AN ABSTRACT OF THE DISSERTATION OF

Jeffrey M. Vargason for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on February 26, 2002. Title: The Effect of Cytosine Methylation on DNA Structure.

Abstract approved: Redacted for Privacy

Pui Shing Ho

DNA methylation is common in prokaryotes and eukaryotes and has been implicated in various biological roles including gene silencing, X-chromosome inactivation, and genomic imprinting. 5-methylcytosine the "fifth base" of the genetic code comprises 1-3% of the human genome and is primarily found on cytosines within the context of the CpG sequence. Although progress has been made in understanding the biological roles of 5-methylcytosine, we are only beginning to uncover how it changes the local structure and global conformation of DNA. This thesis deals with the local perturbations in structure and hydration and the global conformational changes induced by the presence of 5-methylcytosine in DNA as determined by single crystal x-ray diffraction.

5-methylcytosine induces a novel conformation in the structure of duplex DNA. This conformation has characteristics of both the A-DNA and B-DNA conformations as well as some unique defining characteristics. This distinct duplex

provides a structural rationale for the increased rate of deamination in 5-methylcytosine relative to cytosine. In addition to this novel conformation, 5-methylcytosine stabilizes intermediates within the B-DNA to A-DNA transition pathway, thus providing a crystallographic map of the transition from B-DNA to A-DNA.

5-methylcytosine was also used as a tool to probe the stabilizing features of the DNA four-way junction (known as the Holliday junction). The first crystal structures of Holliday junctions were found serendipitously while studying duplex DNA. The DNA four-way junction formation in these crystals was thought to be stabilized by a network of sequence dependent hydrogen bonds at the junction crossover. In this thesis, 5-methylcytosine was used to perturb these hydrogen bonds; however, the junction persisted, suggesting that there is flexibility in the types of sequences that can accommodate junction formation in the crystal, as well as, flexibility in the global structure of the junction. Overall, this work describes the effects of 5-methylcytosine on the local and global structure and hydration of DNA structure, as well as raising some interesting questions regarding the biological impact of methylation induced DNA structure.

The Effect of Cytosine Methylation on DNA Structure

by

Jeffrey M. Vargason

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented Februray 26, 2002
Commencement June 2002

Doctor of Philosophy dissertation of Jeffrey M. Vargason presented on February 26, 2002

APPROVED:

Redacted for Privacy

Major Professor, Representing Biochemistry and Biophysics

Redacted for Privacy

Head of the Department of Biochemistry and Biophysics

Redacted for Privacy

Dean of Graduate School

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

Redacted for Privacy

Jeffrey M. Vargason, Author

ACKNOWLEDGMENTS

I thank my professor and mentor, Dr. P. Shing Ho, whose knowledge, creativity, and guidance provided me with tremendous opportunities to develop as a scientist. Shing, you've created an impressive laboratory environment that encourages collaboration and the pursuit of scientific discovery. To all the members of the Ho Lab that have come before or during my time thanks for the legacy you've left and to those who remain and will come after, continue the excellence! Specifically, I want to thank Blaine Mooers for all the time he spent teaching me in the practical aspects of crystallography and nucleic acid structure. Thanks for spending your time helping me even though I was new to the lab. Beth Basham thanks for sharing your knowledge of programming (even though I have chosen C over your beloved Fortran). Brandt Eichman, thank you for the camaraderie we shared in the lab and the time you spent with me discussing various crystallographic topics. Thanks for continuing to "raise the bar"; your drive is contagious. Thank you Shawn Watts for sharing your knowledge of structure based drug design and your love for organic chemistry. Frank Hays thanks for joining the lab and continuing some of the projects that I left behind. You'll do great.

To the Karplus lab, I want to say a special thank you for your willingness to share beam time at both the ALS and APS. Dr. Andrew Karplus thanks for your expertise in seminar preparation and presentation, as well as your meticulous

approach to science. Savvas Savvides thanks for sharing your methodical approach to crystallography. Zac "DNA, it's just a double-helix, right?" Wood thanks for your larger than life attitude and for passing a little of it on to me. Thanks for your help in each of our synchrotron trips. You will be a great scientist. Thank you Rick Faber for maintaining the x-ray diffractometer and associated instrumentation, as well as, for your endless supply of computer equipment and expertise. Thank you, Ganapathy Sarma for your amazing abilities with the Rubik's cube.

Special thanks to Keith Henderson for taking my crystals and collecting them at the ALS in your spare time. I am greatly indebted to you for the awesome MAD data. Thank you Dennis Lohr for the Gal4 collaboration. I very much enjoyed the opportunity to contribute to your research project.

I'd like to thank my undergraduate professor Dr. Joseph Bohanon at Evangel University for his love of chemistry. Thank you for the education and encouragement to take my academic experience to the next level.

Thank you to all of the teachers in my life (many of whom I also call family) for your patience and dedication to the academic development of others. You make accomplishments like this possible!

Thank you to my parents (Richard and Sherilyn Vargason) for all of your love and support. Together, you provided a wonderful home and a Godly environment where I was encouraged to challenge myself, but at the same time I was reminded to maintain my priorities of my relationship with Christ and my family. Thanks for instilling in me the goal of happiness in whatever life brings.

To my extended family, thank you for always loving and supporting me. Thank you to my Aunt Lori for continuing to be my mathematics mentor.

Saving the best for last, I want to thank my wife and best friend Annette. You have endured every day of the PhD process with me and for that I am forever indebted. Thanks for your commitment to see me finish. You encouraged me when I was down. You motivated me to succeed. You celebrated with me in my accomplishments. You are my muse, and I will always love you.

CONTRIBUTION OF AUTHORS

P. Shing Ho was involved in the design, analysis, and writing of each manuscript. Brandt F. Eichman played a consultative role in the practical aspects of crystallography during the research done in Chapter 2. Keith Henderson collected the multi-wavelength data on crystals of $d(GGBr^5CGCC)_2$ in Chapter 3.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF TABLES

DEDICATION

I dedicate this thesis to my wife Annette. I am grateful for her love, friendship, advice, and steadfast support throughout the PhD process.

The Effect of Cytosine Methylation on DNA Structure

Chapter 1

Introduction

5-methylcytosines have been found to be common in prokaryotic and eukaryotic genomes, representing approximately 1-3% of the bases in human DNA (Spruck *et al.* 1993), and have been implicated in the processes of X-chromosome inactivation (Singer-Sam and Riggs 1993), genomic imprinting (Sasaki *et al.* 1993), and gene inactivation (Graessmann and Graessmann 1993). The four standard bases of DNA are adenine, thymine, guanine, and cytosine. However, various non-standard or modified bases have been found in genomes (e.g. 6-methyladenine, 8-oxoguanine, 5-methylcytosine, etc.), some as a result of DNA damage, and some of these bases have evolved a biological role. 5-methylcytosine was first found to occur in DNA in 1948 (Hotchkiss 1948). Since that time, numerous biological roles have been attributed to DNA methylation. Thus, 5-methylcytosine has been regarded as "the fifth base" of the genetic code.

This modified base, however, has its dark-side–it serves as mutational hotspots in the genome. Both cytosine and 5-methylcytosine are subject to spontaneous deamination to form uracil and thymine, respectively (Wiebauer *et al.*

1993). However, the rate of transformation is 21-fold higher in the methylated compared to the nonmethylated base (Zhang and Mathews 1994). Transition mutations from C·G to T·A base pairs occur through replication if the deaminated products are left unrepaired. The major sites of biological methylation and transition mutations in the human genome occur at CpG steps and have been linked to tumorigenesis in a number of human tissues (Spruck *et al.* 1993).

Transition mutations at specialized sites following deamination of cytosine or 5-methylcytosine have been suggested to be epigenetic; however, the present experimental evidence only suggests an epigenetic role in the conversion of cytosine to 5-methylcytosine (Holliday 1993). Methylation of DNA is a mechanism to change the sequence of the DNA without changing the coding of that sequence. This change is heritable with the help of maintenance methylation, which involves catalysis by methyltransferases that recognize hemimethylated CpG DNA. Since the addition of the methyl group is not involved in coding, the question arises, is the methyl group directly recognized by proteins or does the methyl group change the local structure of the DNA? Hodges-Garcia *et al.* suggests that there is little evidence that the methyl group alone modulates protein-DNA interactions (Hodges-Garcia and Hagerman 1992). Studies on the recognition event of the EcoRI endonuclease for its restriction site suggest that binding is partially regulated by an altered DNA conformation resulting from N6-methyladenine. Thus, the effect of 5-methylcytosine on DNA structure has been of great interest.

In 1953, the first structure of DNA was elucidated by Watson and Crick and given the designation B-DNA (Watson and Crick 1953). This particular duplex is characterized by a 3.4 Å rise of the stacked basepairs and 10 basepairs per helical turn. This DNA duplex conformation is considered the "normal" or "standard" form. In conjunction with Watson and Crick, Franklin and Gosling were working on the first alternative form of DNA, A-DNA (Franklin and Gosling 1953). This conformation arose from the dehydration of B-DNA fibers; and this change was found to be reversible upon rehydration of the DNA duplexes. The A-DNA conformation is characterized by a 2.7 Å rise and 11 basepairs per helical turn. The reversible conversion of B-DNA to A-DNA as a result of hydration prompted research involving the prediction of a sequence's propensity to form A-DNA?

The idea that hydration and dehydration initiated the reversible conversion of B-DNA and A-DNA was used to assign a propensity to form A-DNA to any primary sequence of DNA. Further research on DNA hydration showed a correlation in the position of waters around a DNA duplex and its conformation (Schneider *et al.* 1992). A-DNA is dehydrated with respect to B-DNA and thus its stability should be dependent on the hydration of the DNA surface (Alden and Kim 1979). Based on these observations a set of thermodynamic rules (termed A-DNA Propensity Energies, APEs) were derived to predict the propensity of a sequence to form A-DNA (Basham *et al.* 1995). Given a sequence of interest, the propensities are averaged across a DNA triplet to give an overall APE for the sequence. Based on this propensity, a prediction of the conformation can be made of a particular

sequence in solution. One of the initial tests of these thermodynamic rules was the design, crystallization, and structure determination of short DNA sequences that were predicted to form a particular conformation.

The first set of sequences crystallized to test APEs was $d(GCCGGC)_2$ and $d(GGCGCC)_2$. In this sequence construct, the first three basepairs of one sequence are the last three basepairs of the other sequence. Thus, both sequences have four triplets, two of which they share in common (GGC and GCC). Interestingly, the APE for $d(GCCGGC)_2$ predicts an A-DNA conformation, while the APE for $d(GGCGCC)_2$ predicts a B-DNA conformation. The sequence $d(GCCGGC)_2$ was solved to 2.3 Å as A-DNA (Mooers *et al.* 1995). Described in Chapter 2 of this thesis is the structure of the sequence $d(GGCGCC)_2$ solved to 2.7 Å as B-DNA. Thus, the APEs predicted the structures of these two sequences correctly. To further test the APEs the question was asked, could we perturb this B-DNA conformation to A-DNA by adding something to dehydrate the structure of the DNA. Cobalt hexamine salt was a good candidate as it was shown to promote the formation of A-DNA in solution (Xu *et al.* 1993). Crystals containing this salt were grown and the structure was seen to remain in the B-DNA conformation despite the presence of ordered cobalt cations in the structure. At this point, a cytosine was changed to 5-methylcytosine to produce new sequences. The added hydrophobic DNA surface exposure as a result of adding the methyl group was proposed to change the conformation of this sequence based on the premise that the A-DNA conformation is favored in dehydrating conditions.

5-methylcytosine has not only been shown to influence the bending or flexibility of DNA, but also the conformation of DNA. Recent studies on the Dickerson-Drew dodecamer d(CGCGAATTCGCG)$_2$ (Drew *et al.* 1981) have shown that methylation of this sequence reduces the flexibility of the sequence and underwinds the duplex at the position of the methylated cytosines (Nathan and Crothers 2002). Additional studies have shown that the A-DNA conformation (Frederick *et al.* 1987; Mooers *et al.* 1995) as well as the left-handed Z-DNA conformation (Behe and Felsenfeld 1981; Fujii *et al.* 1982) have been stabilized by the presence of the methyl group on cytosine. However, these conformations do not require 5-methylcytosine to form their respective conformations. Futhermore, there have been no crystal structures of a sequence in the B-DNA conformation that have changed to A-DNA or Z-DNA upon methylation of one or more cytosines.

Described here are the first structures in which 5-methylcytosine has induced not only local perturbations, but also novel conformational changes in both duplex and Holliday junction DNA structure. In Chapter 2, the sequence of d(GGCGm^{5}CC)$_2$ is shown to form a novel conformation with characteristics of both B-DNA and A-DNA. This particular sequence was also crystallized as standard A-DNA when allowed to crystallize over an extended period of time (2-3 months) suggesting that the novel conformation was trapped by crystallization. Presumably, over an extended period of time the crystallization solution is further concentrated by evaporation giving a more dehydrated environment favoring A-DNA formation. Chapter 3 describes the same parent sequence with a different

methylation position to give the sequence d(GGm^{5}CGCC)$_2$. This particular crystal is composed of four duplexes of DNA with three unique conformations of DNA, one standard A-DNA duplex, two A/B type duplexes with primarily A-DNA characteristics, and finally a composite duplex composed of both A-DNA and B-DNA nucleotides. The first three 5' nucleotides on each strand are A-DNA followed by three B-DNA nucleotides. The combined structures from the first two chapters allow us to map out a pathway for the B-DNA to A-DNA transition. In Chapter 4, 5-methylcytosine is used as a tool to probe the stability of the Holliday junction. The Holliday junction was recently crystallized in the sequences of d(CCGGGACCGG) (Ortiz-Lombardía *et al.* 1999) and d(CCGGTACCGG) (Eichman *et al.* 2000). These structures share common sequence dependent hydrogen bonds that were thought to be the reason for the sequence's propensity to form a Holliday junction. These hydrogen bonds were disrupted by the presence of the 5-methylcytosine; however, the junction persisted. General conclusions resulting from this work on cytosine methylation including the potential biological function of these structures are presented in Chapter 5.

Chapter 2

The Extended and Eccentric E-DNA Structure Induced by Cytosine Methylation or Bromination

Jeffrey M. Vargason, Brandt F. Eichman, and P. Shing Ho

Published in *Nature Structural Biology*,

Nature Publishing Group, New York, NY

2000, **7**, 758-761

2.1 Summary

Cytosine methylation or bromination of the DNA sequence $d(GGCGCC)_2$ is shown here to induce a novel extended and eccentric double-helix, which we call E-DNA. Like B-DNA, E-DNA has a long helical rise and bases perpendicular to the helix axis. However, the 3'-*endo* sugar conformation gives the characteristic deep major groove and shallow minor groove of A-DNA. Also, if allowed to crystallize for a period of time longer than that yielding E-DNA, the methylated sequence forms standard A-DNA, suggesting that E-DNA is a kinetically trapped intermediate in the transition to A-DNA. Thus, the structures presented here chart a crystallographic pathway from B-DNA to A-DNA through the E-DNA intermediate in a single sequence. The E-DNA surface is highly accessible to solvent, with waters in the major groove sitting on exposed faces of the stacked nucleotides. We suggest that the geometry of the waters and the stacked base pairs would promote the spontaneous deamination of 5-methylcytosine in the transition mutation of $dm^5C{\cdot}dG$ to $dT{\cdot}dA$ base pairs.

2.2 Introduction

5-Methyldeoxycytidine (dm^5C) is often considered the fifth nucleotide of the genetic code. Prokaryotes use cytosine methylation to distinguish parent from

replicated daughter DNAs, and host from viral DNAs (Noyer-Weidner and Trautner 1993). In eukaryotes, dm^5C has been implicated in processes as varied as X-chromosome inactivation, genomic imprinting, and gene inactivation (Antequera and Bird 1993; Sasaki *et al.* 1993; Singer-Sam and Riggs 1993). The effect of cytosine methylation on double-helical DNA, therefore, has been of great interest over the years. Cytosine methylation and bromination have been shown to stabilize A-DNA (Frederick *et al.* 1987; Mooers *et al.* 1995) and left-handed Z-DNA (Behe and Felsenfeld 1981; Fujii *et al.* 1982). However, neither A-DNA nor Z-DNA requires cytosine methylation for its formation; therefore, methylation serves to facilitate rather than induce these conformations. The question is whether methylation can induce a unique structure.

Here, we show that the sequence $d(GGCGCC)_2$ crystallizes as standard B-DNA, while the sequences $d(GGCGm^5CC)_2$ and $d(GGCGBr^5CC)_2$ form a new conformation, which we call E-DNA, that has structural characteristics of both B-DNA and A-DNA (Figure 2.1a). The B-DNA and E-DNA sequences were crystallized under nearly identical conditions indicating that, unlike A-DNA (Mooers *et al.* 1995), E-DNA does not require dehydration. Interestingly, the sequence $d(GGCGm^5CC)_2$ crystallizes as E-DNA over two to three weeks, but when allowed to crystallize for two to three months forms a standard A-DNA double helix. Finally, we suggest that E-DNA may play a role in the transition mutation of $dm^5C{\cdot}dG$ to $dT{\cdot}dA$ base pairs in the cell.

Figure 2.1. Comparing E-DNA with B-DNA and A-DNA. *a.* View into (top) and down (bottom) the helix axes of 12 base pair models constructed from the crystal structures of $d(GGCGCC)_2$ as B-DNA, and $d(GGCGm^5CC)_2$ as E-DNA and as A-DNA. The phosphodeoxyribose backbones are traced by yellow ribbons. *b.* Helical parameters (calculated by CURVES 5.2 (Lavery and Sklenar 1989)) of hexanucleotide structures containing dC·dG, dm^5C·dG, and dBr^5C·dG base pairs as B-DNA (green circles), A-DNA (red triangles), and E-DNA (blue squares). Parameters determined from the current structures are shown as filled symbols, while those from the previously published A-DNA structures (Mooers *et al.* 1995) are open. Concentric ovals represent 1 and 2 standard deviations from the mean. Parameters for A-DNA and B-DNA fibers are indicated by the boxed A and B, respectively.

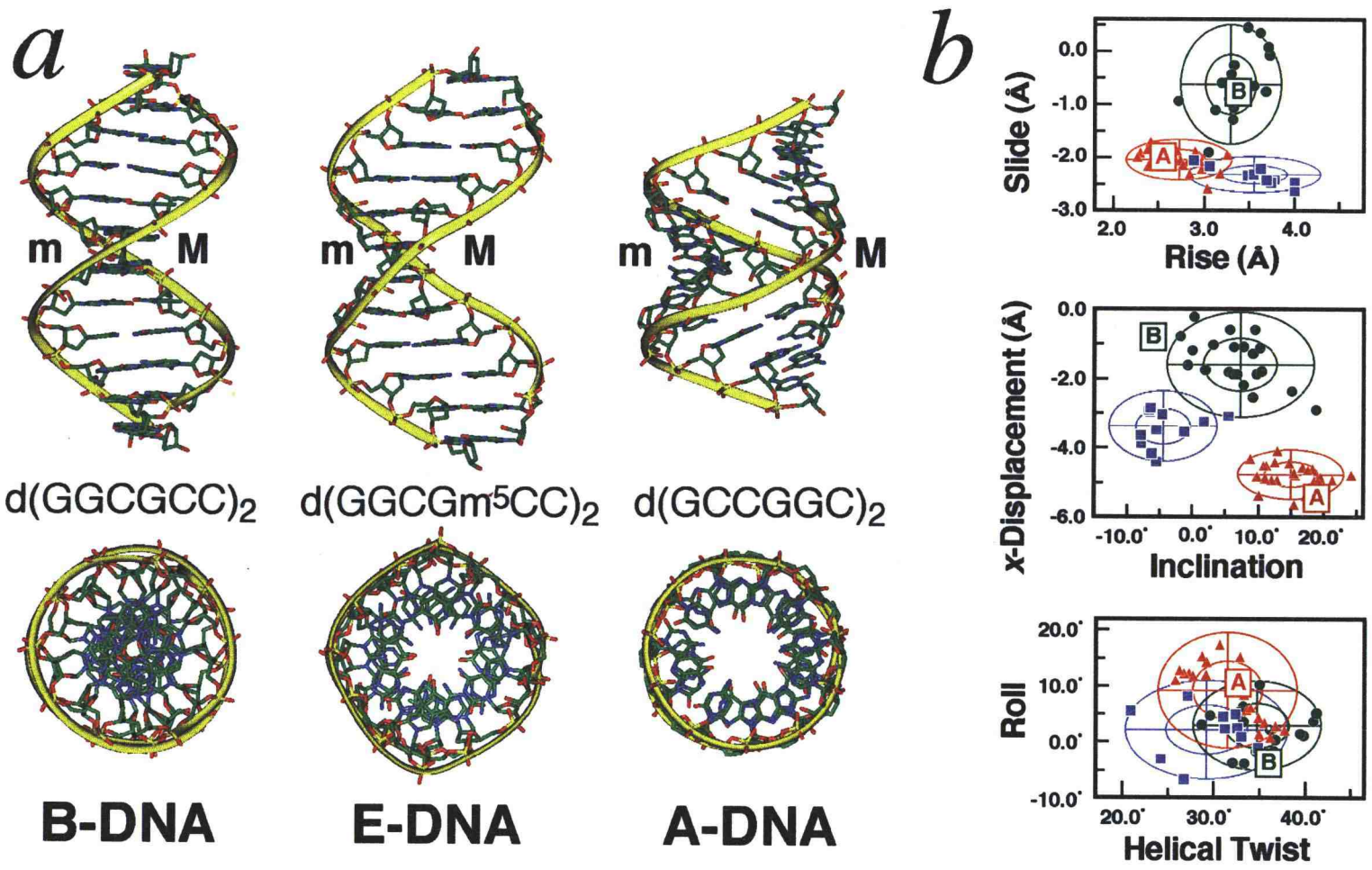
a
m
M
m
M
m
M
d(GGCGCC)2
d(GGCGm5CC)2
d(GCCGGC)2
B-DNA
E-DNA
A-DNA
b
Slide (Å)
Rise (Å)
x-Displacement (Å)
Inclination
Roll
Helical Twist
A
B

Figure 2.1

2.3 Materials and Methods

The sequence $d(GGCGCC)_2$ was crystallized from solutions containing 0.7 mM DNA, 25 mM sodium cacodylate buffer (pH 6), 0.8 mM $MgCl_2$, and 0.75 mM spermine tetrahydrochloride equilibrated against a reservoir of 15% (v/v) 2-methyl-2,4-pentanediol tetrahydrochloride equilibrated against a reservoir of 15% (v/v) 2-methyl-2,4-pentanediol (MPD) and from this same solution with 1-2 mM $Co(NH_3)_6^{+3}$ added. The cobalt form (Table 2.1) was solved using two separate B-DNA d(GGC/GCC) duplexes in a directed real space translation/rotation/rigid body search (XPLOR 3.851 (Brünger 1992) script written in this lab). The refined structure was then used to solve the structures of three stacked B-DNA duplexes in the spermine crystal-form. The solvent content of the spermine form is ~70% greater than typical B-DNA crystals (Dickerson 1992), and could not be accurately modeled in the structure. This explains the moderate resolution and relatively high R-values of this structure.

