuse Ionic Interactions." *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5116-5120.
- Soumpasis, D. M., Robert-Nicoud, M. & Jovin, T.M. (1986) "B-Z DNA Conformational Transition in 1:1 Electrolytes: Dependence Upon Counterion Size." *FEBS Lett.* **213**, 341-344.
- Tanford, C (1980) "The Hydrophobic Effect: Formation of Micelles and Biological Membranes." John Wiley and Sons, New York.
- Tidor, B., Irikura, K. K., Brooks, B. R., & Karplus, M. (1983) "Dynamics of DNA Oligomers." *J. Biomol. Struct. Dynam.* **1**, 231-252.

- Vologodskii, A. V. & Frank-Kamenetskii, M. D. (1984) "Left-handed Z-form in Superhelical DNA: A Theoretical Study." *J. Biomol. Struct. Dynam.* 1, 1325-1333.
- Vorlickova, M., & Jaroslav, K. (1984) *Int. J. of Biol. Macromol.* 6, 77-80.
- Wang, A. H. -J., Gessner, R. V., van der Marel, G., van Boom, J. H., & Rich, A. (1985) "The Crystal Structure of Z-DNA Without an Alternating Purine-Pyrimidine Sequence." *Proc. Natl. Acad. Sci. U.S.A.* 82, 3611-3615.
- Wang, A. H. -J., Hakoshima, T., van der Marel, G., van Boom, J. H., & Rich, A. (1984) "AT Base Pairs Are Less Stable Than GC Base Pairs in Z-DNA: The Crystal Structure of d(m<sup>5</sup>CGTAm<sup>5</sup>CG)." *Cell* 37, 321-331.
- Wang, A. H. -J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979) "Left-handed Double Helical DNA: Variations in the Backbone Conformation." *Science* 211, 171-176.
- Wang, A. H. -J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H., & Rich, A. (1981) "Molecular Structure of a Left-handed Double Helical DNA Fragment at Atomic Resolution." *Nature* 282, 680-686.
- Wang, A. H. -J., Teng, M. -K. (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.
- Whittle P.J. & Blundell, T.L. (1994) "Protein Structure-Based Drug Design." *Annu. Rev. Biophys. Biophys. Chem.* 23, 349-375
- Woolfson, D. N., Cooper, A., Harding, M. M., Williams, D. H., & Evans, P. A. (1993) "Protein Folding in the Absence of the Solvent Ordering Contribution to the Hydrophobic Interaction." *J. Mol. Biol.* 229, 502-511.
- Zamenhof, S., Brawerman, B. & Chargaff, E. (1952) "On the Deoxypentose Nucleic Acids from Several Microorganisms." *Biochim. Biophys. Acta.* 9, 402.
- Zhou, G. & Ho, P. S. (1990) "Stabilization of Z-DNA by Demethylation of Thymine Bases: 1.3-Å Single-Crystal Structure of d(m<sup>5</sup>CGUAm<sup>5</sup>CG)." *Biochemistry* 29, 7229-7236.