The sequences $d(GGCGm^5CC)_2$ and $d(GGCGBr^5CC)_2$ were crystallized from solutions containing 0.7 mM DNA, 25 mM sodium cacodylate (pH 6), 0.8 mM $MgCl_2$, and 0.1-0.5 mM spermine tetrahydrochloride equilibrated against a reservoir of 15-20% (v/v) MPD. The structure of the brominated sequence was solved first by molecular replacement with the program AMoRe (Navaza 1994), using the central four base pairs of an ideal A-DNA structure as the search model.

Table 2.1. Data collection and refinement statistics

	GGCGCC + $Co(NH_3)_6^{+3}$	GGCGCC + Spermine^{+4}	GGCGm^{5}CC E-DNA	GGCGBr^{5}CC E-DNA	GGCGm^{5}CC A-DNA
Data Collection[a]					
Space group	$P4_122$	$P4_12_12$	$P4_32_12$	$P4_32_12$	$C222_1$
Unit cell lengths (Å)	$a = b = 42.6$, $c = 63.3$	$a = b = 71.5$, $c = 59.6$	$a = b = 62.1$, $c = 24.3$	$a = b = 60.4$, $c = 24.7$	$a = 37.1$, $b = 46.8$, $c = 110.7$
Resolution (Å)	42.59-2.6	71.5-2.7	20.0-2.2	60.4-2.25	20.0-2.0
Total reflections (unique)	20150 (2026)	32859 (4551)	23113 (2571)	31302 (2393)	27952 (6053)
Completeness (%)[b]	99.9 (99.9)	98.8 (96.6)	95.7 (91.6)	99.9 (99.9)	88.6 (62.9)
R_{merge} (%)[b,c]	6.1 (25.1)	5.7 (39.5)	4.9 (51.3)	5.0 (34.7)	5.5 (38.4)
Refinement					
Resolution	8.0-2.6	8.0-2.7	8.0-2.2	8.0-2.25	8.0-2.0
R_{cryst} (R_{free}) (%)[d]	20.5 (27.9)	22.6 (28.8)	21.0 (27.3)	19.3 (25.6)	20.5 (24.8)
DNA (Solvent) atoms	240 (22)	600 (37)	242 (29)	242 (28)	693 (126)
RMSD Bond lengths (Å)	0.005	0.006	0.007	0.004	0.003
RMSD Bond angles (°)	1.095	0.981	0.903	0.674	0.811

[a] X-ray diffraction data were collected at room temperature using CuK_α radiation from a Rigaku RUH3R generator and an R-AXIS IV image plate detector, and reduced using the programs Denzo and Scalepack from the HKL package (Otwinowski and Minor 1997) and D*Trek (Pflugrath 1999).

[b] Values in parentheses refer to the highest resolution shell

[c] $R_{merge} = \Sigma_{hkl} \Sigma_i | I_{hkl,i} - \langle I \rangle_{hkl} | / \Sigma_{hkl} \Sigma_i | I_{hkl,i} |$ where I_{hkl} is the intensity of a reflection and $\langle I \rangle_{hkl}$ is the average of all observations of this reflection and its symmetry equivalents.

[d] The B-DNA and E-DNA structures were refined with the program XPLOR 3.851 (Brünger 1992), incorporating nucleic acid specific parameters (Parkinson *et al.* 1996). The A-DNA structure of $d(GGCGm^5CC)_2$ was refined using CNS (Brünger *et al.* 1998). $R_{cryst} = \Sigma_{hkl} | F_{obs} - kF_{calc} | / \Sigma_{hkl} |F_{obs}|$. $R_{free} = R_{cryst}$ for 10% of the reflections that were not used in refinement (Brünger 1992).

This was subsequently subjected to simulated annealing, followed by hand fitting of the terminal base pairs to the residual electron density observed in an Fo-Fc difference map. The refined brominated model was used as a search model to solve the structure of the methylated sequence.

A-DNA crystals of $d(GGCGm^5CC)_2$ that grew after 2-3 months from the same set of crystallization solutions and identical conditions that yielded E-DNA crystals were solved by molecular replacement using the A-DNA structure of $d(GCCGGC)_2$ (Mooers *et al.* 1995) in the program AMoRe (Navaza 1994).

2.4 Results and Discussion

2.4.1 B-DNA structure of $d(GGCGCC)_2$

The sequence $d(GGCGCC)_2$ was crystallized as B-DNA in the presence of spermine hydrochloride ($spermine^{+4}$) alone or with cobalt hexamine ($Co(NH_3)_6^{+3}$) (Figure 2.2a). In both crystal forms, the DNA duplexes stack coaxially to form the continuous columns seen in previous B-DNA crystals (Timsit and Moras 1992). The deoxyribose sugars in the four unique duplex structures determined here all fall in the family of 2'-*endo* conformations of B-DNA. The helical parameters (Figure 2.1b) show sequence-dependent variations that deviate from fiber B-DNA, but are all clearly standard B-DNA, even though $Co(NH_3)_6^{+3}$ has been shown to promote the formation of A-DNA in solution (Xu *et al.* 1993). The polycations in the

Figure 2.2. Electron density maps of $d(GGCGCC)_2$ with Co^{3+} as B-DNA and $d(GGCGm^5CC)_2$ as E-DNA and A-DNA. Electron density maps of $d(GGCGCC)_2$ with Co^{3+} (purple sphere) as B-DNA (*a*), and $d(GGCGm^5CC)_2$ as E-DNA (*b*) and as A-DNA (*c*). The 2Fo-Fc maps (contoured at 1σ) show stereoviews looking into the major groove of each structure. This figure was rendered with Raster3D (Merritt and Bacon 1997).

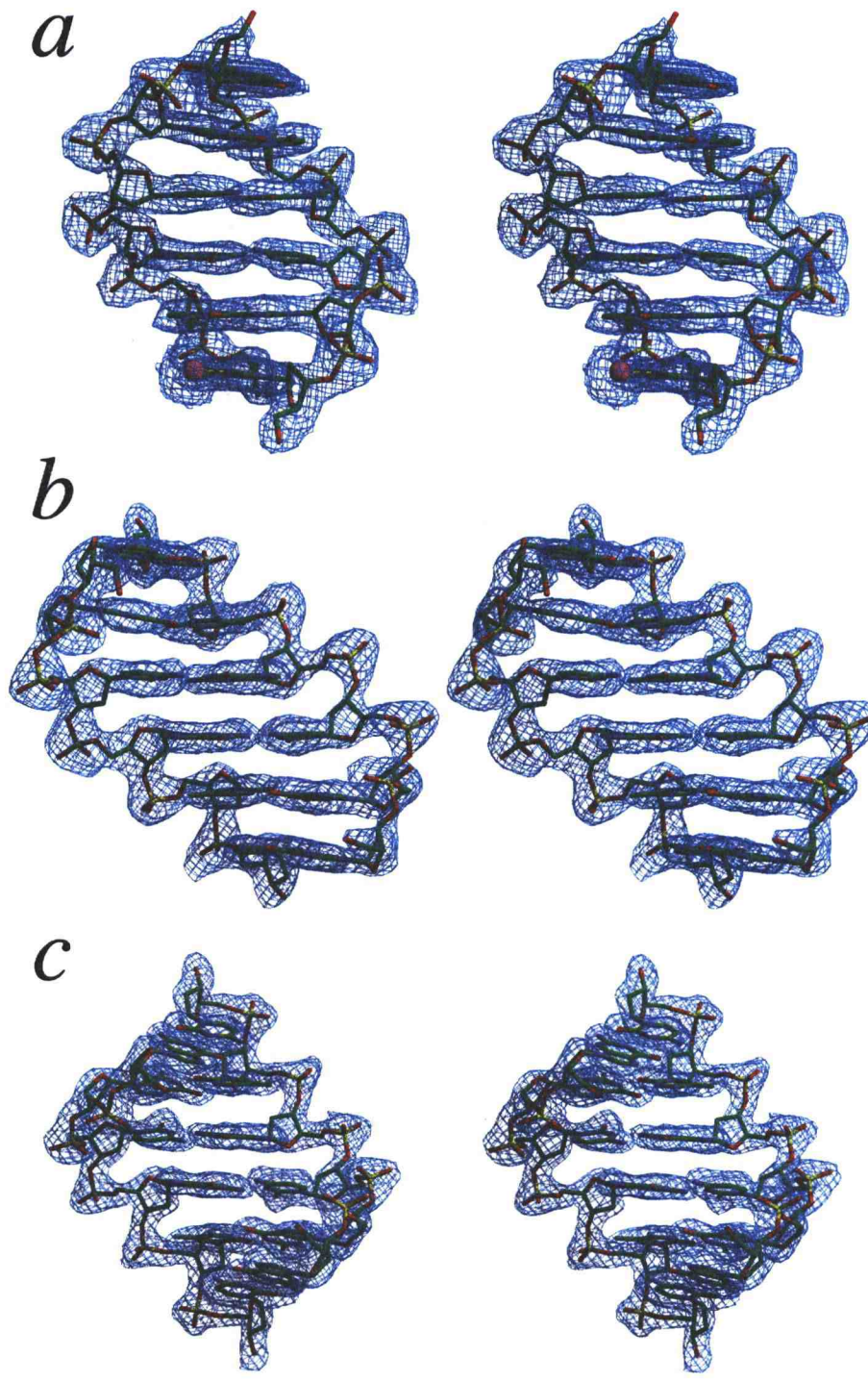

Figure 2.2

crystals act as the molecular glue that holds the lattices together. One spermine^{+4} spans the major groove of two stacked duplexes, while a second molecule bridges the backbones of two unstacked duplexes in the polyamine structure. In contrast, the metal of the cobalt structure directly cross-links the guanines of two symmetry related duplexes.

2.4.2 E-DNA structure of d(GGCGm^{5}CC)$_2$ and d(GGCGBr^{5}CC)$_2$

The sequences d(GGCGm^{5}CC)$_2$ and d(GGCGBr^{5}CC)$_2$ were crystallized from solutions that are nearly identical to the spermine^{+4} form of d(GGCGCC)$_2$. The crystal structures of d(GGCGm^{5}CC)$_2$ and d(GGCGBr^{5}CC)$_2$ are nearly identical, consisting of right-handed antiparallel double-helices (Figure 2.2b). Viewed into the helix (Figure 2.1a), the structures appear to be variants of B-DNA with base pairs lying perpendicular to and extended along the helix axis. However, we see the deep major groove and shallow minor groove that is associated with the 3'-*endo* deoxyribose sugar conformation of A-DNA. Looking down the helix, the backbone traces a squared rather than a circular cylinder. The structure, therefore, has features of both A-DNA and B-DNA; however, it is neither. This new structure is called "E-DNA" to recognize the extended helix and the eccentric trace of the backbone.

A detailed analysis (Figure 2.1b and Table 2.2) shows E-DNA to be distinct from A-DNA and B-DNA. The slight negative inclination of the base pairs is more like fiber B-DNA than even the B-DNA structures of d(GGCGCC)$_2$. However,

Table 2.2. Average helical parameters for the single crystal structures of hexanucleotides containing dC·dG, dm^{5}C·dG, and dBr^{5}C·dG base pairs as A-DNA, B-DNA, and E-DNA.

Parameter[a]	A-DNA	B-DNA	E-DNA
Helical twist	31.5° ± 3.6	34.8° ± 3.5	29.2° ±4.5 (30.6° ± 3.5)
Rise (*z*-Displacement)	2.8Å ± 0.32	3.30Å ± 0.28	3.56Å ±0.36 (3.71Å ± 0.19)
Slide	-1.99Å ± 0.27	-0.63Å ± 0.57	-2.32Å ± 0.17 (-2.39Å ± 0.13)
Roll Angle	9.0° ± 4.7	2.9° ± 4.0	2.1° ± 4.4 (0.9° ± 4.0)
x-Displacement	-4.73Å ± 0.37	-1.60Å ± 0.75	-3.38Å ± 0.51 (-3.43Å ± 0.55)
Inclination angle	15.2° ± 4.3	7.4° ± 5.7	-4.6° ± 4.1 (-6.1° ± 1.9)

[a] Mean values are shown ±1 standard deviation. Values in parentheses were calculated for E-DNA in the absence of the stacked terminal base pairs.

like A-DNA, the large negative *x*-displacement places the helix axis in the major groove. The eccentric backbone results from an even greater protrusion of the modified cytosines away from the helix axis, as evident from the more negative *x*-displacement. In addition, the shorter distance between phosphates of the modified (5.7 ± 0.2 Å) versus unmodified base pairs (6.5 ± 0.3 Å) is comparable to differences between A-DNA and B-DNA. The average rise is longer than that of B-DNA, and the average slide between base pairs is larger than in A-DNA. Thus, E-DNA is more extended and broader than A-DNA and B-DNA, and shows variations associated with the modifications to the cytosine bases.

The E-DNA crystal lattice has one terminal base pair of the duplex sitting in the minor groove of a symmetry-related duplex, which is typical of A-DNA crystals (Mooers *et al.* 1995). The opposite ends of the duplexes, however, are coaxially stacked, similar to the $d(GGCGCC)_2$ crystals. Therefore, rather than the crystal interactions defining the conformation, it appears that E-DNA defines a conglomerate crystal lattice.

Interestingly, it is the B-DNA type stacking interactions that show the greatest lattice distortions to the E-DNA structure. The stacked base pairs at these termini are dramatically under-wound, with an average helical twist of 23.7°. The average helical twist estimated from the remainder of the nucleotides is equivalent to a helical repeat of 11.8 base pairs/turn, making E-DNA an under-wound structure relative to both A- and B-DNA.

E-DNA is not a chimeric structure. *Cis*-platin, for example, sits at a junction between B-DNA and A-DNA (Takahara *et al.* 1995), while DNA/RNA hybrids have distinct A-RNA and a B-DNA strands (Arnott *et al.* 1986). These chimeric structures have the two conformations coexisting in the same molecule. A- and B-DNA have also been shown to coexist as unique structures within a single crystal lattice (Doucet *et al.* 1989), while the sequence $d(CCGCCGGCGG)_2$ has been crystallized as both A-DNA (Mayer-Jung *et al.* 1998) and B-DNA (Timsit and Moras 1994). E-DNA, however, has structural properties of both conformations, but cannot be readily classified as either a variation of B-DNA or of A-DNA, or as a chimera of the two. Finally, with its extended rise and large slide between base pairs, E-DNA distinguishes itself from the intermediate structures of $d(CCCCGGGG)_2$ (Wang *et al.* 1982) and, more recently, $d(CATGGGCCCATG)_2$ (Ng *et al.* 2000), both of which fall within the continuum between B-DNA and A-DNA.

2.4.3 A-DNA structure of $d(GGCGm^5CC)_2$

The sequence $d(GGCGm^5CC)_2$ crystallized as standard A-DNA from identical solutions that yielded E-DNA crystals of this same sequence. The difference was that the A-DNA crystals of this sequence grew after a significantly longer period of time (2-3 months as compared to 2-3 weeks for the E-DNA crystals). The asymmetric unit of the crystals was composed of two complete A-DNA duplexes, and one duplex with one terminal base pair melted. We could not

accurately account for this frayed end, and therefore, did not incorporate this model into our analysis of the structure. The conformation from these "aged" solutions had large positive inclination angles, *x*-displacements and slides, and short helical rises between base pairs (Figure 2.2c). These helical parameters along with the 3'-*endo* sugar conformations are characteristic of A-DNA (Figure 2.1b). Finally, the interduplex interactions were typical of hexanucleotide A-DNA crystals (Mooers *et al.* 1995). In short, there was nothing that set this crystal structure apart from other undistorted A-DNA structures.

2.4.4 The B-DNA to A-DNA transition goes through an E-DNA intermediate

One of the truly unique results from this study is that a single sequence, $d(GGCGCC)_2$, has been crystallized as standard B-DNA and A-DNA double helices, and now shown to form a novel structure, E-DNA, which is intermediate between the two. The requirement of methylation in inducing A-DNA in this sequence is consistent with previous observations that dm^5C nucleotides induce the transition from B-DNA to A-DNA (Frederick *et al.* 1987; Mooers *et al.* 1995). The shorter period of time required to grow crystals of E-DNA from the identical solutions that eventually yielded A-DNA indicates that this intermediate was kinetically trapped by crystallization. We show here that the methyl group facilitates the conversion of the sugar conformation from 2'-*endo* to 3'-*endo*, which is apparently the rate-limiting step for the B- to A-DNA transition. Thus, the transition from B-DNA to A-DNA occurs through E-DNA as a discrete

intermediate, and at least for the methylation induced transition, this intermediate is elongated and under wound relative to either B-DNA or A-DNA. The structural convolutions that lead from B-DNA to E-DNA and finally to A-DNA are apparent from the views into the major grooves of the three crystal structures (Figure 2.2). Crystallization of the A-DNA structure appears to require an additional loss in water activity that would result from the longer equilibrium time during crystal growth. This supports our contention that E-DNA is more hydrated than A-DNA in this sequence.

2.4.5 Solvent interactions in E-DNA

The deep major groove and shallow minor groove, along with the relatively noninclined base pairs result in a more exposed surface for E-DNA. The solvent accessible surface area of a base pair in E-DNA (288 $Å^2$) is >3 $Å^2$ more exposed than in A-DNA (285 $Å^2$) or B-DNA (284 $Å^2$). A well-defined set of water molecules is observed in the major groove of E-DNA, all hydrogen bonded to and in the plane of the base pairs. However, the large rise and slide between base pairs keep these waters from forming a regular spine. Instead, the waters in E-DNA sit on the exposed face of the adjacent stacked base pair (Figure 2.3a). The solvent structure in the minor groove is obscured by the symmetry-related duplex that sits against this groove.

E-DNA provides a structural rationale for the higher rate of spontaneous deamination in methylated cytosines (Figure 2.3b), which leads to the transition

Figure 2.3. Waters in the CpG dinucleotides of B-DNA and E-DNA. *a.* Stereoview of waters (gold spheres) hydrogen bonded (broken lines) to the guanine at the CpG steps in $d(GGCGCC)_2$ as B-DNA, and in $d(GGCGm^5CC)_2$ and $d(GGCGBr^5CC)_2$ as E-DNA. The waters are overlaid relative to the guanine N7 nitrogen of an average structure built from 3 unique CpG steps found in each conformation (the C4 carbon of the cytosine base is colored black). *b.* Mechanism for the spontaneous deamination of cytosine to uracil (Carter 1995). The nucleophilic attack of a water at the C4 carbon forms a hemiaminal intermediate. Release of ammonia results in a tautomer, which subsequently rearranges to uracil. *c.* Model of the hemiaminal intermediate in B-DNA and E-DNA. The hemiaminal intermediate at the CpG step was modeled by adding a hydroxyl group to the C4 carbon (colored black) of the cytosine base, followed by geometry optimization using the AMBER (Weiner *et al.* 1984) forcefield as implemented in the program InsightII (BioSym/MSI). The starting positions of the waters and the starting cytosine base are shown as transparent overlays.

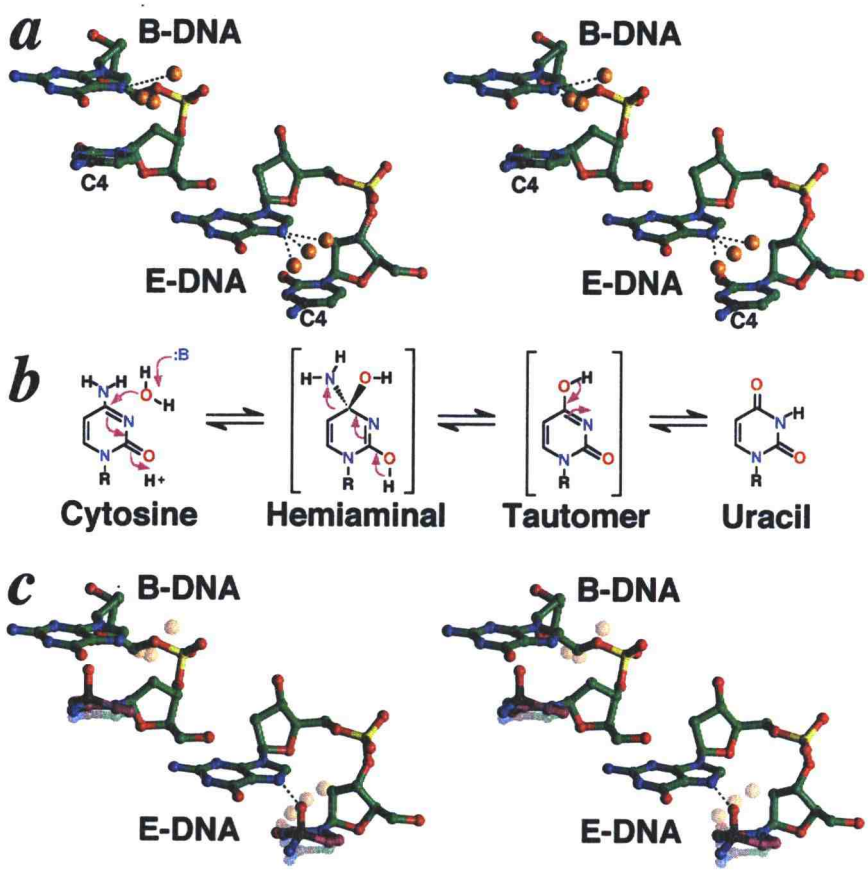
a
B-DNA
C4
E-DNA
C4
B-DNA
C4
E-DNA
C4
b
:B
Cytosine
Hemiaminal
Tautomer
Uracil
c
B-DNA
E-DNA
B-DNA
E-DNA

Figure 2.3

mutation of $dm^5C{\cdot}dG$ to $dT{\cdot}dA$ base pairs. The mutation rate is 21-fold higher when the cytosine bases in duplex DNA are methylated (Zhang and Mathews 1994). We compared the structure of the central d(CpG) step and the associated waters of $d(GGCGCC)_2$ as B-DNA and $d(GGCGm^5CC)_2$ and $d(GGCGBr^5CC)_2$ as E-DNA to illustrate how conformation rather than specific methylation would facilitate the deamination reaction. The CpG step of each structure has a water molecule hydrogen bonded to the basic N7 nitrogen and lying in the plane of the guanine base. The waters in the B-DNA dinucleotides are located 4.5 Å to >6 Å from the edges of the stacked cytosines (Figure 2.3a). However, the shifted base pairs in E-DNA position the solvent molecules just above and within 3.8 Å to 4.2 Å of the exposed C4 carbon of the pyrimidine base. In addition, these waters in E-DNA are nearly perpendicular to the cytosine base plane, a geometry that would facilitate nucleophilic attack of the aromatic ring. When the cytosine of the d(CpG) step in E- DNA is modeled as the hemiaminal intermediate of the deamination reaction, the resulting hydroxyl oxygen is 2.0 to 3.0 Å from the N7 nitrogen and in the plane of the guanine base (Figure 2.3c). The CpG dinucleotide in this E-DNA model, therefore, can potentially stabilize the hemiaminal intermediate through hydrogen bonds.

2.5 Acknowledgments

We thank B. H. M. Mooers and the P. A. Karplus' laboratory for helpful discussion, and K. E. van Holde, C. K. Mathews, and W. C. Johnson, Jr. for reading this manuscript. This work was supported by the National Science Foundation, the Oregon American Cancer Society, and the Environmental Heath Science Center at OSU. X-ray facilities were funded in part by the M.J. Murdock Charitable Trust.

Chapter 3

A Crystallographic Map of the Transition from B-DNA to A-DNA

Jeffrey M. Vargason, Keith Henderson, and P. Shing Ho

Published in *Proc. Natl. Acad. Sci. USA*,

The National Academy of Sciences, Washington, D.C., USA

2001, **98**, 7265-7270

3.1 Summary

The transition between B-DNA and A-DNA was first observed nearly 50 years ago. We have now mapped this transformation through a set of single-crystal structures of the sequence $d(GGCGCC)_2$, with various intermediates being trapped by methylating or brominating the cytosine bases. The resulting pathway progresses through 13 conformational steps, with a novel composite structure that pairs A-nucleotides with complementary B-nucleotides serving as a distinct transition intermediate. The details of each step in the conversion of B-DNA to A-DNA are thus revealed at the atomic level, placing intermediates for this and other sequences in the context of a common pathway.

3.2 Introduction

The transformation from B-DNA in fibers (Watson and Crick 1953) to dehydrated A-DNA (Franklin and Gosling 1953) was one of the first reversible transitions observed between conformations in a biomolecule. Although B-DNA is recognized as the standard form of the double helix, the junction between B- and A-DNA appears as a target for the anticancer drug cisplatin (Takahara *et al.* 1995), and an A-type conformation with exaggerated base pair inclinations is induced by TATA-binding protein, the transcriptional promoter in eukaryotes (Kim *et al.* 1993;

Kim *et al.* 1993). Consequently, an understanding of how B-DNA converts to A-DNA (the B-A transition) is significant both for its history, and for understanding how DNA structure affects cellular function. In the current study, methylated and brominated variants of the sequence $d(GGCGCC)_2$ were crystallized as standard A-DNA and several intermediate conformations, the most important being a composite structure in which each strand is half A-DNA and half B-DNA, with the A-nucleotides paired with complementary B-nucleotides across the duplex. Thus, along with structures from previous work (Vargason *et al.* 2000), we have assembled 13 single-crystal structures that define a pathway for the B-A transition in this single sequence.

In a B-A transition, the long and narrow B-duplex–with its 10 to 10.5 base pair repeat, 3.4 Å rise, and base pairs stacked at the center of the helix–is converted to the underwound and compact A-DNA structure–characterized by an 11 base pair repeat, a 2.6 Å rise, and base pairs that are inclined up to 20° and displaced by ~4 Å so that they essentially wrap around the helix axis (Arnott 1999). The geometries of the deoxyribose sugars are also converted from C2'-*endo* in B-DNA (with the C2'-carbon puckered above the furanose plane towards the nucleobase) to C3'-*endo* in A-DNA. Molecular mechanics simulations suggest that many of these structural features are transformed in a concerted manner (Cheatham and Kollman 1996; Cheatham *et al.* 1997; Cheatham and Kollman 1997), leading to the question of whether a discrete intermediate exists in the B-A transition.

There are a growing number of DNA crystal structures that exhibit properties of both A- and B-DNA. These include $d(CCCCGGGG)_2$, an A-DNA with a relatively extended rise of 3.1 Å (Wang *et al.* 1982), and $d(CGCCCGCGGGCG)_2$, a kinked A-DNA structure with B-type helices at the ends (Malinina *et al.* 1999). More recently, the central six base pairs of $d(CATGGGCCCATG)_2$ were defined as an A/B-intermediate conformation in which the base pairs are displaced, but not highly inclined relative to the helix axis (Ng *et al.* 2000). Our own studies have shown that methylating or brominating cytosines in $d(GGCGCC)_2$ induces a unique extended conformation that is neither B-DNA nor A-DNA (Vargason *et al.* 2000). The sequence $d(GGCGCC)_2$ crystallized as standard B-DNA, while the structures of $d(GGCGm^5CC)_2$ and $d(GGCGBr^5CC)_2$ (m^5C and Br^5C are 5-methylcytosine and 5-bromocytosine respectively) have longer rises than B-DNA and are broader, with a larger slide between base pairs, than A-DNA. All of these conformations, however, have C3'-*endo* sugars, and therefore are generally considered to be allomorphic forms of A-DNA rather than true transition intermediates. We present here a set of single-crystal structures that represent the missing conformational links required to define a common B-A transition pathway for a single DNA sequence.

3.3 Materials and Methods

All deoxyoligonucleotides were synthesized using phosphoramidite chemistry on an Applied Biosystems DNA synthesizer in the Center for Gene Research and Biotechnology at Oregon State University. The oligonucleotides were filtered through a Sephadex G-25 column, lyophilized, redissolved in 15 mM sodium cacodylate buffer (pH 7.0), and used for crystallization without further purification. Crystals were grown at room temperature by vapor diffusion in sitting drop setups with initial solutions containing 0.7 mM DNA, 25 mM sodium cacodylate buffer (pH 6), 0.8 mM $MgCl_2$, and 0.1-1.2 mM spermine tetrahydrochloride, and were equilibrated against a reservoir of 15% (v:v) 2-methyl-2,4-pentanediol. The crystal of $d(GGCGCC)_2$ was grown, as above, with the addition of 2 mM $Co(NH_3)_6^{3+}$. These conditions are similar to those of previous studies (Vargason *et al.* 2000).

X-ray diffraction data for the Co^{3+}-form of $d(GGCGCC)_2$, $d(GGm^5CGm^5CC)_2$, $d(GGm^5CGCC)_2$, and $d(GGBr^5CGBr^5CC)_2$ were collected at room temperature using CuK_α radiation from a Rigaku RUH3R generator and an R-AXIS IV image plate detector. Multiwavelength x-ray diffraction data for $d(GGBr^5CGCC)_2$ were collected on beamline 5.0.2 at the Advanced Light Source in Berkeley, CA at liquid nitrogen temperature using a Quantum IV detector. Three wavelengths corresponding to the peak of bromine absorption (0.92052Å), the inflection (0.92065Å), and a remote (0.90836Å) were used in CNS (Brünger *et*

al. 1998) to determine the phases for d(GGBr^{5}CGCC)$_2$. Diffraction data collected at the peak wavelength was used in SHELXL-97 (Sheldrick and Schneider 1997) for refinement.

The crystals of d(GGm^{5}CGCC)$_2$ and d(GGm^{5}CGm^{5}CC)$_2$ were isomorphous with those of d(GGBr^{5}CGCC)$_2$; consequently, their structures were solved by molecular replacement using the four unique duplexes of the d(GGBr^{5}CGCC)$_2$ structure as the starting model. The bromine atoms were removed and methyl groups added to the appropriate cytosines. The initial values of R_{cryst} and R_{free} were 42.6% and 41.6% respectively for d(GGm^{5}CGCC)$_2$, and 51.2% and 51.0% for d(GGm^{5}CGm^{5}CC)$_2$. The structures were then refined in CNS (Brünger *et al.* 1998) using the maximum likelihood target while maintaining the same cross validation set as d(GGBr^{5}CGCC)$_2$, yielding the final refinement statistics in Table 3.1. The structure of d(GGBr^{5}CGBr^{5}CC)$_2$ was solved by molecular replacement using the previously solved extended structure of d(GGCGBr^{5}CC)$_2$ (Vargason *et al.* 2000) and refined using the same technique. The structure of the Co^{3+}-form of d(GGCGCC)$_2$ was re-refined with higher resolution data using the previously solved and refined structure (Vargason *et al.* 2000). All datasets were reduced using the programs Denzo and Scalepack from the HKL package (Otwinowski and Minor 1997) and structures, with the exception of d(GGBr^{5}CGCC)$_2$, were refined with CNS (Brünger *et al.* 1998) incorporating nucleic acid specific parameters (Parkinson *et al.* 1996).

Table 3.1. Data collection and refinement statistics

Sequence:	GGBr^{5}CGCC	GGm^{5}CGCC	GGm^{5}CGm^{5}CC	GGBr^{5}CGBr^{5}CC	GGCGCC+Co
Structure(s):	**f, i, j, l**	**e, i, k, m**			**a**
Data Collection					
Space Group	$P3_221$	$P3_221$	$P3_221$	$P4_32_12$	$P4_122$
Unit Cell Dimensions	$a = b = 41.1$ Å, $c = 175.1$ Å $\gamma = 120°$	$a = b = 41.9$ Å, $c = 176.7$ Å $\gamma = 120°$	$a = b = 42.3$ Å, $c = 179.4$ Å $\gamma = 120°$	$a = b = 60.7$ Å, $c = 24.6$ Å	$A = b = 42.6$ Å, $c = 63.5$ Å
Resolution (Å)	20.0 - 1.45	20.0 - 1.9	20.0 - 2.4	20.0 - 2.8	20.0 - 2.0
Total reflections (unique)	628716 (58430) including Bijvoets	107358 (14930)	49399 (7269)	10777 (1215)	18846 (4207)
Completeness (%)[a]	99.9 (99.6)	99.0 (96.5)	91.7 (61.5)	95.2 (99.1)	97.5 (95.0)
R_{meas}(%)[a,b]	11.0 (66.8)	4.5 (47.5)	6.3 (40.0)	8.1 (54.4)	6.9 (51.8)
Refinement					
R_{cryst} (R_{free}) (%)[c]	13.8 (18.7)	18.5 (20.4)	17.3 (20.0)	21.1 (25.7)	22.8 (24.5)
DNA (Solvent) atoms	968 (234)	968 (115)	976 (75)	244 (2)	240 (37)
RMSD Bond lengths (Å)	0.015	0.003	0.003	0.005	0.003
RMSD Bond angles (°)	2.29	0.81	0.82	0.95	1.01

[a] Values in parentheses refer to the highest resolution shell

[b] $R_{meas} = \Sigma_{hkl} \sqrt{(n / n-1)} \Sigma_i | \langle I \rangle_{hkl} - I_{hkl,i} | / \Sigma_{hkl} \Sigma_i | I_{hkl,i} |$ where I_{hkl} is the intensity of a reflection, $\langle I \rangle_{hkl}$ is the average of all observations of this reflection and its symmetry equivalents, and n is the multiplicity (Diederichs and Karplus 1997).

[c] $R_{cryst} = \Sigma_{hkl} | F_{obs} - kF_{calc} | / \Sigma_{hkl} |F_{obs}|$. $R_{free} = R_{cryst}$ for 10% of the reflections that were not used in refinement (Brünger 1992).

3.4 Results

3.4.1 Single crystal structures of d(GGBr^{5}CGCC)$_2$ and d(GGm^{5}CGCC)$_2$

For the current study, we have grown crystals of the sequences d(GGm^{5}CGm^{5}CC)$_2$, d(GGm^{5}CGCC)$_2$, and d(GGBr^{5}CGCC)$_2$, which are all isomorphous to each other but differ from all previous oligonucleotide sequences (Table 3.1). The crystal structures of these sequences were seen to include a novel conformation that is a composite of A-DNA and B-DNA. The structure of d(GGBr^{5}CGCC)$_2$ was solved to 1.6 Å resolution using multiwavelength anomalous diffraction (MAD) phasing with bromine as the anomalous scatterer (Hendrickson and Ogata 1997). The resulting experimental electron density maps allow unbiased interpretations of four independent DNA double helices in the asymmetric unit of the crystal, as well as the detailed conformational features of the individual nucleotides (Figure 3.1A). This structure was subsequently used to determine the structures of d(GGm^{5}CGm^{5}CC)$_2$ and d(GGm^{5}CGCC)$_2$ by molecular replacement (Table 3.1). The DNAs of these sequences are arranged with three duplexes forming a planar canopy in the *a-b* plane, while a fourth duplex extends perpendicular from near the center of this planar assembly. One duplex in the plane is a classic A-DNA structure, characterized by C3'-*endo* sugar puckers, and highly inclined and displaced base pairs. The remaining two helices in the plane have C3'-*endo* sugars and basepairs displaced from but not highly inclined relative to the helix axis, like the A/B-intermediate of d(CATGGGCCCATG)$_2$ (Ng *et al.* 2000).

Figure 3.1. Single crystal structures of d(GGBr^{5}CGCC)$_2$ and d(GGm^{5}CGCC)$_2$. A. Experimental electron density map derived from the 1.6 Å MAD phased X-ray diffraction data of d(GGBr^{5}CGCC)$_2$. The electron-density, after density modification, of one base pair is shown with the final refined model of this GC base pair included for reference (figure created with Bobscript (Esnouf 1999)). B. Comparison of the δ and χ-torsion angles and the sugar conformations of the nucleotides in d(GGBr^{5}CGCC)$_2$ (open symbols) and d(GGm^{5}CGCC)$_2$ (closed symbols). Nucleotides with C2'-*endo* type sugars are indicated by circles, with the C3'-*endo* sugars by squares, and those with intermediate O4'-*endo* sugars by diamonds. Nucleotides in the two structures are numbered 1-6 for one strand and 7-12 for the complementary strand (both in the 5' to 3'-directions). The values of δ and χ for previous single-crystal structures (comprising high resolution–better than 2Å–crystal structures deposited in the Nucleic Acid Database (Berman *et al.* 1992)) of B-DNA are defined by the hashed oval and those of A-DNA are defined by the open oval (adapted from Lu *et al.* (Lu *et al.* 2000)).

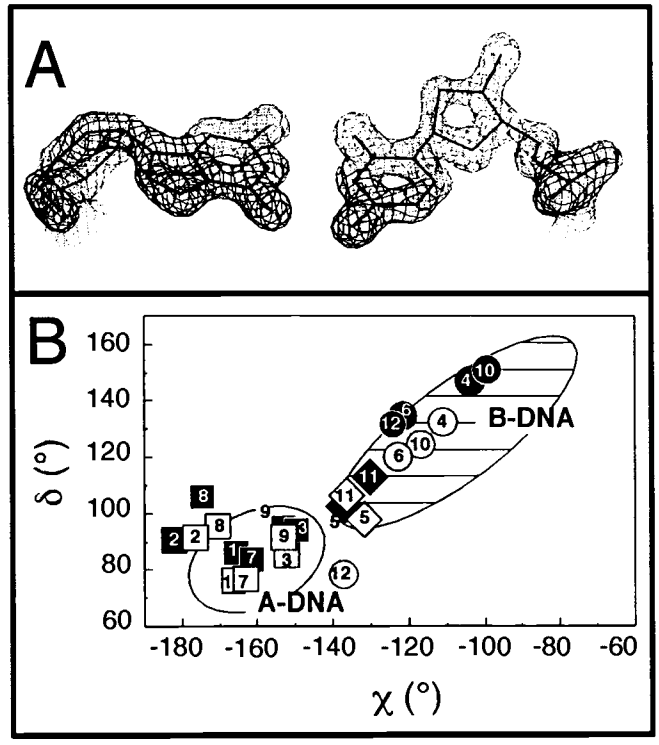

Figure 3.1

The fourth duplex that bridges these planes of A-like structures is a novel composite double-helical conformation. The three nucleotides at the 5'-end of each strand have A-type C3'-*endo* sugars (Figure 3.1B) while the last three nucleotides have C2'-*endo* and C1'-*exo* sugars, which fall within the B-type C2'-*endo* family, and an O4'-*endo* sugar, which is the lowest energy intermediate between the C2'- and C3'-*endo* conformations. In addition, the δ and χ dihedral angles show the three nucleotides at the 5'-end of each strand to be A-DNA, and those of the three 3'-nucleotides to be B-DNA. Thus, the three A-type nucleotides at the 5'-end of each strand are paired with the complementary B-type nucleotides at the 3'-end of the complementary strand. This mixture of A- and B-DNA base pairing is similar to the chimeric A-RNA/B-DNA conformation seen by fiber diffraction (Arnott *et al.* 1986), but differs from the A-RNA structures of DNA/RNA hybrids (Egli *et al.* 1992; Horton and Finzel 1996). The composite conformation seen here is analogous to the junction between neighboring B-DNA and A-DNA conformations first proposed by Arnott (Selsing *et al.* 1979), and to the intermediate for the B-A transition in poly(dG)·poly(dC) suggested from Raman spectroscopy (Nishimura *et al.* 1986).

We have also shown that the sequence $d(GGBr^5CGBr^5CC)_2$ is nearly identical to the previously solved structures of $d(GGCGm^5CC)_2$ and $d(GGCGBr^5CC)_2$, both of which had extended helical rises compared to both A- and B-DNAs (Vargason *et al.* 2000). Finally, we have resolved the B-DNA

structure of $d(GGCGCC)_2$ in its cobalt form to a higher (2.0 Å) resolution than previously reported (Vargason *et al.* 2000).

3.4.2 The B-A transition pathway

The structures of $d(GGm^5CGCC)_2$ and $d(GGBr^5CGCC)_2$, along with the previous B-conformations of $d(GGCGCC)_2$ and extended conformations of $d(GGCGm^5CC)_2$ and $d(GGCGBr^5CC)_2$ form a set of 13 unique structures (designated **a** to **m**, Table 3.2) that define a common transition pathway (Figure 3.2). The structures were sorted according to their increasingly negative *x*-displacement, starting with the structure that most closely resembles the fiber B-DNA conformation (**a**) and ending with the structure that most closely resembles fiber A-DNA (**m**) (Figure 3.3). This yields a continuous and monotonic transformation of duplexes with progressively deeper major grooves. Central to this transition pathway are the composite intermediates (**e** and **f**) that bridge the gap between the B-type and A-type conformations. The structures arranged in this way also show a continuous progression of the conformations away from standard B-DNA. The root-mean-square deviation of the atoms for each conformation relative to standard B-DNA (**a**) systematically increases as the structures proceed from **a** to **m**. Sorting these structures according to other parameters, for example by increasing inclination of the base pairs, results in non-logical pathways that intermix B-DNAs and A-DNAs throughout the transition.

Table 3.2. Crystallized sequences and duplex conformations

Sequence[a]	Resolution, Å	Conformation	Structure	Rise/Patterson Rise[b], Å	Twist (°)
GGCGCC		Fiber B-DNA	**B**	3.38	36.0
GGCGCC + Co^{3+}	2.0	B-DNA	**a**	3.38±0.08/3.5	35.0±2.3
GGCGCC +	2.7	B-DNA	**b**	3.33±0.28/3.4	33. 9±1.1
$spermine^{4+}$			**c**	3.35±0.28/3.5	35. 9±4.7
			d	3.17±0.47/3.5	33.6±2.8
GGm^5CGCC	1.9	Composite	**e**	3.36±0.46/3.1	31.7±4.9
		A/B	**i**	2.93±0.24	31.3±2.5
		A/B	**k**	3.06±0.11	30.4±2.2
		A-DNA	**m**	2.68±0.27	30.9±3.7
$GGBr^5CGCC$	1.4	Composite	**f**	3.07±0.30/3.2	31.9±3.4
		A/B	**i**	2.78±0.26	31.5±3.0
		A/B	**j**	3.02±0.11	31.0±0.9
		A-DNA	**l**	2.63±0.33	31.0±2.9
$GGCGBr^5CC$	2.25	Extended	**g**	3.57±0.40/3.7	30.0±4.7
$GGCGm^5CC$	2.2	Extended	**h**	3.55±0.35/3.7	28.4±4.9
GGCGCC		Fiber A-DNA	**A**	2.56	32.7

[a] The crystal structures of $d(GGCGCC)_2$, $d(GGCGm^5CC)_2$, and $d(GGCGBr^5CC)_2$ were previously described (Vargason *et al.* 2000). We have resolved the Co^{3+} form of $d(GGCGCC)_2$ here to higher resolution than before.
[b] Values for Patterson rise (the rise between base planes in the Patterson maps calculated from the refined structures) are included as measures of an intrinsic helical parameter for comparison with the values determined by CURVES (Lavery and Sklenar 1989).

Figure 3.2. The 13 unique conformations seen in the single crystal structures of $d(GGCGCC)_2$, $d(GGm^5CGCC)_2$, $d(GGBr^5CGCC)_2$, $d(GGCGm^5CC)_2$, and $d(GGCGBr^5CC)_2$. Structures (**a** to **m**) are arranged as described in the text and Table 3.2. Each structure is viewed down the helix axis (purple dot), and into the helix axis (purple line). Conformations **a**-**d** are B-type helices (green labels), **e**-**f** are composite helices (blue labels), **g**-**h** are extended intermediates (purple labels), **i**-**m** are A-DNA allomorphs, starting with A/B-intermediates (red labels) and progressing continuously to standard A-DNA (orange labels).

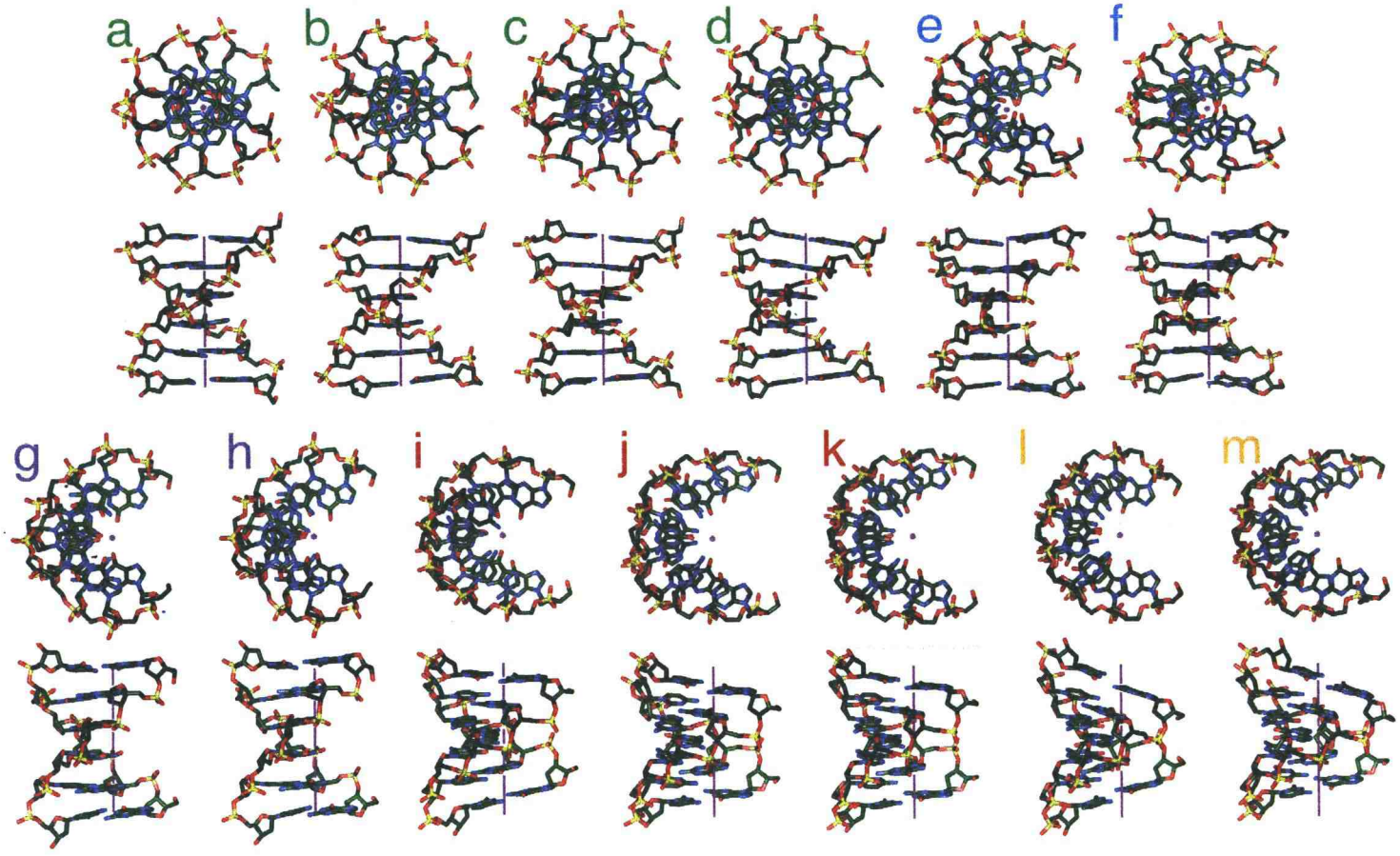

Figure 3.2

Figure 3.3. Helical parameters of the crystal structures of $d(GGCGCC)_2$ and its methylated and brominated analogs. The *x*-displacement, inclination of the base pairs, pseudorotation phase angle, base pair slide, and z_P are plotted, from top to bottom, for the 13 unique helical conformations (**a** to **m**) arranged along a common transition pathway according to their *x*-displacement. Helical parameters were calculated with Curves 5.2 (Lavery and Sklenar 1989) with the exception of z_P, which was calculated by X3DNA (Lu *et al.* 2000). The average values for the parameters are shown for each structure, with the bars extending above and below the average reflecting the values for the individual base pairs or dinucleotide steps. Values for ideal B-DNA and A-DNA in each plot are labeled B and A, respectively. Negative *x*-displacements reflect shifts of the base pairs away from the helix axis towards the minor groove, negative inclination angles arise from base pairs being tipped away from perpendicular of the helix axis and towards the 3'-end of each strand, and slide has the stacked base pairs displaced relative to each other along their long axes (Lavery and Sklenar 1989). Pseudorotation phase angles reflect the conformation of the deoxyribose sugars (labeled according to their respective *endo*-type families), with open circles representing the pseudorotation angle of the end nucleotides and closed circles the internal nucleotides of each strand. The parameter z_P measures the displacement of the phosphates along each strand away from the midpoint between two stacked base pairs (El Hassan and Calladine 1997).

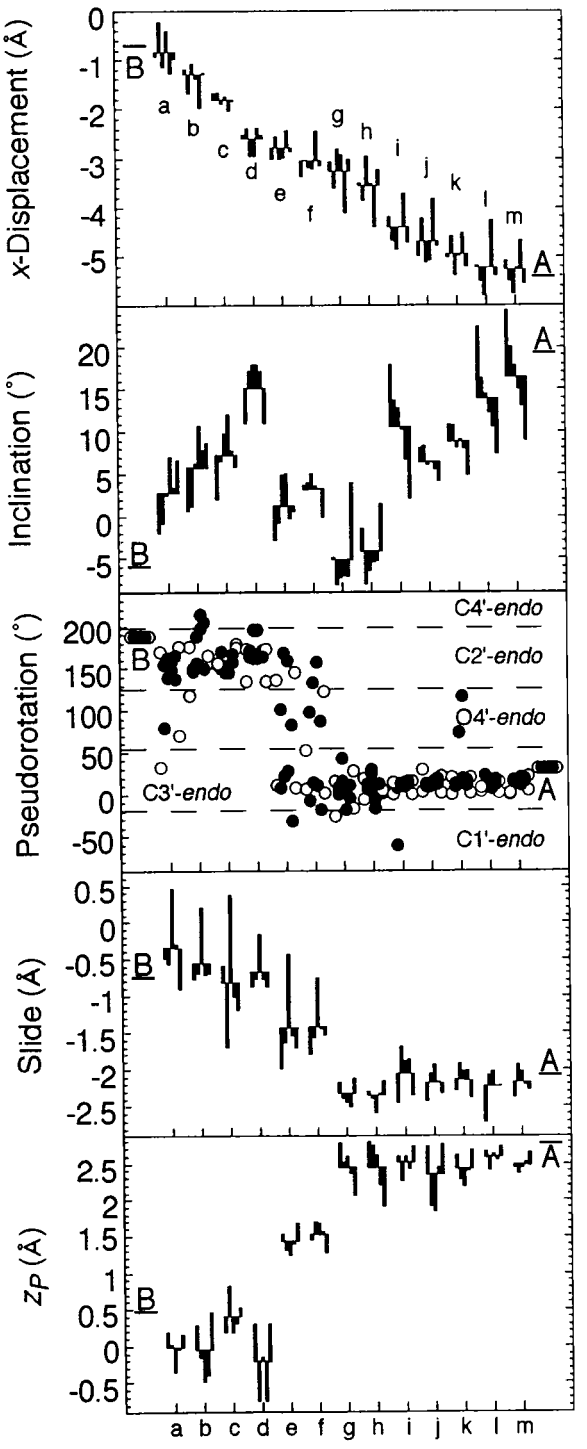
x-Displacement (Å)
0
-1
-2
-3
-4
-5
B
A
a
b
c
d
e
f
g
h
i
j
k
l
m
Inclination (°)
20
15
10
5
0
-5
Pseudorotation (°)
200
150
100
50
0
-50
C4'-endo
C2'-endo
O4'-endo
C3'-endo
C1'-endo
Slide (Å)
0.5
0
-0.5
-1
-1.5
-2
-2.5
zP (Å)
2.5
2
1.5
1
0.5
0
-0.5
a b c d e f g h i j k l m

Figure 3.3

The pathway as we have defined it is consistent with a cooperative B-A transition according to the sugar puckers (as measured by the pseudorotation phase angle), the slide between base pairs, and the displacement of the phosphates reflected in the parameter z_P (El Hassan and Calladine 1997) (Figure 3.3). These three helical parameters, considered to be reliable discriminators between the B- and A-DNA forms (Lu *et al.* 2000), are highly correlated. All of the B-type duplexes (**a** to **d**) have C2'-*endo* type sugar puckers, small base pair slides, and phosphates positioned equidistant between the stacked base pairs of each dinucleotide step ($z_P \approx 0$). The A-type structures (**g** to **m**) have C3'-*endo* sugars, large negative base pair slides, and phosphates pushed towards the 3'-end of each dinucleotide ($z_P \approx 2.5$ Å). The composite conformations (**e** and **f**) are truly transition intermediates, mixing both C2'-*endo* and C3'-*endo* type sugars in the same duplex, and having intermediate slides. The intermediate values of z_P reflect the phosphates being pushed towards the 3'-end for the A-type nucleotides, while the phosphates of the B-nucleotides remain at the midpoint between base pairs.

There is a simple relationship between displacement and inclination for the B-DNA structures (**a** to **d**)–as the base pairs become displaced from the helix axis, they are more inclined. In the extreme case, the structure **d** is nearly halfway to A-DNA in terms of the displacement and inclination of its base pairs. A conformation with A-type base stacking but B-type sugars had been proposed as the intermediate for the B-A transition of $d(CCCCGGGGG)_2$ in solution (Trantírek *et al.* 2000).

Displacement and inclination are also correlated in the A-type structures (**g** to **m**), but the trend is not simply an extension from the B-DNA structures. The first set of A-type structures (**g** and **h**) show no base pair inclination. Indeed, the opposite trend is seen as the structures actually switch from B-DNA to A-DNA (**d** to **h**), with the base pairs reset to the inclination of fiber B-DNA. From **g** to **m**, we see the DNA base pairs becoming more displaced, inclined, and compact in a near continuous manner, passing through a set of conformations that are similar to the A/B-intermediates of $d(CATGGGCCCATG)_2$ (Ng *et al.* 2000).

All variants of this sequence were crystallized from nearly identical solutions, indicating that the conformations are associated with modifications to the cytosine bases and are not artifacts of crystallization. This provides the rationale for placing these structures along a common pathway. The coexistence of the composite and A/B-intermediates, and standard A-DNA within the crystals of $d(GGm^5CGCC)_2$ and $d(GGBr^5CGCC)_2$ indicate that these three forms are energetically related. The crystallization of both $d(GGm^5CGCC)_2$ and $d(GGm^5CGm^5CC)_2$ as the composite intermediate and $d(GGCGm^5CC)_2$ as the extended intermediate show that the two conformations are dependent only on the methylation patterns; therefore, they are energetically similar and should be readily interconvertible. The extended conformation of $d(GGCGm^5CC)_2$ had previously been shown to be a kinetically trapped intermediate that leads ultimately to the formation of A-DNA (Vargason *et al.* 2000). Finally, it is evident that $d(GGCGCC)_2$ and its variants crystallize in conformations that assume several

possible crystal lattices and, therefore, the conformations are not defined explicitly by crystal packing.

3.5 Discussion

We describe here the stepwise transition between two distinct conformations of a macromolecule, in this case from B-DNA to A-DNA. This B-A transition shows many of the features first proposed by Drew and Calladine for this mechanism (Calladine and Drew 1984). Sugar pucker and slide are highly correlated throughout the transition, while inclination is highly variable. Indeed, the base pairs can be highly inclined in B-type DNAs, but they must be reset back to their non-inclined geometries prior to continuing through the transformation to A-DNA. This is generally consistent with the Drew and Calladine mechanism in which slide precedes inclination. However, the intermediate structures seen here also show that the change in sugar pucker is associated with an extension and unwinding of the helix.

The transition mapped through structures **a** to **m** incorporates a set of distinct conformational intermediates. We were able to trap these intermediates because the $d(GGCGCC)_2$ sequence is resistant to forming A-DNA. This sequence was predicted from A-DNA propensity energies (Basham *et al.* 1995) to favor B-

DNA, while a permutation of the sequence–$(GCCGGC)_2$–was predicted to be and was crystallized only as canonical A-DNA (Mooers *et al.* 1995).

In this transition pathway, B-DNA is first transformed to the composite intermediates in which half of each strand in the duplex is A-DNA (Figure 3.4A). The C3'-*endo* sugars are associated with a slide between base pairs and extension of the backbone of the A-type nucleotides at the 5'-end of each strand. The complementary B-type nucleotides, however, oppose these perturbations; consequently, the composite A-B base pairs have only intermediate values for slide, and are buckled to accommodate the backbone extension.

In the next series of steps, the transformation goes through the extended intermediates (**g** and **h**). With all nucleotides now adopting C3'-*endo* sugar conformations, the slide and extension between base pairs are no longer opposed and both reach their maximum values (Figure 3.4B). Although **g** and **h** could be considered A-type conformations, they are extended (Patterson rise, which measures the intrinsic rise between base pairs (Vargason and Ho 2001) = 3.7 Å, Table 2), under wound (helical repeat $\geq$12 base pairs/turn), and have exaggerated base pair slides; therefore, they should be considered to be at the interface between the transition intermediates and true A-DNA structures. It is not unreasonable that the transition between conformational states of DNA goes through an intermediate that bears little resemblance to either the starting or ending states. From this point, the DNA duplex becomes more compact, with a shortened rise, and rewinds itself to an 11 base pairs/turn repeat (Figure 3.4C). In the final stages, the transition to

Figure 3.4. Stereoview of the individual steps of the B-DNA to A-DNA transition at the central base pairs of the double-helix. Each panel shows the view into the major groove of the three base pairs (G2-C11, C3-G10, and G4-C9) from each conformational class superimposed on base pairs from the next conformational class along the transition pathway. The atoms of the G4-C9 base pairs (lower base pairs) serve as the common reference to superimpose the structures. The carbons and phosphorus atoms of the base pairs are colored to distinguish between the conformations. A. The transformation of standard B-DNA structure (**d**, green) to the composite intermediate (**e**, blue) shows the effect of changing half the sugars from C2'-*endo* to C3'-*endo*. B. The transformation from **e** to the extended intermediate (**g**, purple) involves the complete conversion of the sugar puckers, resulting in the increase in the rise and slide, and unwinding of base pairs. C. The transition to an A/B intermediate (**j**, red) results in the reduction of the rise and slide, but with very little inclination of the base pairs. D. The final conversion to canonical A-DNA (**m**, orange) shows the compression of the major groove resulting from the inclination of the base pairs.

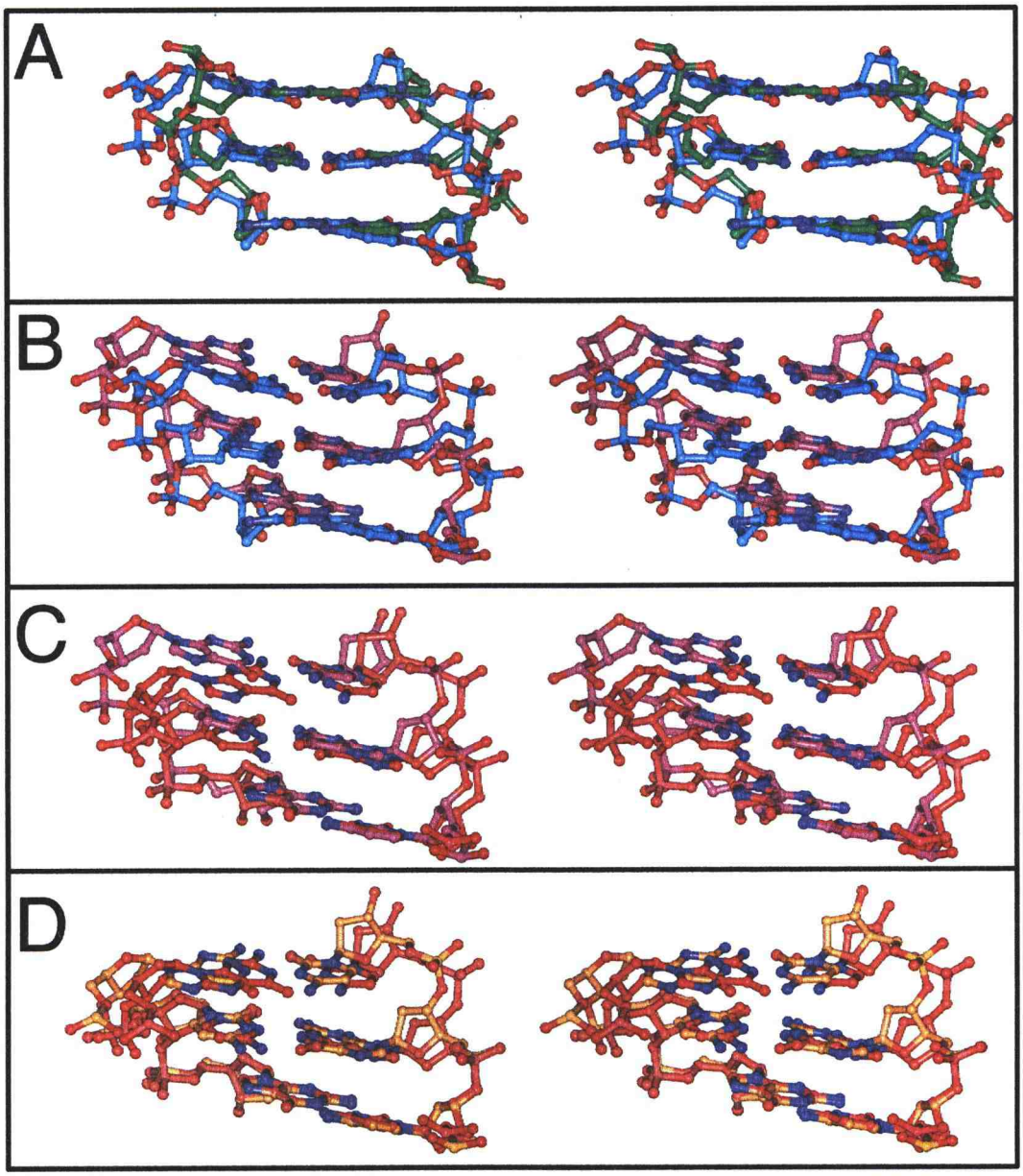

Figure 3.4

the canonical A-DNA structure is nearly continuous, passing through the A/B-type intermediates along the way (Figure 3.4D).

The detailed pathway is explicitly defined here for the B-A transition induced by methylation or bromination, but the common occurrences of a B-type conformation with A-like features and of the A/B-intermediates in the current study and in previous unrelated sequences (Ng *et al.* 2000; Trantírek *et al.* 2000) suggest that this pathway may generally be applicable to B-A transitions in other systems. For example, the A-DNA region of the cisplatin-bound A-B junction is located at the 5'-end of the cross-linked DNA strand (Takahara *et al.* 1995), suggesting that A-type features propagate backwards from the point where the B-DNA helix is perturbed. Comparing $d(GGm^5CGCC)_2$, which has A-type nucleotides at the 5'-end of each strand, with $d(GGCGm^5CC)_2$, which has entirely A-type nucleotides, suggests that the A-conformation is initiated at the point of methylation, and is propagated back towards the 5'-end of the strands. Finally, we see that the DNA conformation induced by TATA-binding protein requires only that the pathway be extended towards A-type structures with even more exaggerated base pair inclinations (from ~20° for standard A-DNA to ~40° for the promoter induced structure (Guzikevich-Guerstein and Shakked 1996)).

3.6 Acknowledgments

We thank P. A. Karplus and Z. A. Woods for a critical reading of this manuscript, and Dr. X.-J. Lu and W. K. Olson at Rutgers University for discussion on the X3DNA analysis program. This work was supported by the National Science Foundation (MCB-0090615) and the National Institutes of Environmental Heath Sciences (ES00210). X-ray facilities were funded in part by the M. J. Murdock Charitable Trust and the Proteins and Nucleic Acids Research Core of the Environmental Health Sciences Center at Oregon State University. Coordinates have been deposited in the Protein Database under accession numbers 1IH6 for $d(GGBr^5CGCC)_2$, 1IH2 for $d(GGBr^5CGBr^5CC)_2$, 1IH4 for $d(GGm^5CGCC)_2$, 1IH3 for $d(GGm^5CGm^5CC)_2$, and 1IH1 for the cobalt structure of $d(GGCGCC)_2$ as B-DNA.

Chapter 4

The Effect of Cytosine Methylation on the Structure and Geometry of the Holliday Junction: The Structure of d(CCGGTACm^{5}CGG) at 1.5 Å Resolution

Jeffrey M. Vargason and P. Shing Ho

Formatted for Submission

4.1 Summary

The single crystal structure of the methylated sequence d(CCGGTACm^{5}CGG) has been solved as an antiparallel stacked-X Holliday junction to 1.5 Å resolution. When compared to the parent nonmethylated d(CCGGTACCGG) structure, each of the duplexes are shifted by 3.4 Å and rolled open by 10.8° relative to each other, rendering the major grooves overall more accessible. A Ca^{2+} complex is seen in the minor groove opposite the junction, but is related to the B-conformation of the stacked arms. At the junction itself, the hydrogen bond from the N4 nitrogen of cytosine C8 to the C7 phosphate at the crossover in the parent structure has been replaced by a water bridge. Thus, this direct interaction is not absolutely required to stabilize the junction at the previously defined ACC-trinucleotide core. The more compact methylated junction forces the Na^+ out of the protected central cavity of the nonmethylated junction and into a solvent cluster that spans the space between the junction crossover and the stacked arms. A series of void volumes within the methylated and the nonmethylated structures suggests that small monovalent cations can fill and vacate this central cavity without the need to completely unfold the four-stranded Holliday junction.

4.2 Introduction

A four-stranded DNA complex was first proposed by Holliday as the key intermediate to effect strand exchange during homologous recombination (Holliday 1964). Consequently, the structure of the Holliday junction is recognized as an important component in a number of cellular processes, including the exchange of genetic material between homologous regions in chromosomes, and during repair of or replication through DNA lesions (reviewed in (Smith 2001) and (Cox 2001)). Furthermore, the product of the human breast cancer associated BRCA2 gene (Xia *et al.* 2001) has been shown to promote homologous recombination, while the BLM gene product associated with Bloom's syndrome displays junction binding and corresponding antirecombinase activity (Karow *et al.* 2000). In recent years, the detailed single crystal structures of junctions have been determined by X-ray crystallography in a series of DNA constructs (reviewed in (Ho and Eichman 2001)). A comparison of these structures shows a certain degree of conformational flexibility and sequence specificity in their formation. We report here a high-resolution single crystal structure of the Holliday junction in a sequence that incorporates a methylated cytosine. The structure shows that both the interactions at the core and the geometry relating the DNA arms that extend from the junction are variable, suggesting flexibility in the types of sequences that can be accommodated by the junction, and flexibility in the detailed conformation of the junction itself.

The structure of Holliday junctions has been the subject of biochemical studies since the 1960s (recently reviewed in (Lilley 2000)). In solution, the conformation is known to adopt an extended open-X form under low salt conditions, but will collapse to the more compact stacked-X form in the presence of both monovalent and multivalent cations (Figure 4.1a). Thus, interaction with ions is important in defining the conformation of these four-stranded complexes. The intrinsic structure of DNA junctions have only recently been seen in detail in the single crystal structures of two DNA/RNA complexes (Nowakowski *et al.* 1999; Nowakowski *et al.* 2000), of the mismatched sequence d(CCGGGACCGG) (Ortiz-Lombardía *et al.* 1999), of the symmetric sequence d(CCGGTACCGG) (Eichman *et al.* 2000), and of the psoralen adducts with the sequences d(CCGGTACCGG) and d(CCGCTAGCGG) (Eichman *et al.* 2001). All of these structures adopt the stacked-X form, but comparisons between the different structures clearly show that there is a large degree of conformational variability in four-way junctions. The DNA/RNA complexes show that the orientation between the stacked arms across the junction (the interduplex angle, Figure 4.1b) can be either right-handed, as seen in solution, or left-handed. This high degree of flexibility may be associated with one arm being extremely short (one base pair) in these constructs, although an atomic force microscopy (AFM) study suggests that this angle may be dependent on the local electrostatic interactions between the arms (Sha *et al.* 2000). This interduplex angle was seen to be affected by drug binding, with a 41° interduplex angle seen in the d(CCGGTACCGG) structure, but a more shallow 38° angle in the

Figure 4.1. Structure of Holliday junctions. a. The open-X (left) and stacked-X (right) forms of DNA four-way junctions. Arrows point from the 5' to the 3' ends of each strand. Continuous duplex regions are shown as cylinders. The stacked-X structure is shown in its antiparallel configuration. b. The interduplex angle. The two duplexes that are linked by the crossover strands of the junction can be related in a true antiparallel fashion (center panel). However, the current crystal structures of DNA junctions show that the duplexes are related either in a left-handed (a negative interduplex angle) or right-handed (positive interduplex angle) sense. c. Translations along the helix axes of adjacent duplexes. The duplexes can slide relative to each other along their helix axes to change the relative positions of base pairs across the junction. We have defined here a positive slide for the antiparallel stacked-X junction as a translation in the 5'- to 3'-direction of the outside strands. d. Rotations of adjacent duplexes about the helix axes. The stacked arms can roll along their helix axes to make the major groove either more or less exposed.

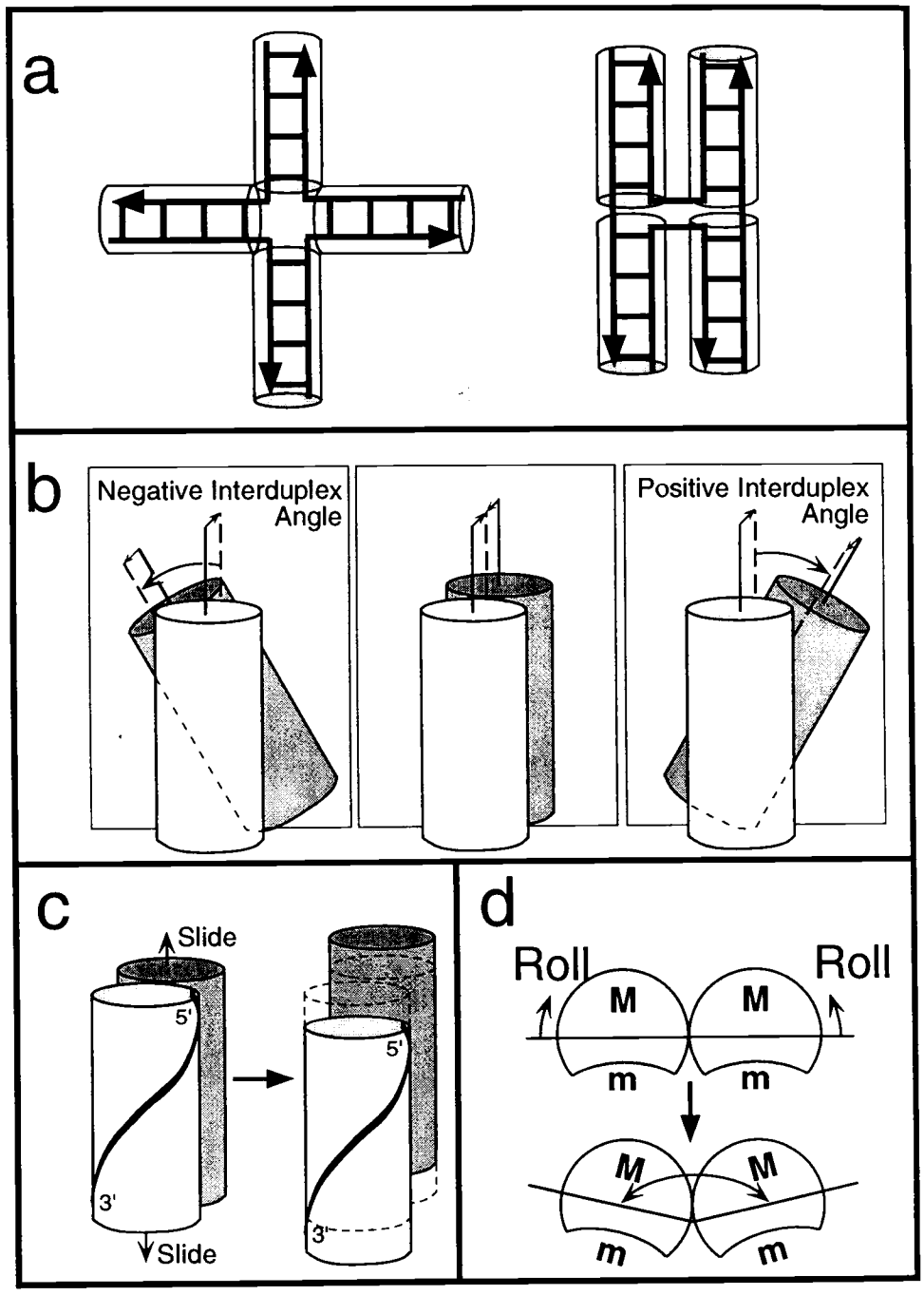
a
b
Negative Interduplex Angle
Positive Interduplex Angle
c
Slide
5'
3'
Slide
5'
3'
d
Roll
Roll
M
M
m
m
M
M
m
m

Figure 4.1

two psoralen adducts of the related sequences. All of these angles are much shallower than the ~60° estimated in solution from anomalous gel migration (Cooper and Hagerman 1987) and flourescence resonance energy transfer studies (Clegg *et al.* 1992), and by AFM (Mao *et al.* 1999). Thus, the Holliday junction shows some degree of conformational flexibility, at least in this angle relating the orientation of the stacked arms.

A comparison of the two junction forming sequences d(CCGGGACCGG) and d(CCGGTACCGG) with similar sequences that have been crystallized as standard B-DNA duplexes identified a common A6pC7pC8 trinucleotide core for the Holliday junction in this crystal system (Eichman *et al.* 2000). These crystal structures suggest that a number of interactions are associated with this core sequence, including a direct hydrogen bonding interaction between the N4 amino group of cytosine C8 and the phosphate of C7 on the major groove side of the junction cross-over, a solvent mediated hydrogen bond between the keto oxygen of guanine G3 (which complements C8) and the phosphate oxygen of A6 at the opposing minor groove side, and a sodium ion that sits within the junction, coordinated to the two phosphate oxygens at A6. It was clear that these interactions, and thus the significance of the ACC trinucleotide core in fixing the junction, are not affected by perturbations from mismatched base pairs and by covalent cross-links to the drug psoralen. Outside of the crossover strands, the conformations of the stacked arms reflect primarily the sequence dependent behavior of the B-DNA double helix (Eichman *et al.* 200x).

The current studies were designed to test the relevance of the hydrogen bonding interactions at the ACC-core in the formation of Holliday junctions. The prediction was that by modifying cytosine C8 with a bulky methyl group, as in the sequence d(CCGGTACm^{5}CGG) (where m^5 represents a methyl modification to the C5-carbon of cytosine at the eighth nucleotide position), the direct hydrogen bond from C8 to the phosphate at the cross-over would be disrupted and, consequently, the junction would be disrupted. Ironically, this sequence has provided us with the highest resolution structure of a Holliday junction to date. The structure shows how the methyl modification can be accommodated by distortions to both the detailed and gross conformational features of the junction, including the targeted hydrogen bonding interactions, rotational and translational relationships between the stacked arms, and the interactions of monovalent and divalent cations at the junctions and the stacked arms.

4.3 Materials and Methods

DNA sequences of d(CCGGTACm^{5}CGG) were synthesized on an Applied Biosystems DNA synthesizer with the trityl-protecting group left intact to facilitate purification by reverse-phase HPLC. The trityl group was removed with 3% acetic acid, and the DNA desalted on a Sephadex G-25 (Sigma-Aldrich) column, lyophilized, resuspended in 15 mM sodium cacodylate buffer (pH 7.0), and used

for crystallization without further purification. Crystals were grown at room temperature by vapor diffusion in sitting drop setups with initial solutions containing 0.5 mM DNA, 25 mM sodium cacodylate buffer (pH 7), 10mM $CaCl_2$, and 2 mM spermine•4HCl and equilibrated against a reservoir of 20% (v/v) 2-methyl-2,4-pentanediol (MPD). X-ray diffraction data were collected at liquid nitrogen temperatures using 1.15 Å wavelength radiation from beamline 5.0.2 at the Advanced Light Source in Berkeley, CA (Table 4.1). The dataset was reduced by using the programs DENZO and SCALEPACK from the HKL package (Otwinowski and Minor 1997).

Crystals of d(CCGGTACm^{5}CGG) grew as thin diamonds (a crystal measured at 0.25x0.25x0.05mm^3 was used for data collection) and belong to the space group *C2* with unit cell dimensions a = 65.5 Å, b = 24.7 Å, c = 37.0 Å, and β = 110.0°. These dimensions are similar to the nonmethylated sequence d(CCGGTACCGG) (a = 66.5 Å, b = 23.5 Å, c = 76.9 Å, β = 114.8°), except that the c-axis in the methylated structure is now nearly half of that of the parent sequence. The asymmetric unit of the methylated sequence is thus half that of d(CCGGTACCGG), consisting of two rather than four strands of DNA. However, the methylated and nonmethylated structures are not isomorphous, even when this difference in the asymmetric unit is taken into account. The structure of the methylated structure was solved by molecular replacement by applying a directed real space translation/rotation/rigid body search (X-PLOR 3.851 (Brünger 1992) script written in this lab) with two separate pentamer B-DNA double helices

Table 4.1. Data collection and refinement statistics

Sequence:	CCGGTACm^{5}CGG
Data Collection	
Space Group	*C2*
Unit Cell Dimensions	a = 65.5 Å, b = 24.7 Å, c = 37.0 Å, β = 110.0°
Resolution (Å)	20.0 - 1.5
Total reflections (unique)	18200 (7988)
Completeness (%)[1]	87.3 (39.3)
R_{meas}(%)[1,2]	4.8 (43.5)
Refinement	
R_{cryst} (R_{free})(%)[3]	21.8 (26.4)
DNA (Solvent) atoms	406 (96)
RMSD Bond lengths (Å)	0.012
RMSD Bond angles (°)	2.08

[1]Values in parentheses refer to the highest resolution shell

[2]$R_{meas} = \Sigma_{hkl} \sqrt{(n / n-1)} \Sigma_i | \langle I \rangle_{hkl} - I_{hkl,i} | / \Sigma_{hkl} \Sigma_i | I_{hkl,i} |$ where I_{hkl} is the intensity of a reflection, $\langle I \rangle_{hkl}$ is the average of all observations of this reflection and its symmetry equivalents, and n is the multiplicity (Diederichs and Karplus 1997).

[3]$R_{cryst} = \Sigma_{hkl} | F_{obs} - kF_{calc} | / \Sigma_{hkl} |F_{obs}|$. $R_{free} = R_{cryst}$ for 10% of the reflections that were not used in refinement (Brünger 1992).

serving as search models. The resulting model consisted of two DNA strands (one cross-over strand and one noncross-over strand), with the complete four-stranded junction generated by a crystallographic two-fold axis.

The methylated Holliday junction was initially refined in CNS (Brünger *et al.* 1998) using rigid body and simulated annealing procedures followed by conventional positional and individual B-factor refinement. A total of 95 waters and one calcium ion were modeled into the electron density, bringing the R_{cryst} and R_{free} down to 24.7% and 27.6%, respectively. The structure was completed using Refmac5 (Murshudov *et al.* 1999), with TLS refinement (Winn *et al.* 2001) defining sugar-phosphates and bases as individual groups, resulting in the final R_{cryst} = 21.8% and R_{free} = 26.4%. The coordinates and structure factors have been deposited in the Protein Data Bank (Berman *et al.* 1992).

4.4 Results

4.4.1 Structure of d(CCGGTACm^{5}CGG) as a Holliday Junction

This study was designed to test the hypothesis that the direct hydrogen bond between the exocyclic amine on the cytosine 8 (C8) nucleotide and the non-esterified oxygen on the phosphate of the cytosine 7 (C7) nucleotide is required for forming the Holliday junction in the symmetric d(CCGGTACCGG) sequence (Eichman *et al.* 2000). Methylating the cytosine base at C8 was predicted to

interfere with this direct hydrogen bond and, consequently, with the formation of the junction. We have solved the crystal structure of the sequence d(CCGGTACm^{5}CGG) and, indeed, the hydrogen bond was seen to be disrupted. However, the four-stranded junction conformation persists in the structure of the methylated sequence (Figure 4.2a,b), but this added substituent group has perturbed the geometry and solvent interactions.

The structure of d(CCGGTACm^{5}CGG) was solved to 1.5 Å by molecular replacement using B-DNA duplexes as the search model because crystals of methylated and nonmethylated d(CCGGTACCGG) sequences were not isomorphous. Electron density maps were discontinuous between the adenine 6 (A6) and C7 of the starting B-DNA duplex model. The densities at these positions were seen to connect the phosphodeoxyribose backbones of adjacent duplexes, consistent with the formation of a four-way junction (Figure 4.2c). The final refined structure consists of two unique strands of DNA that sit on a crystallographic two-fold symmetry axis to generate the complete four-stranded Holliday junction (Figure 4.2b). This represents the highest resolution structure to date of a Holliday junction. As with the structures of d(CCGGGACCGG) (Ortiz-Lombardía *et al.* 1999) and d(CCGGTACCGG) (Eichman *et al.* 2000), the methylated d(CCGGTACm^{5}CGG) junction is in the antiparallel right-handed stacked-X form. The strand crossovers of the junction occur between the A6 and C7 nucleotides, with each 4-basepair arm stacked on a 6-basepair arm to form pseudocontinuous duplexes on either side of the junction (Figure 4.2b). The angle

Figure 4.2. The Holliday junction in the crystal structure of d(CCGGTACm^{5}CGG). a. The antiparallel four-stranded junction is formed by the association of two unique strands (blue and green) and two strands (red and orange) generated by the two-fold symmetry of the crystal space group. b. Atomic structure of the junction. The atoms are shown as stick models, and the phosphodeoxyribose backbone traced with a solid ribbon. The ribbons and atoms are colored according to the strands in (a). In the stacked-X junction, the outside strands (blue and red) wraps around the helix axis in the manner as a standard B-DNA strand, while the cross-over strands (green and orange) leave their respective duplex, make jinked turns, and pair to the end of the adjacent duplexes (figure created and rendered in InsightII (MSI Biosym, Inc.)). c. Electron density at the junction crossover. A 2Fo-Fc map was drawn with 1 σ contours to show the break in the backbone between nucleotides A6 and C7, and the migration of the density to the adjacent duplex to form the cross-over of the junction (figure created with Bobscript (Esnouf 1999)).

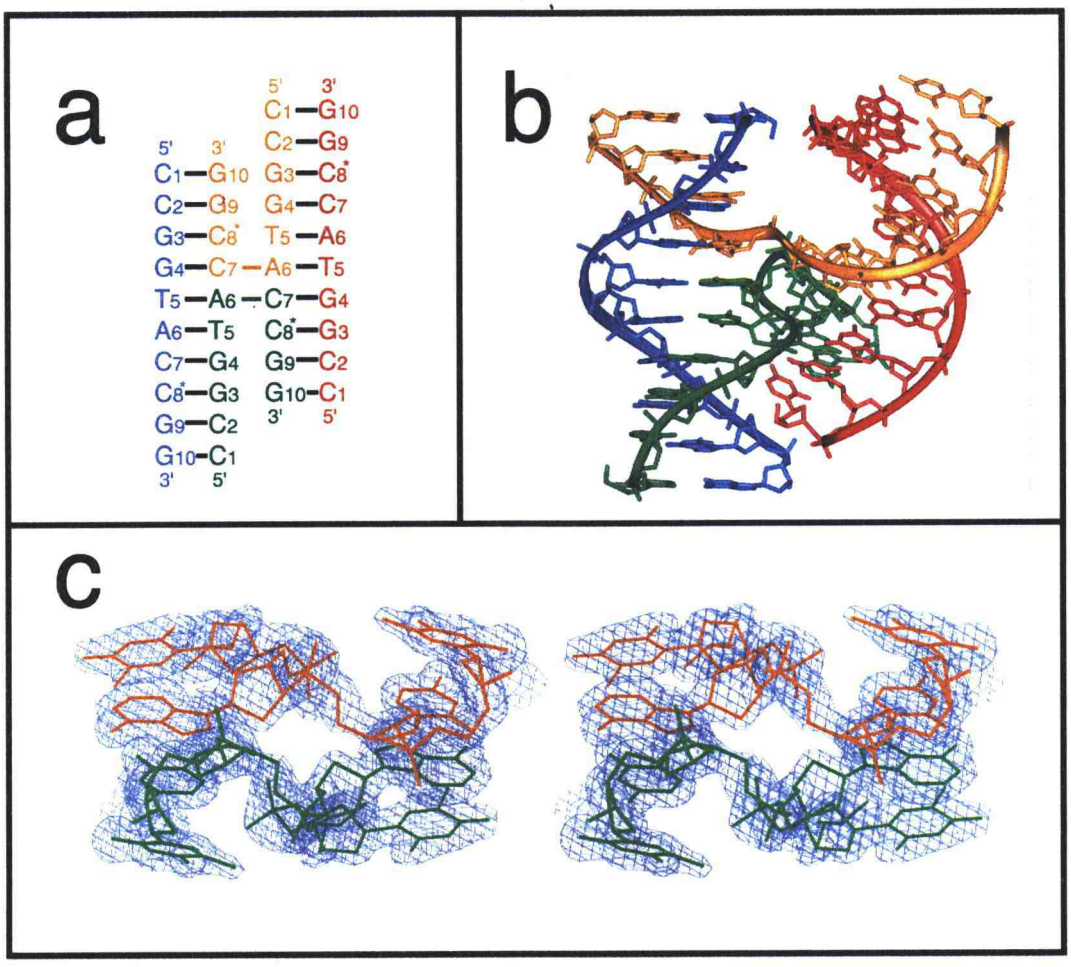
a
5' 3'
C1—G10
C2—G9
G3—C8*
G4—C7
T5—A6
A6—T5
C7—G4
C8*—G3
G9—C2
G10—C1
3' 5'
5' 3'
C1—G10
C2—G9
G3—C8*
G4—C7
T5—A6
A6—T5
C7—G4
C8*—G3
G9—C2
G10—C1
3' 5'
b
c

Figure 4.2

between these stacked duplexes (the interduplex angle, Figure 4.1b) is 41.3°, similar to the structures in both the d(CCGGTACCGG) and d(CCGGGACCGG) sequences. The Holliday junction structure in the current methylated sequence, however, shows perturbations in local conformational features as well as the solvent interactions at the crossover which propagate to affect other global features.

4.4.2 Perturbations to the Local Conformation and Solvent Interactions at the Junction Crossover

The detailed conformation at the junction cross-over was seen to be perturbed when comparing the structures of the d(CCGGTACm^{5}CGG) and parent nonmethylated d(CCGGTACCGG) junctions. Two sets of molecular interactions involving the phosphates of the junction cross-overs were identified in the nonmethylated d(CCGGTACCGG) and d(CCGGGACCGG) structures as being important for defining the structural ACC-core of the junctions. These include i) the direct hydrogen bond between the exocyclic amine N4 of cytosine C8 and the non-esterified oxygen O2P of C7, and ii) a water (W1) mediated hydrogen bonding interaction between the keto O6 oxygen of the complementary G3 nucleotide on the outside strand and the non-esterified oxygen O1P of A6 (Figure 4.3a). The methyl group in the current structure sterically disrupts the direct N4 to O2P hydrogen bond by pushing the phosphate of cytosine C7 away from the stacked arms. The loss of this direct hydrogen bond is compensated by the insertion of an intervening water molecule (S2) which acts to bridge the now widened gap

Figure 4.3. Base pairs flanking the junction viewed down the helix axes of the stacked arms. The duplexes joined by the junction cross-overs in the nonmethylated (a) and methylated (b) sequence in the antiparallel stacked-X are rolled towards each other so that their major grooves (labeled M) are essentially pointing in the same direction relative to the junction. a. View of the nonmethylated junction showing the duplexes rolled towards the major grooves. The relative rotation between the duplexes is shown to be acute in this parent structure. This can be seen by the angle between the vectors that extend from the center of the junction to the two opposite outside strands (red lines). The water molecule observed at the junction in the structure (Eichman *et al.* 2000) is shown as a red colored sphere and labeled W1, while the sodium ion is shown as a violet sphere and labeled Na^+. Dashed lines indicate interactions between the solvent and the DNA atoms. b. The methylated structure shows less roll between duplexes, resulting in a more shallow angle relating the rotation of the duplexes, as seen again by the vectors that extend from the junction center (red lines are from the nonmethylated structure in (a), while the blue lines represent the vectors of the current methylated structure). Solvent molecules are shown as red (water) and violet (sodium) spheres interacting with the DNA atoms (dashed lines). The molecules are labeled S1 to S4 (figure was created and rendered with InsightII (MSI Biosym, Inc.)).

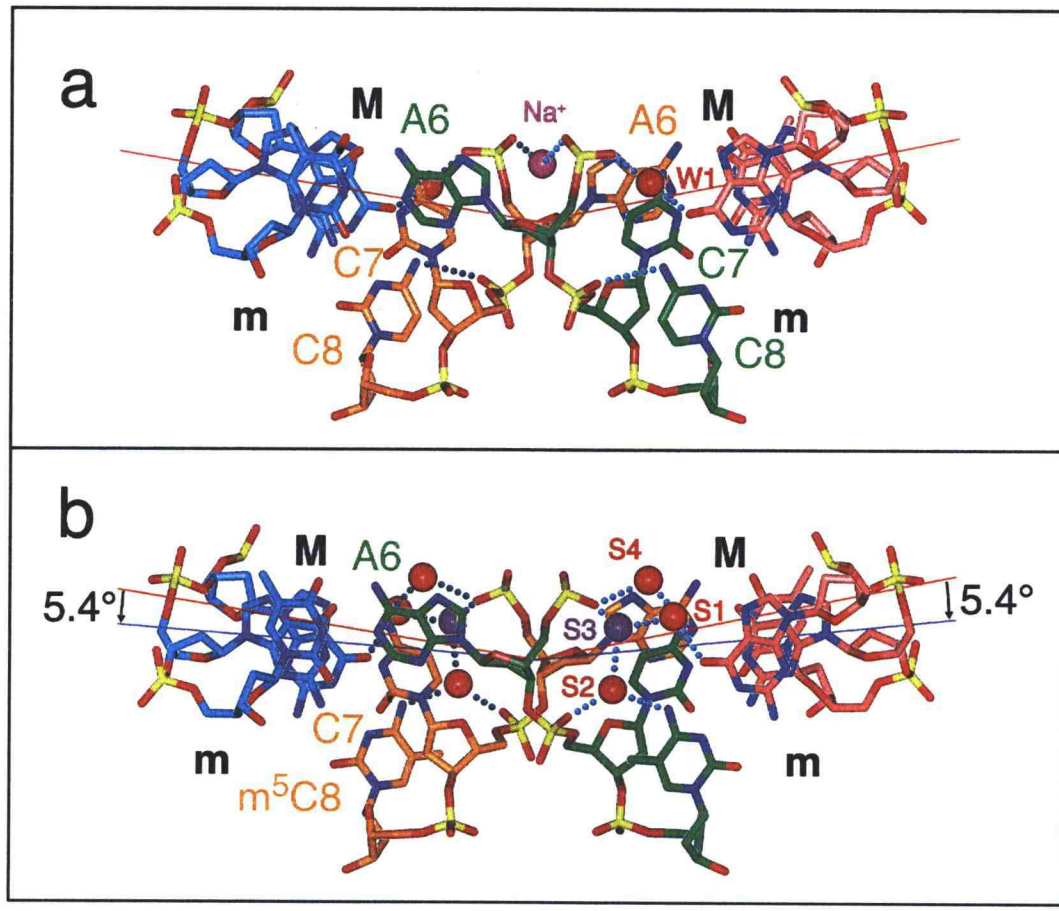

Figure 4.3

between the phosphate of the junction and the base of methylcytosine m^5C8 (Figure 4.3b, Figure 4.4a). This indicates that the direct hydrogen bond from the nucleotide base to the backbone phosphate in itself is not required for the formation of four-way junctions in this crystal system.

The steric interaction of the methyl group with the phosphates at the crossover also increases the phosphate-phosphate distance between C7 and C8 from an average of 6.5 Å in the nonmethylated structure to 6.8 Å in the methylated junction. This causes a shift in the backbone that propagates to affect the conformation of the phosphate of A6 that sits at the opposite face of the junction crossover (Table 4.2). Together, the two symmetry related A6 phosphodeoxyribose groups form a tight cavity that is shielded from solvent by the O2P oxygen. In the non-methylated structure, a sodium ion is coordinated to these A6 phosphate oxygens within the cavity at the center of the crossover (Figure 4.3a). The shift in the position of the A6 phosphate in the methylated structure distorts the geometry and compresses the size of the cavity. No electron density corresponding to the sodium ion is found within the cavity of the methylated junction (Figure 4.2c), indicating that this sodium ion has been dislodged. The volume of the cavities, calculated using the program VOIDOO (Kleywegt and Jones 1994) with standard van der Waals radii for atoms, corresponds to a sphere of radius 1.1 Å in the parent non-methylated junction, but only 0.9 Å in the methylated junction. Thus, a sodium ion, with an ionic radius of 1.13 Å (Huheey 1978), just fits into the space

Figure 4.4. Electron densities of the cations in the junction structure of d(CCGGTACm^{5}CGG). Electron densities of the solvent bridging the junction cross-over to the stacked arms (a) and of the hexaaquo-calcium complex in the minor groove of the stacked arms (b). Electron densities at 1 σ (blue) and 2 σ (orange) contours are shown from the 2Fo-Fc calculated electron density maps calculated for the solvent spaces in the two regions of the structure. a. Solvent at the A6pC7pm^{5}C8 trinucleotide. Dashed lines show the shortest interaction distances to other solvent and to the DNA atoms (distances labeled in Å). b. Calcium complex in the minor groove at base pairs C2·G9 and G3·C8 of the four base pair stacked arm. The complex modeled into the electron density is shown as a ball-and-stick model (figure was created with Bobscript (Esnouf 1999) and rendered with Raster3D (Merritt and Bacon 1997)).

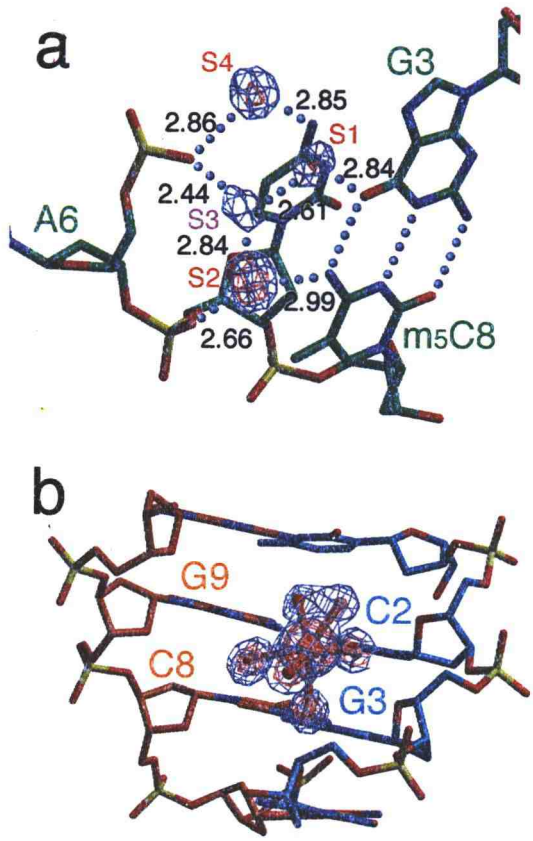
a
S4
G3
2.85
2.86
S1
2.84
2.44
S3
2.61
A6
2.84
S2
2.99
2.66
m5C8
b
G9
C2
C8
G3

Figure 4.4

Table 4.2. Comparison of the dihedral angles at the ACC-core of the junctions in d(CCGGTACm^{5}CGG) and the parent structure d(CCGGTACCGG).

Nucleotide	α	β	γ	δ	ε	ζ	χ
Methylated							
A6	-59.8°	-176.1°	43.5°	142.1°	-80.2°	-78.3°	-84.1°
C7	-59.2°	-171.9°	65.4°	133.1°	-125.3°	-75.2°	-154.6°
m^{5}C8	-80.5°	163.4°	50.5°	142.0°	-102.2°	167.9°	-70.0°
Nonmethylated							
A6	-30.7°	152.3°	34.0°	141.7°	-89.0°	-77.8°	-84.8°
	(4.0)	(3.2)	(1.2)	(1.1)	(0.6)	(0.8)	(2.5)
C7	-75.1°	-157.9°	57.8°	149.8°	-115.2°	-97.9°	-147.4°
	(5.02)	(2.1)	(3.8)	(2.4)	(0.1)	(0.0)	(0.7)
C8	-64.7°	145.6°	45.1°	138.7°	-83.9°	167°	-82.3°
	(2.2)	(6.5)	(2.3)	(3.5)	(1.9)	(2.9)	(1.1)

Values for the nucleotides A6, C7, and C8 are listed only for the strands that crossover to form the junction in both structures. The two unique crossover strands are averaged for the nonmethylated d(CCGGTACCGG) structure (values in parentheses indicate the difference of each value from the mean).

Definitions for the dihedral angles along the phosphodeoxyribose backbone are α O3'$_{i-1}$-P-O5'-C5', β P-O5'-C5'-C4', γ O5'-C5'-C4'-C3', δ C5'-C4'-C3'-O3', ε C4'-C3'-O3'-P$_{i+1}$, and ζ C3'-O3'-P$_{i+1}$-O5$_{I+1}$. The χ-angle relates the base to the deoxyribose about the glycosidic bond, and is defined as the central angle for O4'-C1'-N1-C2 for pyrimidine, and O4'-C1'-N9-C4 for purine nucleotides.

sandwiched between the A6 phosphates in the nonmethylated structure, but would be excluded by the compressed cavity of the methylated junction (Figure 4.5).

The other interaction that was proposed as being important for forming the junction in the nonmethylated d(CCGGTACCGG) sequence (Eichman *et al.* 2000) was a water molecule (W1) that bridged the O1P of A6 (which points away from the central cavity of the junction) and the O6 of the guanine G3 base on the outside strand (Figure 4.3a). In the current methylated junction structure, this water remains hydrogen bonded to the guanine base, but its distance to the phosphate oxygen is greatly extended. The larger distance between the phosphates at the junction and the stacked arms can now accommodate two additional solvent molecules (S3 and S4, Figure 4.3b, 4.4a). We have identified S3 as a sodium ion because of its very close approach to the O1P of A6 (<2.5 Å) (Figure 4.4a). The close proximity of four negatively charged phosphates at the junction crossover suggests the need for positive counterions to balance the electrostatic potential. With the sodium ion now missing from the center of the junction, one would expect to see some compensating counterion in the vicinity of these closely spaced phosphate groups. This putative sodium ion coordinates the O1P phosphate oxygen of A6, the water (S1) hydrogen bonded to the O6 oxygen of guanine G3 on the outside strand, and the water (S2) bridging the amino nitrogen N4 of C8 and the O2P oxygen of the phosphate at C7. Finally, an additional water molecule (S4) bridges the O1P oxygen of the A6 phosphate and the exocyclic amine N4 of C7

Figure 4.5. Surfaces and cavities in the junctions of the nonmethylated d(CCGGTACCGG) and the methylated d(CCGGTACm^{5}CGG). Void volumes at the junctions of the nonmethylated sequence d(CCGGTACCGG) (a) and its methylated analogue d(CCGGTACm^{5}CGG) (b). Accessible surfaces were calculated as Connolly surfaces (Connolly 1983), and displayed as dots. The calculations were performed for the DNA atoms only (with hydrogen atoms added by standard geometries), with all solvent molecules excluded from both models. The surfaces of the exposed DNA and those buried as void volumes were calculated using probe radii that vary from 0.8 to 1.0 Å, in 0.05 Å increments. We should note that the van der Waal's radii of atoms incorporated into the Connolly calculations (Connolly 1983) differ from those in VOIDOO (Kleywegt and Jones 1994), so the values from this figure do not exactly correspond to the radii of the minimum void volumes described in the text. However, the basic trends for the volumes calculated with different probe radii from the two programs are similar. The central cavity seen in the nonmethylated junction (a) is sufficient in size to fit a sodium ion, and is flanked by two unfilled volumes that are approximately 0.05 Å smaller in radius. In the methylated junction (b), two small solvent accessible pockets appear at probe radii 0.05 Å smaller than the sodium filled cavity in (a), and is lined by the O1P and O2P oxygens of the A6 phosphate of the junction crossover (figure was created and rendered with InsightII (MSI Biosym, Inc.)).

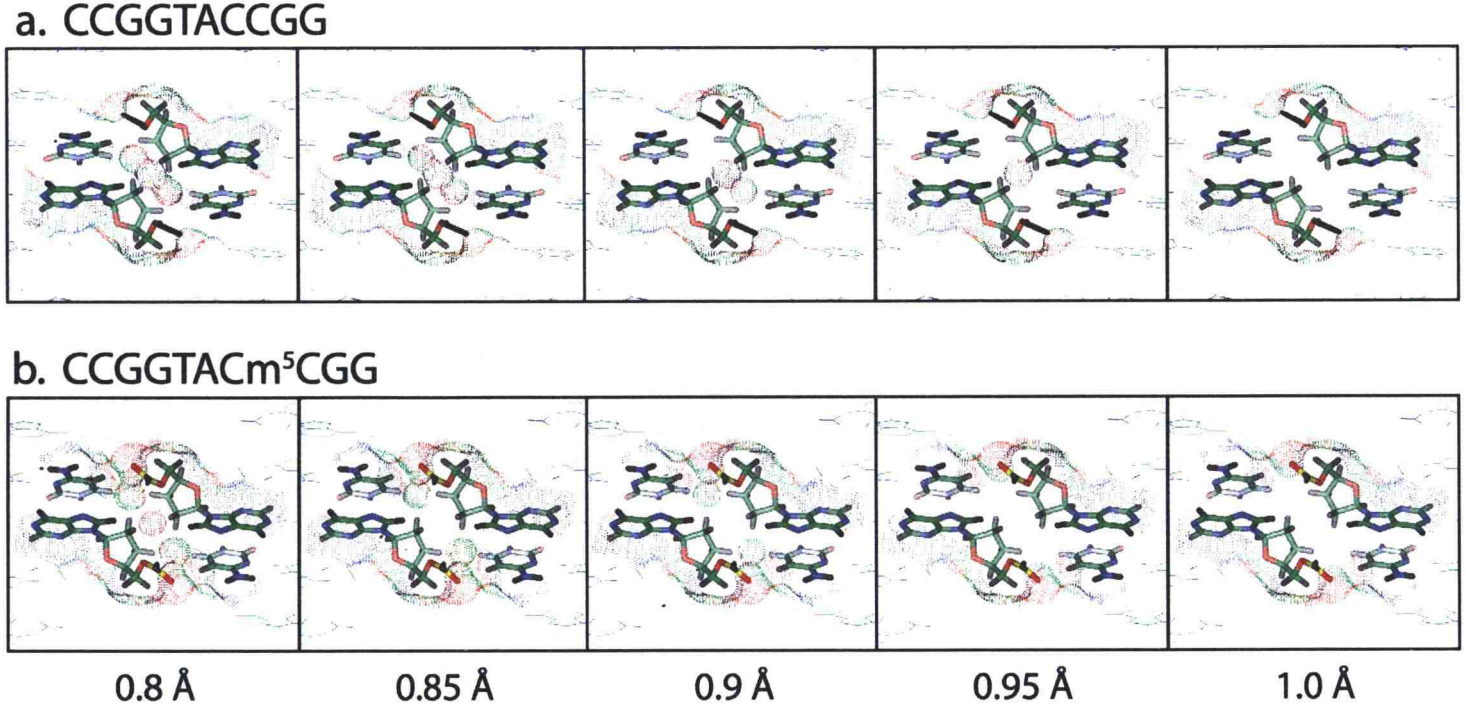

Figure 4.5

providing a complete first shell of solvent molecules that connects the junction to the base pairs of the stacked duplex arms.

More distant from the junction, we see differences in cation interactions at the DNA duplexes of the arms. In addition to the putative sodium counterion that helps to bridge the gap between the stacked arms and the junction cross-over, a calcium ion was placed in a set of strong electron density that lie within the minor groove of the 4 base pair arm between nucleotides C2 and G3 (Figure 4.4b). This ion shows six distinct electron density peaks within 2.1-2.5Å of the calcium center. These have been assigned as the six coordinating waters of a hexaaquo-calcium complex. We would expect to see additional ion interactions to compensate for the large negative potential at the crossover of the Holliday junction structures, but this is the first time a divalent ion has been observed in any of the DNA junction structures and, ironically, it doesn't directly interact with the junction, but sits in the groove on the opposite side of the duplex. Magnesium ions have been foot-printed to specific sites of asymmetric junctions formed in solution (Murchie *et al.* 1989); however, neither the d(CCGGGACCGG) junction grown with Mg^{2+} (Ortiz-Lombardía *et al.* 1999) nor d(CCGGTACCGG) grown with Ca^{2+} (Eichman *et al.* 2000) revealed any evidence for divalent cations. In contrast, a similar well ordered Ca^{2+} complex was found in the minor groove of the 0.74Å B-DNA structure of $d(CCAGTACTGG)_2$ (Kielkopf *et al.* 2000) at nearly the same location as in the current methylated junction structure. This suggests that the calcium complex may not be specifically involved in stabilizing the junction, but rather that

it is associated with the similarity in conformation or in the crystal environment of the DNA duplexes in the stacked arms of this methylated junction and the B-DNA structure.

4.4.3 Perturbations to the Global Geometry of the Junction

The local perturbations to the structure of the junction caused by the methyl group are propagated throughout the structure of the d(CCGGTACm^{5}CGG) junction. Although the conformation of the methylated junction remains a right-handed stacked-X conformation with an interduplex angle of 41.3°, similar to the structures of the previous all DNA junctions (Ortiz-Lombardía *et al.* 1999; Eichman *et al.* 2000), the relative translation along and rotation about the helix axes of the stacked duplex arms across the junction are perturbed (Figure 4.1c,d). These effects are apparent when the methylated and non-methylated junctions are superimposed using only the atoms of one of the outside strands as references. The stacked arms of the non-superimposed strands are shifted approximately 3.4 Å along the helix axis (Figure 4.6) and each duplex has been rotated about its helix axis by approximately 5° towards the minor groove of the adjacent stacked duplex across the junction. Consequently the methylated junction is opened by 10.8°, resulting in a greater exposure of the major groove surfaces relative to the original nonmethylated conformation (Eichman *et al.* 2000).

The shift of the duplex arms along the helix axis can be attributed to the steric contact between the methyl group at cytosine C8 and the phosphate of

Figure 4.6. Global perturbations associated with methylation of cytosine C8 to the translation and orientation of the duplexes joined by the Holliday junction. The outside strand of one duplex from the methylated structure is superimposed onto the analogous strand in the parent nonmethylated structure (left most strands). Cylinders represent the pseudocontinuous DNA double helices formed by stacking the arms of the Holliday junction. The stacked duplexes of the superimposed reference strands (yellow cylinders) are shown on the left-side of the junction in (a) and in the back in (b). a. View into the junction. The nonsuperimposed duplex of the methylated structure (blue cylinder) is seen to be translated by ~1 bp (3.4Å) along the helix axis that runs through the base pairs (in the 3'-direction of the outside strand) relative to that of the nonmethylated structure (red cylinder). Carbons of the methyl group are rendered as blue spheres. b. View along the junction. The slide along the helix axis is seen to be accompanied by a slight shift of the nonsuperimposed duplex towards the minor groove of this duplex. Figure 4.3b shows this apparent shift to actually be a result of the two duplexes rolling about their respective helix axes towards the minor groove (figure created and rendered with InsightII (MSI Biosym, Inc.)).

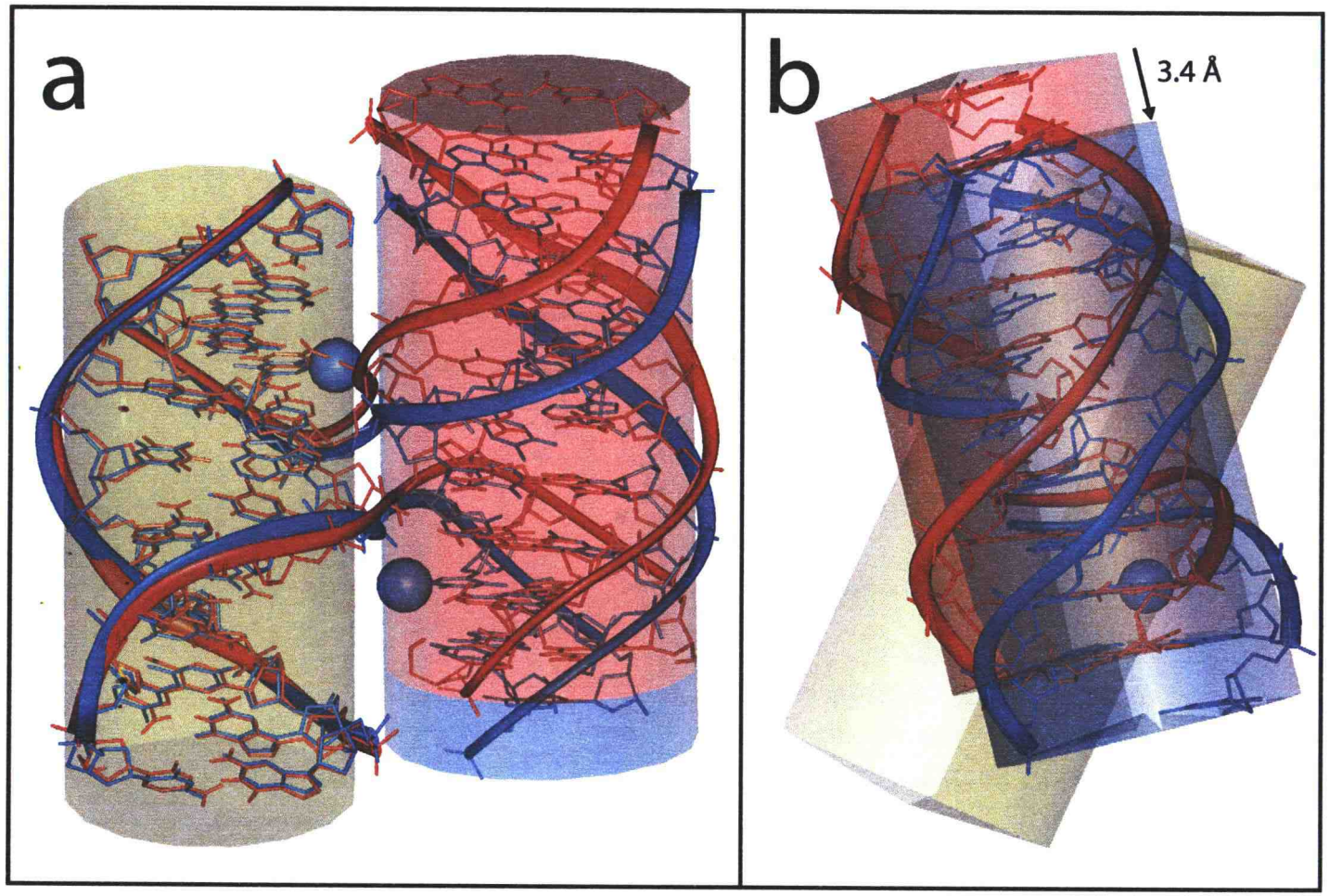

Figure 4.6

cytosine C7 at the junction crossover (methyl carbon to phosphate oxygen distance = 3.54 Å). Modeling this methyl group directly onto the nonmethylated structure results in analogous methyl carbon to phosphate oxygen distances of 2.4 Å. This steric interaction pushes the 4 basepair arm in a vertical direction, up and away from the junction. The stacked 6-baspair arm of this pseudocontinuous duplex follows to maintain the base stacking interactions. The inherent two-fold symmetry of this four-way junction requires the same steric contact to occur on the opposite side, resulting in a concerted shift of the two-stacked duplexes.

A less dramatic effect is seen on the relative rotations along the helix axes of the two sets of stacked duplex arms across the junction. This results in the slight shift of the duplex perpendicular to the helix axis in Figure 4.6b. In both methylated and nonmethylated structures, the duplexes are actually rolled towards the major groove so that the grooves are slightly facing each other (Figure 4.3a). In the methylated d(CCGGTACm^{5}CGG) structure, the rolling of the duplexes is less pronounced, resulting in a junction that is more open towards the major groove (Figure 4.3b). This perturbation in the geometry of the junction can again be attributed to the methyl group on cytosine C8 pushing against the phosphate of C7, but in the lateral direction. This perturbation, however, results in a concerted rotation rather than the apparent shift of the duplexes seen in Figure 4.6b. The angle at which the duplexes roll (Figure 4.1d) across the junction towards the major groove side of each duplex can be estimated by calculating the angle formed by the vectors extending from the center of the junction out to the two opposing outside

strands (Figure 4.3a). The center of the junction was defined by averaging the coordinates of the four phosphates at the junction crossover strands, while the vector ends were determined by averaging the two phosphate positions on both of the outside strands to give two vectors. The resulting roll angle for the nonmethylated junction was estimated to be 159.0°, while that of the methylated junction was 169.8° (Figure 4.3b). Thus the adjacent duplexes in the methylated junction have been rotated by 10.8° to increase the exposure of the major grooves at the crossover.

In general, the conformations of the base pairs along the stacked arms are nearly identical between the methylated and nonmethylated junctions, with the dihedral angles differing by less than 20° between the two structures for the nucleotides that are not in the ACC-core or the ends. An exception is the large discrepancy in the ζ dihedral angle at A6 on the non-crossover outside strand. This is likely a crystal packing effect rather than any perturbation that is intrinsic to the methylated junction. The backbone dihedrals at A6 in the nonmethylated junction fall into the unusual BII category ($\zeta \approx -96.8°$), while they fall in the BI category ($\zeta \approx 159.7$ to $166.1°$) in the methylated junction (Dickerson *et al.* 1987). The A6 nucleotide of the nonmethylated structure is in direct contact with an adjacent, symmetry related molecule. This crystal contact is lost in the current methylated structure because of the shifts and roll of the duplexes induced by the methylated C8 nucleotide. Thus, the nonstandard BII backbone conformation in d(CCGGTACCGG) can be attributed to perturbations associated with crystal lattice

contacts. Furthermore, the detailed structure of the stacked duplex arms reflects the effects of sequence on the structures of the analogous B-DNA duplexes, as we had previously seen (Eichman *et al.* 200x), even when the geometry of the junction is perturbed.

4.5 Discussion

The 1.5Å resolution single crystal structure of the complex formed from 4 strands of d(CCGGTACm^{5}CGG) demonstrates that cytosine methylation does not disrupt the formation, but perturbs the local and global conformation of the Holliday junction. This junction shares some similarities to the previous all DNA structures of d(CCGGGACCGG) (Ortiz-Lombardía *et al.* 1999)and d(CCGGTACCGG) (Eichman *et al.* 2000). All three form right-handed stacked-X junctions with interduplex angles of ~41°. The methylated junction is composed of all Watson-Crick basepairs like the parent nonmethylated junction in d(CCGGTACCGG), but the presence of the methyl group significantly affects the local structure and hydration at the junction crossovers, which are associated with a global shift along and roll around the helix axes of the stacked duplexes across the Holliday junction.

Simply, the observation that this methylated sequence forms a junction obviates the absolute necessity of a direct hydrogen bond between the base of

cytosine C8 and the phosphate of C7 in ACC-core containing junctions. This would now appear to lift the requirement that a cytosine occupy the third position of the ACC-core, which was thought to be important in the crystallization of d(CCGGGACCGG) and d(CCGGTACCGG) as junctions. In introducing a water molecule to mediate this hydrogen bonding interaction, we would expect that the specificity for the hydrogen bond donating N4 amino group of the cytosine would be lost, and could be readily replaced by a hydrogen bond accepting keto oxygen (*i.e.*, this cytosine should be replacable by a thymine base). However, the single crystal structure of the sequence d(CCAGTACTGG), which replaces cytosine C8 at the ACC-core with a thymine, has been solved to 0.74Å resolution as a standard B-DNA double-helix (Kielkopf *et al.* 2000). Thus, other interactions, perhaps unique base pair stacking interactions or other base pair specific effects (for example, the inherent large propeller twist of a TA base pair (El Hassan and Calladine 1996)), are important in defining the sequence specificity for fixing the position of the junction in these symmetric DNA sequences.

One of the unexpected results from the current structure is that the sodium ion at the center of the nonmethylated d(CCGGTACCGG) junction (Eichman *et al.* 2000) has been displaced from its protected cavity. The placement of this counterion, which would help to balance the negative potential of the very closely positioned phosphates at the junction crossover, is, as with the direct hydrogen bonding interaction, not absolutely required. Rather, this sodium ion is now apparently expelled to the gap between the junction and the stacked duplex arms,

and is now coordinated to the outer O1P oxygens of the same phosphates. This raises two interesting questions. Is the charge neutralization at the crossovers exactly the same in the methylated and nonmethylated junctions, and can a sodium ion enter and exit the four-stranded junction without unfolding the complex?

A comparison of the electron densities of the solvent molecules (S1 to S4) at the methylated junction indicates that they are all well ordered, but not equally occupied (Figure 4.4a). The density around S3 (whether refined as water or as a Na^+ ion) appears to be significantly weaker, and its associated temperature- or B-factor significantly higher (B-factor > 50 $Å^2$) than the surrounding atoms, including those for S1, S2, and S4 (average B-factor ≈ 30 $Å^2$). When we assume that the thermal disorder, and thus the B-factor of S3 should be similar to the other three solvent molecules in this cluster (which is reasonable because it is interacting with the other solvent molecules and the DNA), the occupancy of S3 can be estimated to be approximately half that of the associated solvent molecules. Furthermore, if our assignment of S3 as a sodium ion is correct, then the total number of monovalent cations at the methylated junction will effectively be identical to that of the nonmethylated junction, which is one sodium ion, either at the central cavity or shared between two sites. There does not appear to be any net change in the neutralization of the negative charges at the junction, with apparently only a single monovalent cation to counterbalance the four negatively charged phosphates. One would expect that sodium ions would be capable of fully occupying both sites available in the methylated junction, thereby reducing the net negative charge of

the phosphates at the junction, but we see only a single sodium essentially sharing time in one or the other of the potential sites. Thus, we can ask what is the effective charge of the phosphates at the crossover of the junction in the crystals? At the very most, the two phosphates at A6 would have effective charges of -0.5 to counterbalance the single sodium. The other -0.5 charge must be accounted for somehow in order to maintain an electrostatically neutral environment. This could come from other, more disordered counterions or, an interesting alternative, from a dramatic shift in the pK_a of the phosphates so that they are partially protonated under neutral conditions. Bringing four negatively charged phosphates in such very close proximity in a solvent depleted environment could provide sufficient electrostatic energy to perturb the pK_a of the nonesterified oxygens. The more likely explanation is that there are counterions that have not been located even at the high resolution of the current structure. However, each phosphate appears to have a full complement of solvent molecules at its first shell. Thus, additional counterions must either be currently assigned as water molecules, or must act from a distance (beyond the first hydration shell) and thereby suggest that direct cation interactions at the center or the junction are not essential to the formation of a stacked-X conformation.

The question of whether the sodium ion is trapped in the nonmethylated d(CCGGTACCGG) structure as the four-strands are being assembled cannot be directly addressed here, but a comparison of the void volumes within the methylated and nonmethylated structures (Figure 4.5) presents a possible pathway

by which a sodium can enter or leave the internal cavity without the need to completely disassemble the Holliday junction. When the junctions were probed with spheres of different sizes, the pocket where the sodium resides in the nonmethylated structure was seen to be flanked by two slightly smaller pockets (Figure 4.5a). These additional cavities are also present in the methylated structure (Figure 4.5b); however, the collapse of the junction associated with methylation pushes these cavities away from the center and towards the surface, making them accessible to the solvent in the space between the crossover strands and the stacked arms. Interestingly, the sodium ions assigned to the current methylated structure (Figure 4.4b) sit at the mouths of these invaginations. A mechanism therefore is suggested in which a sodium ion can enter or exit a fully folded junction through rearrangements between the collapsed methylated and expanded nonmethylated states, in conjunction with less drastic conformational breathing. Figure 4.7 shows how a cation that is initially coordinated to the O1P phosphate oxygen at the surface of the crossover strands can be captured and eventually coordinate at the central cavity. Starting with the junction in its collapsed state (that of the methylated junction, Figure 4.7a), a pocket can form with only minor conformational breathing (Figure 4.7b). The suggestion from the surface calculations shown in Figure 4.5 is that the surface can be invaginated with ~0.1 Å shifts in the positions of the atoms in the collapsed junction. Pinching off this pocket to making it now inaccessible requires a more dramatic conformational rearrangement from the collapsed to the expanded form of the nonmethylated

Figure 4.7. Pathway to encapsulate a sodium ion into the central cavity of the junction. The exposed and buried surfaces of the collapsed methylated (from Figure 4.5b) and the open nonmethylated junction (from Figure 4.5a) are traced with curved lines. The volumes filled by the DNA atoms are shown shaded, with the bonds of the nucleotide adenine A6 and the base of cytosine C7 shown as simple lines. The O1P and O2P oxygen atoms of the A6 phosphate are drawn as small open circles. The sodium ions are rendered as spheres, while the coordination of the cations to the oxygens of the A6 phosphates seen in the crystal structures of the methylated and nonmethylated junctions are shown as dashed lines. The open double-headed arrows indicate minor conformational breathing, while the filled arrow indicates the need for conformational rearrangement of the collapsed methylated junction to the open nonmethylated form. a. Surface of a collapsed junction with no ions at the central cavity (corresponding to Figure 4.5b, 0.8 Å probe), with cations coordinated to the O1P oxygen at the surface. b. Solvent accessible pockets form by invagination of the collapsed junction surface (from Figure 4.5b, 0.85 Å probe). c. The pockets in (b) are pinched off from the surface with the conformational rearrangement that forms a more expanded junction at the center (from Figure 4.5a, 0.85 Å probe). d. The sodium ion falls into the central cavity and is coordinated to the O2P oxygens (Figure 4.5a, 0.95 Å probe). The ion can be captured from either side of the junction, but only one can fit into the central cavity.

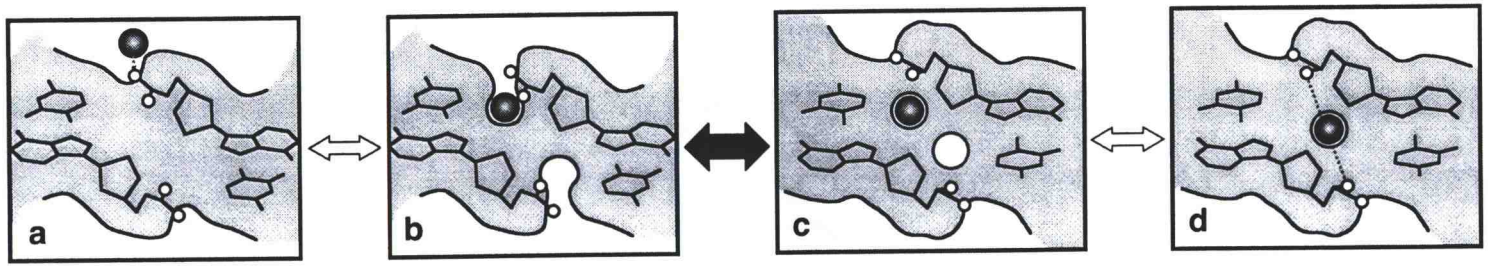

Figure 4.7

junction (Figure 4.7c). This larger conformational change, however, also opens the central cavity, allowing the ion to sandwich itself between and become coordinated to the two O2P oxygens of the phosphates that line the internal cavity (Figure 4.7d). In solution, any dynamic sliding or rolling between the linked duplexes could readily expand or collapse the junction to facilitate the movement of a sodium cation in or out of the central core of the structure. Thus, it is possible for a small cation to enter and leave the buried cavity within the junction without the need to completely disassemble the complex. Furthermore, it is also likely that, in solution, these motions are coordinated with the conformational dynamics of the junction, and a single ion may fluctuate between the central cavity and the solvent surrounding the junction.

Aside from the structural effects, does a methylated Holliday junction play a significant biological role? Whitehurst, *et al.*, have suggested that cytosine methylation is involved in regulating V(D)J recombination (Whitehurst *et al.* 2000). These effects could be attributed to the influence of CpG methylation on local chromatin accessibility (similar to the effect on transcriptional regulation), or on binding and recognition by the associated recombinase. The structural changes induced by the methyl group could affect how proteins may recognize the junction. The rotations of the duplexes increase exposure of the major grooves of the stacked arms, allowing base pairs in the area immediately adjacent to the junction to be more readily accessible and recognized. In contrast, the collapse of the junction makes the phosphates less accessible. The current structure shows that there is

conformational flexibility in the geometry of the junction in parameters besides twisting about the standard interduplex angle. This is apparently an area that is not well studied, and it would be interesting to determine whether the perturbations to the Holliday junction associated with cytosine methylation have any effect on cellular events that rely on homologous recombination.

4.6 Acknowledgments

We would like first to thank Dr. Z. A. Wood for help in collecting the diffraction data for this study. This work has been funded by grants from the National Science Foundation (MCB 0090615) and the Environmental Health Science Center at Oregon State University (NIEHS ES00210).

Chapter 5

Conclusion and Discussion

5-methylcytosine has been implicated in various roles such as transcriptional regulation (Antequera and Bird 1993), X-chromosome inactivation (Singer-Sam and Riggs 1993), and genomic imprinting (Sasaki *et al.* 1993) in eukaryotes and in distinguishing parent from replicated daughter DNA in prokaryotes (Noyer-Weidner and Trautner 1993). Somewhat antithetical to these biological roles, 5-methylcytosine is prone to a greater rate of spontaneous deamination than cytosine (Zhang and Mathews 1994). The biological relevance of this, the "fifth base" of DNA, has prompted us to ask, what is the impact of cytosine methylation on the structure of DNA. Chapters 2-4 explore some of the structures that can be induced by the presence of the methyl group on carbon 5 of cytosine and how these local structures can propagate a global change. From these structural studies, three general conclusions or points of discussion about methylation emerge.

The presence of cytosine methylation perturbs the local DNA structure and solvent interactions. Methylated cytosine in the sequence $d(GGCGm^5CC)_2$ causes an increase in the negative *x*-displacement and the negative slide of the basepairs. In addition, the basepairs are less inclined than either A-DNA or B-DNA. As a

result, the basepair planes are more exposed to solvent. The hydration around guanines in this structure are characterized by waters which both hydrogen bond to the N7 position and stack on the exposed face of the adjacent stacked base. Comparisons between the methylated and nonmethylated structures of d(GGCGCC) suggest that these perturbations are a result of the methylation. When the methylation is moved to another cytosine to give the new sequence $d(GGm^5CGCC)_2$, different distortions can be seen.

In the sequence $d(GGm^5CGCC)_2$, the 5-methylcytosine bases adopt an A-DNA conformation while the guanine bases that stack on the 3' side adopt a B-DNA conformation (because the sequence is self-complementary, these guanines are the basepair complements of the 5-methylcytosine bases). This change in conformation within basepair complements causes a large basepair buckle (~15°). The adjacent m^5C•G basepairs buckle in opposite directions yielding a large distortion at the center of the duplex. The hydration pattern around these basepairs is consequently affected. Waters hydrogen bonded to the N4 nitrogens of the two methyl cytosines are now close enough to hydrogen bond due to the large basepair buckle. This provides added stability at the location of the conformational change. The previous two examples of methylation related perturbations occur in hexamer length, duplex DNA, but the last example shows perturbations arising in longer duplexes.

The local structure of the methylated Holliday junction $d(CCGGTACm^5CGG)$ has been disrupted by cytosine methylation. Comparisons

between the non-methylated Holliday junction d(CCGGTACCGG) (Eichman *et al.* 2000) and the methylated variant show that the methyl group has disrupted an hypothesized important DNA-DNA hydrogen bond. The methyl group sterically disrupts the hydrogen bond between the exocyclic amine N4 nitrogen of cytosine 8 and the non-esterified O2P oxygen of cytosine 7 by pushing the phosphate away from the cytosine. The loss of the hydrogen bond is compensated by the addition of an intervening water molecule, thus changing the hydration in the environment of the methylation. In each of the methylated structures, the methyl on the cytosine causes a local structural perturbation resulting in a change in the hydration pattern.

The presence of cytosine methylation propagates a global conformation in the DNA. In the structure of $d(GGCGm^5CC)_2$, the global conformation of the sequence is no longer the B-DNA conformation of the unmethylated d(GGCGCC) sequence, rather it forms a conformation which shares features of both A-DNA and B-DNA. The basepairs of the duplex have relatively small inclinations like B-DNA, but have an extended rise. Unlike B-DNA, the basepairs have a large displacement similar to A-DNA. As a result of the local distortions caused by the methyl group, the double helices trace out an eccentric path rather than the circular path of B-DNA or A-DNA helices. This intermediate conformation of DNA has changed as a result of the methyl group on cytosine.

In the sequence $d(GGm^5CGCC)_2$, the local A-DNA conformation induced by methylation has propagated from the methylcytosine to the 5' end of the strand resulting in a duplex where the three 5' nucleotides (GGm^5C) are A-DNA and the

three 3' nucleotides (GCC) remain in the B-DNA conformation. Since the sequence is self-complementary, the A-DNA nucleotides are basepaired to B-DNA nucleotides. This particular structure has been predicted by theoretical calculations (Selsing *et al.* 1979) and has been observed in Raman spectroscopy (Nishimura *et al.* 1986). The conformation is a composite of both B-DNA and A-DNA and was induced by cytosine methylation. In this hexamer length duplex as well as in $d(GGCGm^5CC)_2$, the local distortion of the methylation has propagated a global conformational change.

The structure of the methylated junction $d(CCGGTACm^5CGG)$ displays global shifts and rolls of the stacked duplexes across the Holliday junction. The methyl group on cytosine C8 sterically contacts the phosphate of C7 perturbing the DNA backbone. This shift is propagated to the adenine A6 on the crossed over strand. The adenine A6 and its symmetry equivalent are pushed along the helix axis causing each duplex to shift 1.7 Å along their respective helix axes. They are also pushed in a direction perpendicular to the crystallographic two-fold and duplex axes causing a roll of each duplex by 5.4°. This results in a greater exposure of the major grooves at the junction crossover. The methylation in this structure has changed the global geometry of the Holliday junction.

The presence of the methylation and the resulting local and global perturbations to the DNA may provide an explanation for the implicated biological roles of 5-methylcytosine. First, the methylation-induced conformation may explain the elevated rate of deamination of 5-methylcytosine relative to cytosine in

the context of duplex DNA. 5-methylcytosine is shown to induce a novel conformation in DNA, which is characterized by waters that stack on the exposed faces of base pairs. In this position, the waters can act as a general nucleophile to initiate the deamination reaction. This suggests that the structure conveyed to duplex DNA as a result of the methylation increases the rate of deamination. Previous studies have suggested that the rate of deamination is strongly dependent upon DNA structure (Frederico *et al.* 1990). Experiments to test the rate of deamination in the environment of the crystal and therefore within the conformation of interest, are under way.

A methylation-induced conformation may alter a protein's affinity for that particular sequence of DNA. Little evidence exists that methyl groups modulate protein-DNA interactions solely through a direct protein-methyl group contact (Hodges-Garcia and Hagerman 1992). In addition, research on the EcoRI restriction endonuclease suggests that binding is partially regulated by an altered conformation of DNA induced by adenine N6-methylation (Diekmann and McLaughlin 1988). This precedence for a protein recognizing a unique conformation could suggest a role for methylation in transcriptional regulation. The demethylation of 5-methylcytosine has been coupled to transcriptional accessibility, and is thought to influence chromatin structure (Jones and Wolffe 1999). One hypothesis is that the methylation-induced conformation favors the nucleosome bound state. When the DNA is methylated it is preferentially bound to the nucleosome and is no longer accessible for transcription. Unfortunately, there

is no data to suggest that the nucleosome favors binding to methylated DNA or to a methylation induced conformation of DNA.

The structural changes conveyed to the Holliday junction upon methylation could provide evidence for a direct methylation effect on recombination. Methylation has been shown to influence the recombination event needed to assemble variable (V), diversity (D), and joining (J) gene segments associated with the variable exons of antigen receptor genes (V(D)J recombination) (Whitehurst *et al.* 2000). However, it is still unclear whether methylation regulates V(D)J recombination at a general DNA accessibility level analogous to transcriptional regulation or if there is a direct recognition of methylated junctions. Based on the increased exposure of the major groove and shift of the duplex arms in the methylated junction, it would be interesting to know whether the V(D)J recombinase protein complex prefers non-methylated relative to methylated junction conformations.

BIBLIOGRAPHY

Alden, C. J. and Kim, S. H. (1979). "Solvent-accessible surfaces of nucleic acids." *J. Mol. Biol.* **132**, 411-434.

Antequera, F. and Bird, A. (1993). CpG islands. DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz. Boston, Birkhäuser Verlag, 169-185.

Arnott, S. (1999). Polynucleotide secondary structures: an historical perspective. Oxford Handbook of Nucleic Acid Structure. S. Neidle. New York, Oxford University Press, 1-38.

Arnott, S., Chandrasekaran, R., Millane, R. P. and Park, H.-S. (1986). "DNA-RNA hybrid secondary structures." *J. Mol. Biol.* **188**, 631-640.

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

Behe, M. and Felsenfeld, G. (1981). "Effects of methylation on a synthetic polynucleotide: the B-Z transition in poly(dG-m5dC)·poly(dG-m5dC)." *Proc. Natl. Acad. Sci. USA* **78**, 1619-1623.

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

Brünger, A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." *Nature* **355**, 472-475.

Brünger, A. T. (1992). X-PLOR Version 3.1: A system for X-ray Crystallography and NMR. New Haven, Yale University Press.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." *Acta Crystallogr.* **D54**, 905-921.

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

Carter, C. W., Jr. (1995). "The nucleoside deaminases for cytidine and adenosine: structure, transition state stabilization, mechanism, and evolution." *Biochimie* **77**, 92-98.

Cheatham, T. E., 3rd, Crowley, M. F., Fox, T. and Kollman, P. A. (1997). "A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions." *Proc. Natl. Acad. Sci. USA* **94**, 9626-9630.

Cheatham, T. E., 3rd and Kollman, P. A. (1996). "Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution." *J. Mol. Biol.* **259**, 434-444.

Cheatham, T. E., 3rd and Kollman, P. A. (1997). "Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]$_2$ in the presence of hexaamminecobalt(III)." *Structure* **5**, 1297-1311.

Clegg, R. M., Murchie, A. I., Zechel, A., Carlberg, C., Diekmann, S. and Lilley, D. M. (1992). "Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction." *Biochemistry* **31**, 4846-4856.

Connolly, M. L. (1983). "Solvent-accessible surfaces of proteins and nucleic acids." *Science* **221**, 709-713.

Cooper, J. P. and Hagerman, P. J. (1987). "Gel electrophoretic analysis of the geometry of a DNA four-way junction." *J. Mol. Biol.* **198**, 711-719.

Cox, M. M. (2001). "Recombinational DNA repair of damaged replication forks in Escherichia coli: questions." *Annu. Rev. Genet.* **35**, 53-82.

Dickerson, R. E. (1992). "DNA structure from A to Z." *Methods Enzymol.* **211**, 67-111.

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

Diederichs, K. and Karplus, P. A. (1997). "Improved R-factors for diffraction data analysis in macromolecular crystallography." *Nature Struct. Biol.* **4**, 269-275.

Diekmann, S. and McLaughlin, L. W. (1988). "DNA curvature in native and modified EcoRI recognition sites and possible influence upon the endonuclease cleavage reaction." *J. Mol. Biol.* **202**, 823-834.

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

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1981). "Structure of a B-DNA dodecamer: conformation and dynamics." *Proc. Natl. Acad. Sci. USA* **78**, 2179-2183.

Egli, M., Usman, N., Zhang, S. and Rich, A. (1992). "Crystal structure of an Okazaki fragment at 2-Å resolution." *Proc. Natl. Acad. Sci. USA* **89**, 534-538.

Eichman, B. F., Mooers, B. H., Alberti, M., Hearst, J. E. and Ho, P. S. (2001). "The crystal structures of psoralen cross-linked DNAs: drug-dependent formation of Holliday junctions." *J. Mol. Biol.* **308**, 15-26.

Eichman, B. F., Ortiz-Lombardía, M., Aymamí, J., Coll, M. and Ho, P. S. (200x). "The inherent properties of DNA four-way junctions: comparing the crystal structures of Holliday junctions." *Submitted.*

Eichman, B. F., Vargason, J. M., Mooers, B. H. and Ho, P. S. (2000). "The Holliday junction in an inverted repeat DNA sequence: sequence effects on the structure of four-way junctions." *Proc. Natl. Acad. Sci. USA* **97**, 3971-3976.

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

El Hassan, M. A. and Calladine, C. R. (1997). "Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps." *Phil. Trans. R. Soc. Lond. A* **355**, 43-100.

Esnouf, R. M. (1999). "Further additions to MolScript version 1.4, including reading and contouring of electron-density maps." *Acta Crystallogr.* **D55**, 938-940.

Franklin, R. E. and Gosling, R. G. (1953). "Molecular configuration in sodium thymonucleate." *Nature* **171**, 740-741.

Frederick, C. A., Saal, D., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. and Rich, A. (1987). "The crystal structure of d(GGm5CCGGCC): the effect of methylation on A-DNA structure and stability." *Biopolymers* **26**, S145-S160.

Frederico, L. A., Kunkel, T. A. and Shaw, B. R. (1990). "A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy." *Biochemistry* **29**, 2532-2537.

Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J. H. and Rich, A. (1982). "Molecular structure of $(m5dC\text{-}dG)_3$: the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA." *Nucleic Acids Res.* **10**, 7879-7892.

Graessmann, M. and Graessmann, A. (1993). DNA methylation, chromatin structure and the regulation of gene expression. DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz. Boston, Birkhäuser Verlag, 404-424.

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

Hendrickson, W. A. and Ogata, C. M. (1997). "Phase determination from multiwavelength anomalous diffraction measurements." *Methods Enzymol.* **276**, 494-523.

Ho, P. S. and Eichman, B. F. (2001). "The crystal structures of DNA Holliday junctions." *Curr. Opin. Struct. Biol.* **11**, 302-308.

Hodges-Garcia, Y. and Hagerman, P. J. (1992). "Cytosine methylation can induce local distortions in the structure of duplex DNA." *Biochemistry* **31**, 7595-7599.

Holliday, R. (1964). "A mechanism for gene conversion in fungi." *Genet. Res.* **5**, 282-304.

Holliday, R. (1993). Epigenetic inheritance based on DNA methylation. DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz. Boston, Birkhäuser Verlag, 452-468.

Horton, N. C. and Finzel, B. C. (1996). "The structure of an RNA/DNA hybrid: a substrate of the ribonuclease activity of HIV-1 reverse transcriptase." *J. Mol. Biol.* **264**, 521-533.

Hotchkiss, R. D. (1948). "The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography." *J. Biol. Chem.* **168**, 315-332.

Huheey, J. E. (1978). Inorganic Chemistry: Principles of Structure and Reactivity, 2nd Ed. New York, Harper and Row, Publishers.

Jones, P. L. and Wolffe, A. P. (1999). "Relationships between chromatin organization and DNA methylation in determining gene expression." *Semin. Cancer. Biol.* **9**, 339-347.

Karow, J. K., Constantinou, A., Li, J. L., West, S. C. and Hickson, I. D. (2000). "The Bloom's syndrome gene product promotes branch migration of holliday junctions." *Proc. Natl. Acad. Sci. USA* **97**, 6504-6508.

Kielkopf, C. L., Ding, S., Kuhn, P. and Rees, D. C. (2000). "Conformational flexibility of B-DNA at 0.74 Å resolution: $d(CCAGTACTGG)_2$." *J. Mol. Biol.* **296**, 787-801.

Kim, J. L., Nikolov, D. B. and Burley, S. K. (1993). "Co-crystal structure of TBP recognizing the minor groove of a TATA element." *Nature* **365**, 520-527.

Kim, Y., Geiger, J. H., Hahn, S. and Sigler, P. B. (1993). "Crystal structure of a yeast TBP/TATA-box complex." *Nature* **365**, 512-520.

Kleywegt, G. J. and Jones, T. A. (1994). "Detection, delineation, measurement and display of cavities in macromolecular structures." *Acta Crystallogr.* **D50**, 178-185.

Lavery, R. and Sklenar, H. (1989). "Defining the structure of irregular nucleic acids: conventions and principles." *J. Biomol. Struct. Dyn.* **6**, 655-667.

Lilley, D. M. J. (2000). "Structures of helical junctions in nucleic acids." *Quart. Rev. Biophys.* **33**, 109-159.

Lu, X.-J., Shakked, Z. and Olson, W. K. (2000). "A-form conformational motifs in ligand-bound DNA structures." *J. Mol. Biol.* **300**, 819-840.

Malinina, L., Fernandez, L. G., Huynh-Dinh, T. and Subirana, J. A. (1999). "Structure of the d(CGCCCGCGGGCG) dodecamer: a kinked A-DNA molecule showing some B-DNA features." *J. Mol. Biol.* **285**, 1679-1690.

Mao, C., Sun, W. and Seeman, N. C. (1999). "Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy." *J. Am. Chem. Soc.* **121**, 5437-5443.

Mayer-Jung, C., Moras, D. and Timsit, Y. (1998). "Hydration and recognition of methylated CpG steps in DNA." *EMBO J.* **17**, 2709-2718.

Merritt, E. A. and Bacon, D. J. (1997). "Raster3D: photorealistic molecular graphics." *Methods Enzymol.* **277**, 505-524.

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

Murchie, A. I., Clegg, R. M., von Kitzing, E., Duckett, D. R., Diekmann, S. and Lilley, D. M. (1989). "Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules." *Nature* **341**, 763-766.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. and Dodson, E. J. (1999). "Efficient anisotropic refinement of macromolecular structures using FFT." *Acta Crystallogr.* **D55**, 247-255.

Nathan, D. and Crothers, D. M. (2002). "Bending and Flexibility of Methylated and Unmethylated EcoRI DNA." *J. Mol. Biol.* **316**, 7-17.

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

Ng, H.-L., Kopka, M. L. and Dickerson, R. E. (2000). "The structure of a stable intermediate in the A<–>B DNA helix transition." *Proc. Natl Acad. Sci. USA* **97**, 2035-2039.

Nishimura, Y., Torigoe, C. and Tsuboi, M. (1986). "Salt induced B–>A transition of poly(dG)·poly(dC) and the stabilization of A form by its methylation." *Nucleic Acids Res.* **14**, 2737-2748.

Nowakowski, J., Shim, P. J., Prasad, G. S., Stout, C. D. and Joyce, G. F. (1999). "Crystal structure of an 82-nucleotide RNA-DNA complex formed by the 10-23 DNA enzyme." *Nature Struct. Biol.* **6**, 151-156.

Nowakowski, J., Shim, P. J., Stout, C. D. and Joyce, G. F. (2000). "Alternative conformations of a nucleic acid four-way junction." *J. Mol. Biol.* **300**, 93-102.

Noyer-Weidner, M. and Trautner, T. A. (1993). Methylation of DNA in prokaryotes. DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz. Boston, Birkhäuser Verlag, 39-108.

Ortiz-Lombardía, M., González, A., Eritja, R., Aymamí, J., Azorín, F. and Coll, M. (1999). "Crystal structure of a DNA Holliday junction." *Nature Struct. Biol.* **6**, 913-917.

Otwinowski, Z. and Minor, W. (1997). "Processing of x-ray diffraction data collected in oscillation mode." *Methods Enzymol.* **276**, 307-326.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. and Berman, H. M. (1996). "New parameters for the refinement of nucleic acid-containing structures." *Acta Crystallogr.* **D52**, 57-64.

Pflugrath, J. W. (1999). "The finer things in x-ray diffraction data collection." *Acta Crystallogr.* **D55**, 1718-1725.

Sasaki, H., Allen, N. D. and Surani, M. A. (1993). DNA methylation and genomic imprinting in mammals. DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz. Boston, Birkhäuser Verlag, 469-486.

Schneider, B., Cohen, D. and Berman, H. M. (1992). "Hydration of DNA bases: analysis of crystallographic data." *Biopolymers* **32**, 725-750.

Selsing, E., Wells, R. D., Alden, C. J. and Arnott, S. (1979). "Bent DNA: visualization of a base-paired and stacked A-B conformational junction." *J. Biol. Chem.* **254**, 5417-5422.

Sha, R., Liu, F., Millar, D. P. and Seeman, N. C. (2000). "Atomic force microscopy of parallel DNA branched junction arrays." *Chem. Biol.* **7**, 743-751.

Sheldrick, G. M. and Schneider, T. R. (1997). "SHELXL: High-resolution refinement." *Methods Enzymol.* **277**, 319-343.

Singer-Sam, J. and Riggs, A. D. (1993). X-chromosome inactivation and DNA methylation. DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz. Boston, Birkhäuser Verlag, 358-384.

Smith, G. R. (2001). "Homologous recombination near and far from DNA breaks: alternative roles and contrasting views." *Annu. Rev. Genet.* **35**, 243-274.

Spruck, C. H., III, Rideout, W. M., III and Jones, P. A. (1993). DNA methylation and cancer. DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz. Boston, Birkhäuser Verlag, 487-509.

Takahara, P. M., Rosenzweig, A. C., Frederick, C. A. and Lippard, S. J. (1995). "Crystal structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin." *Nature* **377**, 649-652.

Timsit, Y. and Moras, D. (1992). "Crystallization of DNA." *Methods Enzymol.* **211**, 409-429.

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

Trantírek, L., Stefl, R., Vorlícková, M., Koca, J., Sklenár, V. and Kypr, J. (2000). "An A-type double helix of DNA having B-type puckering of the deoxyribose rings." *J. Mol. Biol.* **297**, 907-922.

Vargason, J. M., Eichman, B. F. and Ho, P. S. (2000). "The extended and eccentric E-DNA structure induced by cytosine methylation or bromination." *Nature Struct. Biol.* **7**, 758-761.

Vargason, J. M. and Ho, P. S. (2001). "Response to Ng and Dickerson." *Nature Struct. Biol.* **8**, 107-108.

Wang, A. H.-J., Fujii, S., van Boom, J. H. and Rich, A. (1982). "Molecular structure of the octamer d(G-G-C-C-G-G-C-C): modified A-DNA." *Proc. Natl Acad. Sci. USA* **79**, 3968-3972.

Watson, J. D. and Crick, F. H. C. (1953). "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid." *Nature* **171**, 737-738.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. J. and Weiner, P. (1984). "A new force field for molecular mechanical simulation of nucleic acids and proteins." *J. Am. Chem. Soc.* **106**, 765-784.

Whitehurst, C. E., Schlissel, M. S. and Chen, J. (2000). "Deletion of germline promoter PD beta 1 from the TCR beta locus causes hypermethylation that impairs D beta 1 recombination by multiple mechanisms." *Immunity* **13**, 703-714.

Wiebauer, K., Neddermann, P., Hughes, M. and Jiricny, J. (1993). The repair of 5-methylcytosine deamination damage. DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz. Boston, Birkhäuser Verlag, 510-522.

Winn, M. D., Isupov, M. N. and Murshudov, G. N. (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." *Acta Crystallogr.* **D57**, 122-133.

Xia, F., Taghian, D. G., DeFrank, J. S., Zeng, Z. C., Willers, H., Iliakis, G. and Powell, S. N. (2001). "Deficiency of human BRCA2 leads to impaired homologous recombination but maintains normal nonhomologous end joining." *Proc. Natl. Acad. Sci. USA* **98**, 8644-8649.

Xu, Q., Shoemaker, R. K. and Braunlin, W. H. (1993). "Induction of B-A transitions of deoxyoligonucleotides by multivalent cations in dilute aqueous solution." *Biophys. J.* **65**, 1039-1049.

Zhang, X. and Mathews, C. K. (1994). "Effect of DNA cytosine methylation upon deamination-induced mutagenesis in a natural target sequence in duplex DNA." *J. Biol. Chem.* **269**, 7066-7069.